Large-scale comparison of immunoassay- and aptamer-based plasma proteomics through genetics and disease

Grimur Hjorleifsson Eldjarn^{1,7}, Egil Ferkingstad^{1,7}, Sigrun H. Lund¹, Hannes Helgason¹, Olafur Th. Magnusson¹, Thorunn A. Olafsdottir¹, Bjarni V. Halldorsson^{1,2}, Pall I. Olason¹, Florian Zink¹, Sigurjon A. Gudjonsson¹, Gardar Sveinbjornsson¹, Magnus I. Magnusson¹, Agnar Helgason^{1,3}, Asmundur Oddsson¹, Gisli H. Halldorsson¹, Magnus K. Magnusson^{1,4}, Saedis Saevarsdottir^{1,4}, Thjodbjorg Eiriksdottir¹, Gisli Masson¹, Hreinn Stefansson¹, Ingileif Jonsdottir^{1,4}, Hilma Holm¹, Thorunn Rafnar¹, Pall Melsted^{1,5}, Jona Saemundsdottir¹, Gudmundur L. Norddahl¹, Gudmar Thorleifsson¹, Magnus O. Ulfarsson^{1,6}, Daniel F. Gudbjartsson^{1,5}, Unnur Thorsteinsdottir^{1,4}, Patrick Sulem¹✉ & Kari Stefansson^{1,4}✉

¹deCODE genetics/Amgen, Inc., Reykjavik, Iceland. ²School of Science and Engineering, Reykjavik University, Reykjavik, Iceland. ³Department of Anthropology, University of Iceland, Reykjavik, Iceland. ⁴Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland. ⁵School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland. ⁶Faculty of Electrical and Computer Engineering, University of Iceland, Reykjavik, 101, Iceland. ⁷These authors contributed equally: Grimur Hjorleifsson Eldjarn, Egil Ferkingstad.

✉ e-mail: patrick.sulem@decode.is; kstefans@decode.is

18 Summary

19 High-throughput proteomics platforms measuring thousands of proteins in blood combined with
 20 genomic information have the power to bridge the gap between the genome and diseases and in that
 21 capture some of the environmental contributions to their risk and pathogenesis. Although such
 22 methods have already demonstrated their utility¹⁻³, the validation of their actual protein targets is
 23 lacking. Here we present a large-scale analysis of levels of proteins in plasma and protein quantitative
 24 trait loci (pQTLs) detected using the Olink Explore 1536 (1,459 immunoassays) in 47,151 European
 25 participants from the UK Biobank with 57.7 million imputed sequence variants. We compared the
 26 results with those of a large-scale SomaScan v4 study² (35,559 participants and 4,907 aptamer-based
 27 assays) in order to assess and compare the qualities of these two platforms. The correlation between
 28 levels of proteins targeted by the two platforms is modest (median Spearman correlation 0.46). The
 29 vast majority of assays on the Olink Explore platform had *cis* pQTLs, evidence that they correctly target
 30 their intended proteins (84%), while the assays on the SomaScan v4 platform were half as likely to have
 31 *cis* pQTLs (38%). We also highlight novel pQTLs discovered using the Olink Explore platform, not
 32 captured by SomaScan v4, and describe their colocalization with disease-associated sequence variants
 33 as well as associations between protein levels and diseases. Our results further underscore the value
 34 of proteomics data and highlight the major differences in quality between the two most commonly
 35 used high-throughput proteomics platforms.

Introduction

The adoption of high-throughput proteomics and its integration with genomic data has helped to develop a deeper understanding of the relationship between sequence variants and diseases and other traits²⁻⁶. This has uncovered sequence determinants of plasma protein levels (protein quantitative trait loci, pQTLs) and biomarkers of diseases and their progression. Although associations between protein levels and diseases are rarely sufficient to separate cause from effect, by combining pQTLs with associations between diseases and sequence variants it is often possible.

The SomaScan platform, using aptamer-based assays, has until recently been the dominant high-throughput method for plasma proteomics, with the largest study reported to date being the one we reported using the SomaScan v4 platform on 36 thousand Icelanders with rich health-related and genotype information². With the competing Olink platform, using immunoassays, reaching maturity, large-scale orthogonal validation has become possible for a subset of the proteins measured using each platform⁷. Such large-scale validation is important since validation of the measurements of such a large number of individual proteins is a prohibitively difficult undertaking.

pQTLs are classified as *cis* if the pQTL is close to the gene encoding the protein or otherwise *trans*. Replication of pQTLs detected using SomaScan v4 has been attempted using the Olink Target platform³ for a limited set of proteins (n=90 and 163). In those studies 65% (81% for *cis* and 52% for *trans*) and 67% (93% for *cis* and 64% for *trans*) of pQTLs identified in the Interval⁶ and deCODE^{2,3} studies replicated, respectively. Reciprocally, 64% (75% for *cis* and 61% for *trans*) of the pQTLs identified using the Olink Target platform³ were replicated using the SomaScan v4 platform (n=163 proteins)^{2,3}. A recent study⁸ involved a larger number of proteins targeted by both platforms (n=871), but was limited by the low number of participants measured using the Olink Target platform (n=485).

Here we report on the analysis of the first large-scale study using the Olink Explore platform to measure plasma samples from 47,151 participants in the UK Biobank study (UKB) with in-depth health information⁹. A total of 57.7 million sequence variants identified through whole-genome sequencing were imputed into the chip genotyped genomes of all of the participants, allowing for analysis of the effects of both rare and common sequence variants on protein levels. This data set is comparable to the SomaScan v4 36K Icelandic data set² in size and depth of genotyping and provides well-powered orthogonal mutual validation of the SomaScan v4 and Olink Explore platforms.

Results

Data sets

We used the Olink Explore 1536 platform with 1,459 assays targeting 1,450 unique proteins (Supplementary Table 1) to measure plasma protein levels of 54,303 UKB participants with 57.7 million imputed sequence variants discovered through whole-genome sequencing of 150,119 individuals from UKB¹⁰. Of the UKB participants with protein level measurements, 47,151 individuals were of British or Irish ancestry. We refer to this data set as UKB 47K (Figure 1).

We had previously measured plasma protein levels of 35,559 Icelanders using the SomaScan v4 platform with 4,907 assays targeting 4,719 unique proteins and with 27.2 million sequence variants discovered through whole-genome sequencing of 49,708 Icelanders². We refer to this data set as Iceland 36K (Figure 1).

A total of 1,097 proteins are targeted by both platforms, with 1,180 assays on SomaScan v4 and 1,108 assays on Olink Explore making up 1,193 pairs of assays with a matching target (Supplementary Table 2). To directly assess the correlation between protein levels measured with these matching assays, we measured 1,075 Icelanders from the 36K data set with the Olink Explore platform using the same plasma sample. This data set will be referred to as Iceland 1K (Figure 1). We refer to the outcome of a particular assay as the level of the protein, noting that the assay may not in fact measure the targeted protein.

Comparison of precision of the two platforms

We assessed the precision of the assays by calculating the standard deviation (SD) of repeated measurements of the same sample (Olink Explore: 575, SomaScan v4: 429) after standardizing the protein levels for each protein. The median SD for Olink Explore assays was smaller than for SomaScan v4 (0.49 and 0.60, Mann-Whitney $p=2.2 \times 10^{-64}$). However, the distribution of SDs was wider for the Olink Explore assays (first and third quartiles 0.36 and 0.67) than for SomaScan v4 assays (first and third quartiles 0.54 and 0.67) (Figure 2). This indicates that Olink Explore has on average higher precision but more outliers than SomaScan v4. However, the difference in median precision was smaller when restricting to the matching assays on the two platforms (Olink Explore median SD 0.52, SomaScan v4 median SD 0.58, Wilcoxon signed rank $p=9.0 \times 10^{-4}$).

Between-platform correlation of protein levels for shared targets

Using the Iceland 1K data set, the median correlation between levels from matching assays on the two platforms was 0.46 (Spearman correlation from -0.32 to 0.93, first and third quartiles 0.17 and 0.65).

The distribution of the correlation coefficients appeared to have two modes, one just above 0 and the other around 0.7 (Figure 2b). In comparison, the median pairwise correlation for all possible assay pairs

from the set of matching assays ($n=1,108 \times 1,180=1,307,440$) was 0.08 (first and third quartiles 0.01 and 0.18).

Five matching pairs of assays have a strong negative correlation (<-0.18) with those targeting PILRA having the strongest negative correlation (-0.32). As previously described, PILRA has a *cis* pQTL with opposite directions of effect on the Olink Explore and SomaScan v4 platforms⁸, a result that we also observe in our data (Arg78Gly; MAF=32%; Olink Explore $p=7.9 \times 10^{-7823}$, effect=-1.01 SD and SomaScan v4 $p=1.0 \times 10^{-9045}$, effect=1.31). We observe that the range of levels for matching assays was similar on the two platforms (IQR correlation 0.51, $p=8.2 \times 10^{-81}$). Furthermore, if either of a pair of matching assays had a low median level, or a low variation in level, the correlations between the levels of proteins measured with the matching assays tended to be low (Supplementary Note S1).

Olink Explore and SomaScan v4 both use several dilution groups for their assays based on the reported abundance of the proteins in plasma¹¹⁻¹³. The median correlation between levels for matching assays was dependent on the dilution group, as defined by Olink, with a median of 0.30, 0.49, 0.54 and 0.53 for no dilution, 10-fold, 100-fold and 1000-fold dilutions, respectively (p -value between no dilution vs. other $< 1.9 \times 10^{-7}$, Figure 2c). This is consistent with more difficulties with measuring lower-abundance proteins than the higher-abundance ones.

Between-platform ranking of assays by correlation of levels

To assess whether the matching assays on the two platforms are measuring the same proteins, we compared the levels of all pairs of targeted proteins between platforms. For each assay on both platforms, we rank it based on correlations with all assays on the other platform. In the case of the SomaScan v4 the matching assay ranked within the top 1% in 592 of 1,180 cases. Conversely, on the Olink Explore platform the matching assay ranked within the top 1% in 607 of 1,108 cases (Extended Data Figures 1 and 2). This indicates that at least half of matching assays are measuring the same protein.

Detection of pQTLs using the Olink platform in UK Biobank

Using Olink Explore in the UKB 47K data set, we identified *cis* and *trans* pQTLs, including sentinel and secondary associations based on conditional analysis, using the same approach as we previously applied to the Iceland 36K data set measured with SomaScan v4². We found 17,724 sentinel pQTL associations ($p < 1.8 \times 10^{-9}$) (Supplementary Table 3). pQTLs were found for proteins measured with 96% of 1,459 Olink Explore assays. *Cis* pQTLs were present for 84% of assays and *trans* pQTLs for 94% of assays. Of the sentinel pQTL associations, 1,227 were in *cis* and 16,497 were in *trans*. Each variant may be associated with the level of more than one protein; the sentinel variants called pQTLs below are

6,584 in total if associated variants in high linkage disequilibrium (LD) ($r^2 > 0.80$) are called as one. In addition, we found 10,249 secondary pQTL associations, of which 6,534 were in *cis* and 3,715 in *trans* (Supplementary Note S2).

Of all 27,973 pQTL associations, 1,685 (6%, involving 934 pQTLs) were with variants with a minor allele frequency (MAF) below 0.1% and 2,285 (8%, involving 962 pQTLs) were with variants with MAF between 0.1% and 1%.

For each *cis* pQTL, we determined whether the variant is in high LD ($r^2 > 0.80$) with a coding or splice variant (collectively, protein-altering variants; PAVs) and whether it is a sentinel *cis* expression quantitative trait locus (eQTL) in one or more of 73 tissue types as previously described². Of the 1,227 sentinel *cis* pQTLs, 24% were in high LD with PAVs only, 30% were in high LD with *cis* eQTL only, and 9% were in high LD with both (Supplementary Table 4).

Of the 6,584 pQTLs, 35 pQTLs are associated with the levels of 50 or more proteins, including 14 pQTLs that are associated with levels of 100 or more proteins (Supplementary Note S3).

pQTL comparison between the Olink and SomaScan platforms

The presence of a *cis* pQTL for a particular protein provides evidence that the assay is, in fact, measuring the targeted protein. The fraction of assays with a *cis* pQTL for the targeted protein can therefore serve as an indicator of how many of the assays on a given platform are in fact measuring the intended protein.

We compared the pQTLs found with Olink Explore in the UK, as described above, to those we previously reported using SomaScan v4 data on 36K Icelanders² (Table 1, Extended Data Table 2). Of the 4,907 SomaScan v4 assays, 38% have a *cis* pQTL, 90% have a *trans* pQTL, and 94% have at least one pQTL, either *cis* or *trans*.

Olink Explore assays are 2.2 times more likely to have a *cis* pQTL (84% of 1,459) than SomaScan v4 assays (38% of 4,907). The greater fraction of assays with *cis* pQTLs on the Olink Explore platform indicates that a higher fraction of Olink Explore assays are measuring the intended protein. There are more *trans* than *cis* sentinel pQTLs on both platforms, but a larger number of secondary associations in *cis* than in *trans* (Table 1).

We compared the pQTLs detected with matching assays on the two platforms (Table 1). A total of 1,097 unique proteins are targeted by both platforms, with 1,180 assays on SomaScan v4 and 1,108 assays on Olink Explore. Restricting to those assays, the Olink Explore assays were still 1.5 times more likely to have a *cis* pQTL (87% of 1,108) than the SomaScan v4 assays (59% of 1,180).

Of the pQTLs identified using Olink Explore we replicated 69% of sentinel *cis* pQTLs and 55% of sentinel *trans* pQTLs in the Icelandic SomaScan v4 data (Supplementary Table 5). Conversely, we replicated 86% of sentinel *cis* pQTLs and 65% of sentinel *trans* pQTLs detected using SomaScan v4 in the UKB Olink Explore data (Supplementary Table 6). Thus, in addition to higher overall fraction of proteins with *cis* pQTLs in the Olink Explore data, the replication rate for *cis* pQTLs detected using SomaScan v4 in Olink Explore data was 1.3 times higher than for those detected using Olink Explore and replicated using SomaScan v4 (95% CI: 1.20-1.32). Assays targeting proteins with *cis* pQTLs that do not replicate between platforms have lower correlation between measured protein levels than assays for proteins with replicating *cis* pQTLs (Figure 3a, b, d, e, Supplementary Tables 5 and 6, Supplementary Note S4). This indicates that a greater fraction of these assays is not measuring the same protein on the two platforms.

The fraction of *cis* pQTLs correlating with PAVs (33% and 35% for Olink Explore and SomaScan v4, respectively) or *cis* eQTLs (39% and 34%, respectively) in high LD ($r^2 > 0.80$) was similar on the two platforms. We have previously concluded that the presence of a PAV and the absence of eQTL could be evidence for the association resulting from a binding artefact². When attempting to replicate sentinel *cis* pQTLs between platforms, we observed that when *cis* pQTLs are in high LD with PAV, the correlations between effect estimates were lower in both directions (Supplementary Note S4 and Figure 3), indicating that the antibody binding in the Olink Explore assays is also affected by PAVs. Of the 1,097 proteins targeted with both SomaScan v4 and Olink Explore, 616 (56%) had a *cis* pQTL detected using the two platforms. For 248 of these 616 proteins (40%), the sentinel *cis* pQTLs detected using SomaScan v4 and Olink Explore were in high LD ($r^2 > 0.80$) (Supplementary Table 7). Thus, even when *cis* pQTLs for a protein are detected by both platforms, they are not necessarily the same.

A large fraction of the pQTLs associating with the levels of many proteins (>50) are specific to either of the two data sets (Supplementary Note S3 and Supplementary Tables 8 and 9).

***Cis* pQTLs on the SomaScan and Olink platforms and their colocalization with disease-associated variants**

Colocalization of a variant associating with a disease and a *cis* pQTL makes it likely that the variant is mediating risk through the associated protein. We identified all variants reported in the NHGRI-EBI Catalog of human genome-wide association studies (GWAS catalog¹⁴) colocalizing ($r^2 > 0.80$) with sentinel pQTLs (Supplementary Table 10), based on either Olink Explore in UKB 47K, as described above, or SomaScan v4 in Iceland 36K, as we have previously reported². For the analysis of disease-colocalizing *cis* pQTLs we first focused on the proteins targeted by both platforms.

Proteins targeted by both platforms

With the Olink Explore data, we demonstrated independent replication of colocalization of *cis* pQTLs with disease-associating variants that we previously reported using SomaScan v4². These include a variant associating with lower SULT2A1 levels and less risk of gallstones and a variant associating with lower CHRD2 levels and less risk of colorectal cancer (Supplementary Note S5).

For the proteins targeted by both platforms we observed, in the Olink Explore data only, *cis* pQTLs that colocalized with several disease-associating variants. At the *IL10* (interleukin-10) locus, the minor allele (A) of rs3024493 (MAF=15%) and its correlate rs3024505 ($r^2=1$) associate with risk of inflammatory bowel disease (IBD)¹⁵. However, the variants have not been reported to correlate strongly with coding variants, sentinel *cis* eQTL or *cis* pQTL, but *IL10* has been considered the most likely candidate gene in the region based on its anti-inflammatory function and the fact that IL10-deficient mice develop chronic enterocolitis¹⁶. In the UKB data, the IBD risk allele of rs3024493 is the top *cis* pQTL for IL10 and associates with lower plasma level of IL10 ($p=2.3 \times 10^{-51}$, effect=-0.14 SD), but not with any other protein measured using the Olink Explore platform (Figure 4). Interestingly, the missense variant Ser159Gly (rs3135932-G; MAF=17%) in *IL10RA* (interleukin 10 receptor subunit alpha) associates in *trans* with greater IL10 levels measured on the Olink Explore platform (effect=0.26 SD; $p=3.6 \times 10^{-194}$). Ser159Gly also associates in *cis* with higher IL10RA levels (effect=0.18 SD; $p=1.2 \times 10^{-93}$). *IL10RA* is the most commonly reported gene for monogenic IBD¹⁷ and in a meta-analysis of 20,295 IBD cases, we found Ser159Gly to associate with greater risk of IBD (OR=1.05, $p=1.6 \times 10^{-5}$) (Supplementary Note S6).

Other examples of *cis* pQTLs on Olink Explore that colocalize with disease-associating variants include a variant associating with higher TNFSF11 levels and lower risk of autoimmune thyroid disease (AITD), a variant associating with higher ERBB4 levels and lower BMI, and a variant associating with lower IL34 levels and higher risk of Alzheimer's disease (Supplementary Note S6). We did not detect any *cis* pQTLs for IL10, IL10RA, ERBB4, IL34, and TNFSF11 on SomaScan v4. The presence of *cis* pQTLs for these proteins on Olink Explore and the low between-platform correlation of the levels of these proteins indicate that the Olink Explore assays are successfully measuring these proteins whereas the SomaScan v4 assays are not. In the subset of proteins targeted by both platforms, we did not find any clear examples of sentinel *cis* pQTLs only detected using SomaScan v4 that colocalized with disease variants (Supplementary Table 10).

In some instances, we detected *cis* pQTLs for a given protein on both platforms but they had little correlation and only one of them colocalized with a disease-associated variant. As an example, the minor allele of the sentinel *cis* pQTL for CD58 associated with lower protein levels on Olink Explore and

less risk of multiple sclerosis (MS). However, the sentinel *cis* pQTL on SomaScan v4 is in low LD ($r^2=0.04$) with the one on Olink Explore and does not associate with MS (Supplementary Note S7).

Proteins targeted by either SomaScan or Olink

Olink Explore targets 1,475 proteins, of which 300 are not targeted by SomaScan v4, including ITGA11. The minor allele of Val433Met in *ITGA11* (rs2306022; MAF=9%) is the sentinel *cis* pQTL for ITGA11 ($p=2.1 \times 10^{-212}$, effect=-0.36SD) and is reported to confer risk of Dupuytren contracture (OR=1.27) and uterine fibroids (OR=1.12)^{18,19} (Extended Data Figure 5). The fact that the variant associates with ITGA11 levels is a PAV, and not an eQTL, raises the concern that the observed reduction in ITGA11 may be the result of a probe binding artefact. Validation of these result thus needs to be performed using an alternative method.

Of the 4,907 proteins targeted on SomaScan v4, about 3,000 are not targeted by Olink Explore, of which about 1,000 have *cis* pQTLs. We have previously described the *cis* pQTLs for some of these proteins and their colocalization with disease-associated variants (e.g. GREM1 and colorectal cancer; ASIP and melanoma; F11 and venous thromboembolism and STAT3 and MS (Extended Data Figure 6)²).

Predicting disease phenotypes with protein levels

We determined correlations between levels of proteins measured using the two proteomics platforms with disease diagnosis using heart failure and Alzheimer's disease as examples.

In heart failure, we looked for proteins with levels that correlate with the first heart failure diagnosis following plasma collection (n=1369 in Iceland and n=676 in UK). On both platforms, the BNP protein (encoded by *NPPB*, Natriuretic Peptide B) predicted heart failure most significantly (UK (Olink Explore): HR=2.21 per SD; $p=2.0 \times 10^{-251}$, Iceland (SomaScan v4): HR=1.76 per SD; $p=6.1 \times 10^{-80}$), consistent with BNP being a good biomarker of heart failure progression. On Olink Explore the levels of ADM (adrenomedullin) ranked as the fourth best predictive protein for heart failure (HR=2.04 per SD; $p=9.0 \times 10^{-186}$) but had a much weaker association with heart failure based on SomaScan v4 (HR=1.13 per SD; $p=4.8 \times 10^{-6}$). The correlation between levels of ADM measured on the two platforms was low ($r=0.05$). The association of ADM levels and heart failure based on Olink Explore is consistent with reports of elevated circulating ADM in heart failure²⁰. This indicates that while the two platforms agree in their measurement of BNP, SomaScan v4 is either not measuring ADM or measuring a form of ADM not associated with heart failure.

In Alzheimer's disease we looked for correlations of protein levels with the first Alzheimer's disease diagnosis following plasma collection (n=389 in Iceland and n=224 in UK). Among the proteins targeted by both platforms, and most significantly associated with Alzheimer's disease, is NEFL (Neurofilament

Light) with a strong effect on both platforms (rank 2 and 4 of all assays on Olink Explore and SomaScan v4, respectively), but with opposite direction of effect (HR=1.73 per SD; $p=9.7 \times 10^{-38}$ on Olink Explore, HR=0.70 per SD; $p=6 \times 10^{-13}$ on SomaScan v4). The correlation between NEFL on the two platforms was low ($r=-0.05$). Neurofilament light is known to form polymers and it remains to be elucidated what forms are measured with the two platforms. Consistent with the NEFL results on Olink Explore, high NEFL levels in CSF and blood have been reported to associate with advanced Alzheimer's disease²¹. The opposite effect on SomaScan v4 indicates that the two platforms are either targeting different forms of the same protein, or different proteins entirely.

Discussion

The large-scale validation of assays on high-throughput proteomics platforms is important to verify both that the assays are in fact targeting the intended proteins and that the levels measured are precise. By analysis of repeated sample measurements, we show that the Olink Explore assays have higher precision than assays on the SomaScan v4 platform, although the difference is smaller among the set of proteins that are targeted by both platforms. Direct comparison of the levels measured with matching assays on the two platforms gave a median correlation of 0.46, with a bimodal distribution with modes around 0 and 0.7. Similar correlation has been reported between proteins measured using Olink Target and SomaScan v4⁸, highlighting a clear difference between the platforms.

The presence of a *cis* pQTL for a protein measured using a particular assay provides evidence that the assay is in fact measuring the intended protein. The large sample size presented here with genotypes and protein levels measured using Olink Explore allowed for such an analysis not previously possible due to small sample size in previous studies comparing Olink and SomaScan v4⁸. Here we show that a twofold higher fraction of Olink Explore assays have *cis* pQTLs than that of SomaScan v4 assays (84% vs 38%). The low fraction of *cis* pQTLs in the SomaScan v4 data we have previously observed² is comparable to that of the second largest SomaScan v4 study to date (10,708 individuals), that detected *cis* pQTLs for 32% of the targeted proteins⁵. Our results are thus consistent with the hypothesis that vast majority of Olink Explore assays measure the intended proteins. In contrast, less than half of the SomaScan v4 assays have *cis* pQTL evidence of measuring the targeted protein.

When restricting the analysis to 1,097 proteins targeted with both platforms, a higher fraction of the proteins on both platforms have *cis* pQTLs and the difference between Olink Explore and SomaScan v4 in the fraction of *cis* pQTLs remains, but is smaller (87% vs. 59%). In keeping with these data, we also show that sentinel pQTLs detected using SomaScan v4 replicated better in Olink Explore data (86% of *cis* pQTL and 62% of sentinel *trans*) than Olink Explore pQTLs in SomaScan v4 data (68% of *cis* pQTLs and 59% of *trans* pQTLs).

Of the 1,097 proteins targeted by both SomaScan v4 and Olink Explore, 616 (56%) had a *cis* pQTL on both platforms. For 248 only of these 616 proteins (40%), the sentinel *cis* pQTLs for Olink Explore and SomaScan v4 were in high LD ($r^2 > 0.80$). Thus, even when *cis* pQTLs for a protein are detected by both platforms, suggesting correct targeting, the measured protein levels are not the same.

We and others have reported that pQTL colocalization with variants associating with diseases and other traits can shed light on how the variants affect the risk of the diseases and shed light on their pathogenesis.

We identified novel pQTLs using Olink Explore that colocalize with disease-associating variants not previously detected using SomaScan v4 although the corresponding protein is allegedly also measured with SomaScan v4. In all these examples, the protein levels correlated poorly between the platforms. This suggests that the Olink Explore assays are measuring the intended proteins but the SomaScan v4 assays are not. In other instances, different sentinel *cis* pQTLs in low LD were detected for a given protein using Olink Explore and SomaScan v4, but only one of them colocalized with a disease-associated variant.

When incorporating protein levels in studies of diseases and other traits, differences between platforms in targeting accuracy and measurement precision will influence the results. When directly associating protein levels with diseases we observed associations based on one platform but not the other (e.g., ADM and heart failure on Olink Explore only). We also observed associations with opposite effect for the two platforms (e.g., NEFL and Alzheimer's disease).

In recent years, platforms that simultaneously measure a large number of proteins have become available. Here, taking advantage of two populations with a large number of genotyped individuals with plasma protein measurements, we were able to compare and contrast two high-throughput proteomic platforms, the SomaScan v4 and the Olink Explore platforms. While the assays on the two platforms appear to have comparable precision, we present evidence indicating that the Olink Explore platform is more likely to measure the intended protein than the SomaScan v4 platform, but for proteins of particular interest, orthogonal validation of the measurements is prudent to verify the results.

Methods

Study populations

Three different data sets from Iceland and UK Biobank were used for the protein measurements. A population of 40,034 Icelanders was measured using the SomaScan v4 platform, of which genotype data was available for 35,559. This data set will be referred to as Iceland 36K and has been published previously².

The UK Biobank resource is a large-scale prospective study that includes data from 500,000 volunteer participants who were recruited between the age of 40-69 years in 2006-2010 across the United Kingdom (<https://www.ukbiobank.ac.uk/>). Various health records and health related information is available and regularly updated for these 500,000 participants. Plasma samples were taken from these individuals at the baseline visit, and for a subset of 54,265 individuals the levels of 1,472 proteins were measured with the Olink Explore 1536 platform as a part of The UK Biobank – Pharma Proteomics Project (UK Biobank Application Number 65851). Large majority of the samples were randomly selected across the UK Biobank. Of the 54,265 participants, 47,151 were of British Irish ancestry and with available genotype data¹⁰, and were used in the study. Sequence- and genotype data can be accessed via the UKB research analysis platform (RAP)¹⁰, while proteomics data will be made publicly available in accordance with UKB policy. This data set will be referred to as UKB 47K.

A subset of 1,075 individuals from the Iceland 36K set were measured using the Olink Explore 1536 platform in addition to the SomaScan v4 platform. This subset will be referred to as Iceland 1K.

Plasma protein measurements

The SomaScan v4 platform was used for measurement of the Iceland 36K data set as previously described². It provides measurements for 4,907 aptamer-based assays targeting 4,719 proteins. The assays on the SomaScan v4 platform are based on *aptamers*: short, single-stranded oligonucleotides that bind to specific protein targets. The bound aptamers are then quantified using DNA microarray technology^{13,22}. The SomaScan v4 platform has been extensively used in proteomics studies, including the study of the correlations between sequence variants and protein levels^{2,5,6,23}. The largest study reported to date in up to 36 thousand individuals used the SomaScan v4 platform, which targets around 5000 proteins².

The Olink Explore 1536 platform was used for measurement of the UKB 47K data. It provides measurements for 1,521 antibody-based assays targeting 1,475 proteins. The Iceland 1K data set was measured on both platforms. The assays on the Olink Explore platform are each based on two antibodies. The antibodies bind to different sites on the target protein and are labelled with single-stranded complementary oligonucleotides. If matching pairs of antibodies bind to the protein, the attached oligonucleotides hybridize, and are then measured using next generation sequencing (NGS)^{11,24}. The Olink platform is offered either as 15 panels, each with 45 or 92 assays analyzed by qPCR

(Olink Target), or as four panels of 384 assays analyzed by NGS (Olink Explore 1536). The Olink Target panels have been used both as a primary research instrument^{3,25} and for comparison with the SomaScan v4 platform^{6,8,26}. The Olink Explore panels were released in 2020 and the only previous study using that platform²⁷ had a relatively small population size (n=101 individuals).

Dilution groups

Both the Olink Explore and SomaScan v4 platforms use dilutions of the plasma samples to compensate for different concentrations of proteins in plasma^{11–13}. As the two classifications are in good agreement for the set of proteins targeted by both platforms, we adopt the classification used by Olink Explore (Extended Data Table 1).

Sample preparation for Olink proteomics measurements

The UK Biobank plasma samples were measured using the Olink Explore 1536 platform (<https://www.olink.com/products-services/explore/>) at Olink's facilities in Uppsala Sweden. All samples were randomized and plated by the UK Biobank laboratory team prior to delivery. Samples were processed across three NovaSeq 6000 Sequencing Systems, leveraging the Olink Explore technology. Extensive quality control measures and normalization of protein concentration was performed at Olink's facilities, producing Normalized Protein eXpression (NPX) values for each protein per participant. NPX is Olink's relative protein quantification unit on log2 scale.

The plasma samples in 1K Icelandic dataset were measured with Olink Explore at deCODE facility in accordance with the Olink Explore manual²⁸. Quality control measures were the same as used by Olink for the UK Biobank samples.

Normalization of proteomics measurements

Measurements using the Olink Explore platform use the NPX values recommended by the manufacturer, which include normalization²⁹.

When normalizing the SomaScan v4 data, the final step of scaling the protein measurements to levels from reference samples was not performed³⁰. This step has been shown to have a detrimental effect on the correlation between protein levels when comparing measurements made using the Olink Explore and SomaScan v4 platforms⁸.

Genotyping

The 47K participants in the UK Biobank with Olink Explore measure plasma protein levels in this study had 57.7 million imputed sequence variants discovered through whole-genome sequencing of 150,119 individuals from UKB¹⁰.

The 36K Icelanders with SomaScan v4 measurements had 27.2 million imputed sequence variants discovered through whole-genome sequencing of 49,708 Icelanders².

Additional annotation of assay targets

Genomic coordinates were assigned to assay targets using UniProt IDs³⁶ for each assay provided by the manufacturer. In cases where multiple UniProt IDs were assigned to the same assay, the assay was excluded from further analysis. Out of 4,963 valid assays on the SomaScan v4 platform (excluding non-human proteins and assays marked as defective by the manufacturer), this resulted in 4,961 assays getting assigned the genomic coordinates of their intended targets. Out of 1,473 valid assays on the Olink Explore platform, this resulted in 1,472 assays getting assigned the genomic coordinates of their intended targets.

Assays targeting the same protein were identified using their UniProt IDs. This resulted in 1,193 triads of assay IDs and UniProt ID, targeting 1,097 distinct UniProt IDs. This was done by 1,108 distinct Olink Explore assays and 1,180 distinct SomaScan v4 assays.

Comparison of assay intensities

The SD of the protein levels measured using each assay were estimated using repeated samples from the Iceland 36K and UKB 47K data sets. While the protein levels are not directly comparable between the platforms, the standard deviation of each assay can be used to obtain information about the relative difference between the signal strength and the noise level of each assay, and thus evaluate the precision of the measurement³⁷ (Supplementary Tables 11 and 12).

The protein level measurements for each assay were standardized by setting $\bar{v}_i = (v_i - \bar{v})/\sigma(v)$ where σ is the standard deviation and both mean and standard deviation are calculated using all measurement values for the assay, i.e. not only the repeated samples.

After standardization, standard deviation for measurements using each assay was estimated using only the m pairs of repeated samples using the formula $SD = \sqrt{\sum_{j=1}^m d_j^2 / 2m}$, where d_j is the difference between the measurements for sample j ³⁷.

Correlation of assays across platforms

All correlations of protein levels measured in the same sample were evaluated using Spearman correlation. Correlations between protein levels measured using matching assays and correlations for all pairs were calculated without adjusting for age and sex (Supplementary Table 2).

pQTL analysis

Analysis of pQTLs for the Olink Explore results on the UKB 47K data set was carried out in the same way as described for the Iceland 36K data set in Ferkingstad *et al.*² The following three sections briefly describe this process.

Genome-wide Association Study

The measured intensities for each assay were subjected to rank-inverse normal transformation, after which they were adjusted for age, sex, and sample age. The residuals were standardized again using rank-inverse normal transformation and the standardized values used as phenotypes for genome-wide association testing using a linear mixed model (BOLT-LMM³⁸). LD score regression was used to account for inflation in test statistics due to cryptic relatedness and stratification³⁹.

The *p*-values were computed using a likelihood-ratio test and multiple testing significance threshold was determined as 1.8×10^{-9} using Bonferroni correction for all 27.2 million variants tested in the Icelandic data set.

We define a pQTL association to be *cis* if the pQTL is located within 1MB from the transcription start site for the gene that encodes the target protein, as reported by UniProt, and *trans* otherwise.

Conditional analysis

Recursive conditional analysis was performed separately for each assay and each chromosome. For computational efficiency, this analysis was restricted to the candidate set of markers associating with the assay with a *p*-value lower than 5×10^{-6} . If the marker *v1* with the lowest *p*-value had a *p*-value below the global threshold 1.8×10^{-9} , the marker was removed from the candidate set and the association of all other markers in the candidate set was recomputed, conditional on the first marker *v1*. If any marker in the candidate set had *p*-value below the global threshold 1.8×10^{-9} , the marker with the lowest *p*-value was assigned the label *v2*, removed from the candidate set, and the conditional association of the markers remaining in the candidate set calculated, given *v1* and *v2*. This process was repeated until none of the markers in the candidate set associated with the assay with a *p*-value below 1.8×10^{-9} .

Merging pQTLs

Markers from the conditional analysis were considered to belong to the same region if they were within 2MB of each other. Furthermore, the MHC region (chr6:25.5-34.0MB) was treated as a single region. The most significant marker in each region is referred to as the *sentinel marker* for the assay in the region, and other markers as *secondary*.

The „LD-based clumping approach“ proposed by Sun and co-workers⁶ was used to identify pQTLs associating with multiple assays. Using this approach, sentinel variants associating with different assay are considered to belong to the same pQTL if they are in high linkage disequilibrium ($r^2 > 0.8$).

WITHDRAWN
see manuscript DOI for details

Platform	Somascan v4	Olink Explore	Somascan v4	Olink Explore
Assays	All tested assays	All tested assays	Assays targeting proteins also targeted on Olink Explore	Assays targeting proteins also targeted on Somascan v4
Number of assays	4,907	1,459	1,180*	1,108*
Number of individuals	35,559	47,151	35,559	47,151
Population	Iceland	UK biobank	Iceland	UK biobank
% assays with <i>cis</i> pQTL associations	38	84	59	87
% assays with <i>trans</i> pQTL associations	90	94	90	97
% assays without pQTL associations	6	4	7	0.3
% sentinel <i>cis</i> pQTL associations with PAV in high LD ($r^2 > 0.80$)	35	33	34	32
% sentinel <i>cis</i> pQTL associations with <i>cis</i> eQTL in high LD ($r^2 > 0.80$)	34	39	35	39

Table 1: Summary of pQTLs detected on the Olink Explore and SomaScan v4 platforms

*A total of 1,097 proteins are targeted by assays on both platforms

Figure 1: Properties of the data sets used in the proteomics analysis. Proteomics measurements on the UKB 47K data set were performed on the Olink Explore 1536 platform, while measurements on the Iceland 36K data set were performed on the SomaScan v4 platform. The Iceland 1K data set is a subset of the Iceland 36K data set, on which the same samples were measured using the Olink Explore 1536 platform in addition to the SomaScan v4 platform. Measurements in duplicated samples were used to evaluate precision of the assays.

464

465

466 **Figure 2: Protein levels measured by individual assays. a:** Precision of the measurement of repeated
467 samples by platform. After standardizing the levels of each protein using all available data, the
468 standard deviation (SD) of the repeated measurements were calculated for each assay. The median SD
469 for the Olink Explore assays (orange) was lower than the median SD for the SomaScan v4 assays (blue),
470 but with a wider IQR. The Olink Explore assays were evaluated on 429 duplicate measurements from
471 the UKB 47K data set, while the SomaScan v4 assays were evaluated on 575 duplicate measurements
472 from the Iceland 36K data set.

473 **b:** Correlation between measurements for protein levels measured using assays on the Olink Explore
474 and SomaScan v4 platforms in the Iceland 1K data set (Spearman correlation). The distribution is
475 bimodal with modes around 0 and 0.7.

476 **c:** Correlation between protein levels for matching assays broken down by Olink Explore dilution
477 groups. The median correlation for assays targeting proteins that belong to the undiluted group is
478 lower than for assays belonging to other dilution groups.

479

481

482 **Figure 3: Replication of pQTLs between platforms. a, b:** Replication of sentinel *cis* (a) and *trans* (b)
 483 pQTLs detected using Olink Explore (UK biobank) in SomaScan v4 (Iceland) data. For each pQTL, the
 484 plot shows the effect (in units of SD) in SomaScan v4 (y-axis) vs the effect in Olink Explore (x-axis). The
 485 assays were matched on the UniProt ID of their targeted protein. Each point is colored based on the
 486 Spearman correlation between measured protein levels using SomaScan v4 and Olink Explore. The
 487 yellow lines show values where the effect is equal based on SomaScan v4 and Olink Explore, while the
 488 blue lines show a linear regression estimate with shaded 95% pointwise confidence intervals.

489 **c:** Replication of sentinel *cis* pQTLs detected using Olink Explore (UK Biobank) in SomaScan v4 (Iceland)
 490 data, stratified on whether the *cis* pQTL is in high LD with PAV (red) or not (blue). For each pQTL, the
 491 plot shows the effect (in units of SD) in SomaScan v4 (y-axis) vs the effect in Olink Explore (x-axis).
 492 Each point is colored based on whether or not the associated variant has a protein altering variant in
 493 high LD ($r^2 > 0.80$). The yellow line shows values where the effect is equal based on SomaScan v4 and
 494 Olink Explore, while the blue and red lines show linear regression estimates with shaded 95% pointwise
 495 confidence intervals for each group (PAV in high LD; No PAV in high LD).

496 **d, e:** Replication of sentinel *cis* (d) and *trans* (e) pQTL associations detected using SomaScan v4
 497 (Iceland) in Olink Explore (UK biobank) data. For each pQTL association, the plot shows the effect (in
 498 units of SD) in SomaScan v4 (x-axis) vs the effect in Olink Explore (y-axis). Each point is colored based
 499 on the Spearman correlation between measured protein levels using SomaScan v4 and Olink Explore.
 500 The yellow lines show values where the effect is equal based on SomaScan v4 and Olink Explore, while
 501 the blues line show a linear regression estimate with shaded 95% pointwise confidence intervals.

502 **f:** Replication of sentinel *cis* pQTL associations detected using SomaScan v4 (Iceland) in Olink Explore
 503 (UK biobank) data, stratified on whether the *cis* pQTL is in high LD with PAV (blue) or not (red). For
 504 each pQTL association, the plot shows the effect (in units of SD) in SomaScan v4 (x-axis) vs the effect
 505 in Olink Explore (y-axis). Each point is colored based on whether or not the associated variant has a
 506 protein altering variant in high LD ($r^2 > 0.80$). The yellow line shows values where the effect is equal
 507 based on SomaScan v4 and Olink Explore, while the blue and red lines show linear regression estimates
 508 with shaded 95% pointwise confidence intervals for each group (PAV in high LD; No PAV in high LD).

Figure 4: Olink IL10 *cis* pQTL and IBD colocalization. Association at IL10 locus between sequence variants and measured levels of IL10 (a) on Olink Explore, (b) on SomaScan, and (c) inflammatory bowel disease risk (IBD). All r^2 are shown to the same marker. The IL10 protein is targeted by assays on both platforms, but no *cis* pQTLs were observed using the SomaScan platform.

Extended Data

	Olink dilution					
SomaScan dilution		1:1	1:10	1:100	1:1000	total
	0.005	2	2	8	23	35
	0.5	66	79	102	43	290
	20	554	223	80	11	868
	total	622	304	190	77	1,193

Extended Data Table 1: Assay dilution groups

Platform	Somascan v4	Olink Explore	Somascan v4	Olink Explore
Assays	All tested assays	All tested assays	Assays targeting proteins also targeted on Olink Explore	Assays targeting proteins also targeted on Somascan v4
Number of assays	4,907	1,459	1,180*	1,108*
Number of individuals	35,559	47,151	35,559	47,151
Population	Iceland	UK biobank	Iceland	UK biobank
% assays with <i>cis</i> pQTL associations	38	84	59	87
% assays with <i>trans</i> pQTL associations	90	94	90	97
% assays without pQTL associations	6	4	7	0.3
Number of sentinel <i>cis</i> pQTL associations	1,881	1,227	697	998
Number of sentinel <i>trans</i> pQTL associations	16,203	16,497	4,734	13,912
Number of secondary <i>cis</i> pQTL associations	5,691	6,534	2,565	5,650
Number of secondary <i>trans</i> pQTL associations	4,416	3,715	1,366	3,249
% sentinel <i>cis</i> pQTL associations with PAV in high LD ($r^2 > 0.80$)	35	33	34	32
% sentinel <i>cis</i> pQTL associations with <i>cis</i> eQTL in high LD ($r^2 > 0.80$)	34	39	35	39
% sentinel <i>cis</i> pQTL associations without <i>cis</i> eQTL or PAV in high LD ($r^2 > 0.80$)	39	37	39	37

Extended Data Table 2: Summary of pQTLs detected on the Olink Explore and SomaScan v4 platforms

*A total of 1,097 proteins are targeted by assays on both platforms

522

523 Extended Data Figure 1: Histogram of the rank of the matched assay compared to all possible pairs.

524 For an assay on a given platform, the matching assay on the other platform is among the highest-

525 ranked assays on that platform in a majority of cases.

526

527

Extended Data Figure 2: The difference between the highest correlation coefficient for an assay and the correlation coefficient for the matched assay. The histogram shows the distribution of the differences in correlation coefficient between the matching assay and the assay with the strongest correlation. The right-hand plot shows the distribution of the correlation coefficients. For assays on the diagonal, the matching assay is the assay showing the highest correlation in protein levels. For the majority of assays, the difference between the strongest correlation and the correlation with the matching assay is small.

IBD

Olink Explore IL10RA *cis* pQTL

Olink Explore IL10 *trans* pQTL

Extended Data Figure 3: Association at IL10RA locus between variants and protein levels of IL10RA and IL10 measured using Olink Explore and IBD risk. All r^2 are shown to the same marker.

560 Extended Data Figure 4: Association at CD58 locus between variants and CD58 levels measured using
561 Olink Explore, CD58 levels measured using SomaScan v4 and multiple sclerosis risk. All r^2 are shown to
562 the same marker.

Olink Explore

Dupuytren's disease

Extended Data Figure 5: Association at *ITGA11* locus between variants and *ITGA11* levels measured using Olink Explore, and Dupuytren's disease. All r^2 are shown to the same marker.

Supplementary Note

S1. Between-platform correlation of protein levels for shared targets

Both the median protein levels and the protein level IQRs were correlated between the two platforms (protein level correlation coefficient 0.14, p -value 7.6×10^{-7} , IQR correlation coefficient 0.51, p -value 8.2×10^{-81}). The variation in IQR is only to a limited degree explained by variation in median level (R^2 for Olink Explore 0.23, SomaScan v4 0.06).

The correlation coefficient between levels of a protein measured with the two platforms was correlated with the median of the protein levels (Olink Explore: 0.22, p -value 1.9×10^{-14} , SomaScan v4: 0.53, p -value 1.8×10^{-87}) as well as the IQR of the protein levels (Olink Explore: 0.25, p -value 5.19×10^{-18} , SomaScan v4: 0.31, p -value 1.9×10^{-29}). In particular, this was true for the minimum of the protein level medians and IQRs over the two platforms (median level: 0.43, p -value 2.21×10^{-54} , IQR: 0.43, p -value 1.45×10^{-55}). This indicates that if either assay of a pair of matching assays has a low level, or a low variation in level, the correlation with the levels measured with the matching assay tends to be low.

S2. Detection of pQTLs using the Olink platform in UK Biobank

For 19% of the 17,724 sentinel pQTL associations, a secondary variant in the same region associated with the same protein based on conditional analysis (6,534 secondary associations in *cis* and 3,715 in *trans*). Combined, we found a total of 27,973 sentinel and secondary pQTL associations.

Most Olink *cis* pQTLs had secondary pQTLs (1,065 of 1,227; 87%), in contrast to only 2,351 (14%) of the 16,497 *trans* pQTLs.

Of the Olink pQTLs, 4,833 (73%) associated with the level of a single protein, 1523 (23%) with levels of two to nine proteins and 228 (3%) with levels of ten or more proteins. We refer to pQTLs that were associated with levels of fewer than ten proteins as *specific* and the rest as *nonspecific*. Specific variants were associated with 1,352 (93%) of the Olink assays.

Of the 1,459 Olink assays, 2% had only *cis* pQTLs, 12% had only *trans* pQTLs and 82% had both *cis* and *trans* pQTLs. Only 4% had no pQTLs, whereas 5% had a single pQTL and 91% had two or more pQTLs.

S3. Variants associating with a large number of proteins

We assessed the proximity of the variants that associate with the largest number of proteins on SomaScan (38 variants associating with at least 50 of the 5K proteins) and on Olink (228 variants associating with at least 10 of the 1.5K proteins) (Supplementary Tables 8 and 9). A total of 26 of the

38 variants from SomaScan are less than 2Mb from one of the 228 Olink variants. Conversely, 43 of the 228 variants from Olink are less than 2Mb from one of the 38 SomaScan variants.

The three variants associating with the levels of the largest number of proteins on SomaScan but not on Olink are close to *CFH*, *BCHE* and *IGFBP2*, and rank 1,5 and 6 out of 38 based on the number of associated proteins. The three variants associating with the largest number of proteins on Olink but not on SomaScan are close to *SLC22A4*, *COPZ1*, and *GP6*, and rank 6, 7 and 9 out of 264 based on the number of associated proteins. The variant at *GP6* is in high LD with a missense variant (Pro219Ser) in *GP6*, it is an eQTL and a pQTL for GP6, and associates with platelet volume⁴⁰ and aggregation⁴¹. GP6 is the primary glycoprotein receptor that mediates collagen responses in platelets involved in thrombus formation and the difference in association between the platforms may reflect differences in plasma sample preparation and handling^{2,42}, as prolonged storage of blood samples may lead to platelet activation and subsequent secretion of proteins into plasma^{43,44}.

S4. pQTL replication between the Olink and SomaScan platforms

When attempting to replicate sentinel Olink pQTLs using SomaScan, the correlation between effect estimates was lower for cis pQTLs ($r=0.42$; 95% CI: 0.36-0.48) than for trans pQTLs ($r=0.59$; 95% CI: 0.57-0.60) (Figure 3a, b). Furthermore, the correlation between effects was substantially higher for cis pQTLs not in high LD with a PAV ($r=0.56$; 95% CI: 0.50-0.62) than for cis pQTLs in high LD with a PAV ($r=0.31$; 95% CI: 0.19-0.42) (Figure 3c). In contrast, for pQTLs detected using SomaScan, the correlation with effect estimates using Olink was higher for cis pQTLs ($r=0.47$; 95% CI: 0.41-0.53) than for trans pQTLs ($r=0.39$; 95% CI: 0.37-0.42) (Figure 3d, e). However, as in the other direction of replication, the correlation was substantially higher for cis pQTLs not in high LD with a PAV ($r=0.69$; 95% CI: 0.63-0.73) than for cis pQTLs in high LD with a PAV ($r=0.33$; 95% CI: 0.20-0.44) (Figure 3f).

Assays targeting proteins with cis pQTLs that do not replicate between platforms have lower correlation between measured protein levels than assays for proteins with replicating cis pQTLs (Figure 3a, b, d, e, Supplementary Tables 5 and 6). For cis pQTLs detected using the SomaScan v4 platform and replicated on the Olink Explore platform, this is based on 541 replicating and 85 non-replicating pQTLs, while for cis pQTLs detected using the Olink Explore platform and replicated using the SomaScan v4 platform, this is based on 519 replicating and 238 non-replicating pQTLs.

S5. Proteins targeted by both platforms, same cis pQTL for both platforms

The protein encoded by *SULT2A1* (bile salt sulfotransferase) is targeted by both Olink and SomaScan platforms and the cis pQTL from both platforms, rs212100 ($P=10^{-415}$; effect=-0.38 SD) and chr19:47871693 ($P=8 \times 10^{-88}$; effect=-0.21 SD), are in high LD ($r^2=1$; MAF=16%). SULT2A1 levels are

correlated between the platforms ($r=0.47$). We and others have reported association of the minor allele of rs212100 with less gallstone risk⁵⁶ and lower plasma levels of SULT2A1 using the SomaScan v4 platform^{2,5}. The results here using the Olink Explore platform in UKB confer an independent replication. These variants are reported to be in high LD with the sentinel *SULT2A1* eQTL (Lung and Liver effect on RNA=-0.9 SD, Supplementary Table 4), but not with any PAV.

Similarly, we had reported colocalization of a *CHRD2* *cis* pQTL with colorectal cancer risk variant using SomaScan, implicating *CHRD2* at the locus as the gene through which the effect is likely mediated in the absence of any correlated eQTL or PAV². Consistent with these results, the sentinel *cis* pQTL (rs11236228; MAF=4%) for *CHRD2* in Olink is in high LD ($r^2=1$) with the disease associated variant (rs4944940) with similar effect ($P=10^{-373}$; effect=-0.68 SD) to the one observed on Somascan (effect=-0.62SD), thus providing an independent replication.

S6. Proteins targeted by both platforms, *cis* pQTL only detected by one platform

The missense variant Ser159Gly (rs3135932-G; MAF=17%) in *IL10RA* (interleukin 10 receptor subunit alpha) associated in *trans* with greater IL10 levels measured on the Olink platform (effect=0.26 SD; $P=3.6\times 10^{-194}$). Ser159Gly also associated with higher IL10RA levels (effect=0.19 SD; $P=1.2\times 10^{-93}$). *IL10RA* encodes a high affinity homodimer that is specific for IL10 binding and together with the IL10RB (interleukin 10 receptor subunit beta) homodimer forms the signaling receptor complex for IL10. *IL10RA* is the most commonly reported gene for monogenic IBD¹⁷. Ser159Gly reduces the signaling via the IL10 receptor⁴⁷ and has been studied in relation to risk of IBD in candidate gene studies with conflicting results⁴⁷⁻⁵⁰. In a meta-analysis of 20,295 IBD cases, Ser159Gly associated with greater risk of IBD (OR=1.05; $P=1.6\times 10^{-5}$) (Extended Data Figure 3).

The only pQTLs for IL10 and IL10RA on SomaScan were in *trans* together with 20 or more proteins and as such do not constitute evidence of accurate targeting of either protein on the platform. Consistently, the levels of IL10 and IL10RA as measured using Olink Explore and SomaScan v4 have low correlation (IL10 correlation=0.02, IL10RA correlation=0.02 and 0.06 for the two SomaScan v4 assays with the protein as intended target).

The sentinel *cis* pQTL for TNFSF11 (Tumor necrosis factor ligand superfamily member 11, also known as RANKL) using Olink (chr13:42478744; MAF=47% $P=9.1\times 10^{-76}$; effect=0.12 SD) is in high LD ($r^2=0.93$) with an autoimmune thyroiditis (AITD) associated variant (rs9533118; $P=1\times 10^{-14}$ ⁵¹). The minor allele associates with a reduction in AITD risk by 7% and higher TNFSF11 level and is not in LD with a PAV or a *cis* eQTL. The correlation between platforms of the level of TNFSF11 was low ($r=-0.01$). TNFSF11 is involved in establishing self-tolerance⁵² and as such may play a role in autoimmune diseases such as AITD.

The sentinel *cis* pQTL for ERBB4 on Olink associates with higher ERBB4 levels (rs6735267; MAF=29%; $P=1.0 \times 10^{-195}$, effect=0.22 SD) and a highly correlated variant ($r^2=0.99$) associates with lower BMI (rs7599312; $P=2 \times 10^{-23}$, effect=-0.01SD)⁵³. The correlation between the levels of ERBB4 measured by the two platforms was low ($r=-0.01$). We did not observe any eQTL or PAV in high LD with these variants. Supporting the possible implication of *ERBB4* in obesity, this gene encodes a receptor tyrosine kinase expressed in liver and pancreas, and ERBB4 disruption has been linked to impaired glucose tolerance and reduced insulin response in mice⁵⁴.

The sentinel *cis* pQTL on Olink for IL34 is a stop gained variant in *IL34* (rs4985556), located in the in the last exon (position 213 of 243 amino acids) and associates with greater Alzheimer's disease (AD) risk (OR=1.08, p -value= 2×10^{-8})⁵⁵. We did not detect colocalization with any eQTL. The stop gained variant associates with less IL34 ($P=10^{-3279}$; effect=-1.09 SD; MAF=12%) which is consistent with the predicted loss of function. The correlation between the levels of IL34 measured by the two platforms was low ($r=-0.06$). IL34 promotes microglial proliferation and clearance of soluble amyloid- β (A β) and as such confers neuroprotection against the neurotoxicity of A β . It has been proposed that the stop gained variant associating with less IL34 could increase AD risk by reducing the neuroprotective properties of IL34⁵⁵.

S7. Proteins targeted by both platforms, different *cis* pQTL detected by the platforms

In some instances, *cis* pQTLs for a given protein are detected on both platforms but they have little correlation and only one of them colocalizes with a disease associated variants. In the case of the CD58 (Lymphocyte function-associated antigen 3) protein we observed *cis* pQTLs on both the Olink Explore and SomaScan v4 platforms, however, the *cis* pQTLs were different. The sentinel *cis* pQTL on Olink Explore (rs10801908, MAF=12%) and SomaScan v4 (rs61789227; MAF=2.1%) have very different minor allele frequencies and are in low LD ($r^2=0.04$) (Extended Data Figure 4). However, a secondary *cis* pQTL for CD58 on SomaScan v4 (rank 2, rs10754443; MAF=11%; effect=-0.10 SD, $p=3 \times 10^{-16}$) is in high LD ($r^2=0.81$) with the sentinel *cis* pQTL on Olink Explore. None of the *cis* pQTLs on the two platforms correlates with a PAV. The correlation between the levels of CD58 measured by the two platforms was low ($r=0.04$). The sentinel *cis* pQTL (rs10801908) for CD58 on Olink Explore ($p=2.8 \times 10^{-148}$, effect=-0.26 SD) is reported to associate with 25% less risk of multiple sclerosis (MS)^{14,45}. The minor allele of this variant associated with lower CD58 levels on Olink Explore and less MS risk. The glycoprotein CD58, also known as lymphocyte-function antigen 3 (LFA-3), is a costimulatory receptor distributed on a broad range of human tissues including cells of the immune system, making it relevant for autoimmune diseases, including MS⁴⁶. Based on these results both platforms are likely targeting CD58 and the

722 discrepancies observed could result from the two platforms targeting different CD58 isoforms, or
723 multimeric versus monomeric forms.

724

725

726

727

WITHDRAWN
see manuscript DOI for details

References

1. Correa Rojo, A. *et al.* Towards Building a Quantitative Proteomics Toolbox in Precision Medicine: A Mini-Review. *Front. Physiol.* **12**, 723510 (2021).
2. Ferkingstad, E. *et al.* Large-scale integration of the plasma proteome with genetics and disease. *Nat. Genet.* 1–10 (2021) doi:10.1038/s41588-021-00978-w.
3. Folkersen, L. *et al.* Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat. Metab.* **2**, 1135–1148 (2020).
4. Folkersen, L. *et al.* Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. *PLOS Genet.* **13**, e1006706 (2017).
5. Pietzner, M. *et al.* Mapping the proteo-genomic convergence of human diseases. *Science* **374**, eabj1541 (2021).
6. Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
7. Suhre, K., McCarthy, M. I. & Schwenk, J. M. Genetics meets proteomics: perspectives for large population-based studies. *Nat. Rev. Genet.* **22**, 19–37 (2021).
8. Pietzner, M. *et al.* Synergistic insights into human health from aptamer- and antibody-based proteomic profiling. *Nat. Commun.* 2021 121 **12**, 1–13 (2021).
9. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nat.* 2018 5627726 **562**, 203–209 (2018).
10. Halldorsson, B. V. *et al.* The sequences of 150,119 genomes in the UK biobank. 2021.11.16.468246 <https://www.biorxiv.org/content/10.1101/2021.11.16.468246v1> (2021) doi:10.1101/2021.11.16.468246.
11. Assarsson, E. *et al.* Homogenous 96-Plex PEA Immunoassay Exhibiting High Sensitivity, Specificity, and Excellent Scalability. *PLOS ONE* **9**, e95192 (2014).
12. Candia, J. *et al.* Assessment of Variability in the SOMAscan Assay. *Sci. Rep.* **7**, 1–13 (2017).
13. Gold, L. *et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLOS ONE* **5**, e15004 (2010).

14. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
15. Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986 (2015).
16. Kühn, R., Löhler, J., Rennick, D., Rajewsky, K. & Müller, W. Interleukin-10-deficient mice develop chronic enterocolitis. *Cell* **75**, 263–274 (1993).
17. Nambu, R. *et al.* A Systematic Review of Monogenic Inflammatory Bowel Disease. *Clin. Gastroenterol. Hepatol. Off. Clin. Pract. J. Am. Gastroenterol. Assoc.* S1542–3565(21)00331–1 (2021) doi:10.1016/j.cgh.2021.03.021.
18. Ng, M. *et al.* A Genome-wide Association Study of Dupuytren Disease Reveals 17 Additional Variants Implicated in Fibrosis. *Am. J. Hum. Genet.* **101**, 417–427 (2017).
19. Sakaue, S. *et al.* A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* **53**, 1415–1424 (2021).
20. Jougasaki, M., Wei, C. M., McKinley, L. J. & Burnett, J. C. Elevation of circulating and ventricular adrenomedullin in human congestive heart failure. *Circulation* **92**, 286–289 (1995).
21. Teunissen, C. E. *et al.* Blood-based biomarkers for Alzheimer’s disease: towards clinical implementation. *Lancet Neurol.* **21**, 66–77 (2022).
22. 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).
23. Suhre, K. *et al.* Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun.* **8**, 1–14 (2017).
24. Lundberg, M., Eriksson, A., Tran, B., Assarsson, E. & Fredriksson, S. Homogeneous antibody-based proximity extension assays provide sensitive and specific detection of low-abundant proteins in human blood. *Nucleic Acids Res.* **39**, e102 (2011).
25. Graumann, J. *et al.* Multi-platform Affinity Proteomics Identify Proteins Linked to Metastasis and Immune Suppression in Ovarian Cancer Plasma. *Front. Oncol.* **9**, 1150 (2019).

26. Finkernagel, F. *et al.* Dual-platform affinity proteomics identifies links between the recurrence of ovarian carcinoma and proteins released into the tumor microenvironment. *Theranostics* **9**, 6601–6617 (2019).
27. Zhong, W. *et al.* Next generation plasma proteome profiling to monitor health and disease. *Nat. Commun.* **12**, 2493 (2021).
28. Olink Proteomics AB. *Olink Explore 1536 User Manual*.
<https://www.olink.com/content/uploads/2021/12/olink-explore-1536--expansion-user-manual-2021-12-13-v1.1.pdf> (2021).
29. Wik, L. *et al.* Proximity Extension Assay in Combination with Next-Generation Sequencing for High-throughput Proteome-wide Analysis. *Mol. Cell. Proteomics* **20**, (2021).
30. SomaLogic, Inc. *SOMAscan v4 Data Standardization and File Specification Technical Note*.
https://www.mcgill.ca/genepi/files/genepi/bqc19_jgh_prt_tech_note_0.pdf (2018).
31. Gudbjartsson, D. F. *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* **47**, 435–444 (2015).
32. Wain, L. V. *et al.* Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir. Med.* **3**, 769–781 (2015).
33. Welsh, S., Peakman, T., Sheard, S. & Almond, R. Comparison of DNA quantification methodology used in the DNA extraction protocol for the UK Biobank cohort. *BMC Genomics* **18**, 26 (2017).
34. Eggertsson, H. P. *et al.* GraphTyper enables population-scale genotyping using pangenome graphs. *Nat. Genet.* **49**, 1654–1660 (2017).
35. Kong, A. *et al.* Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat. Genet.* **40**, 1068–1075 (2008).
36. The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).

37. Glüer, C.-C. *et al.* *Osteoporosis International Accurate Assessment of Precision Errors: How to Measure the Reproducibility of Bone Densitometry Techniques*. vol. 5 262–270 (1995).
38. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
39. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
40. Chen, M.-H. *et al.* Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. *Cell* **182**, 1198–1213.e14 (2020).
41. Johnson, A. D. *et al.* Genome-wide meta-analyses identifies seven loci associated with platelet aggregation in response to agonists. *Nat. Genet.* **42**, 608–613 (2010).
42. Elliott, P., Peakman, T. C., & on behalf of UK Biobank. The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int. J. Epidemiol.* **37**, 234–244 (2008).
43. Mussbacher, M. *et al.* Optimized plasma preparation is essential to monitor platelet-stored molecules in humans. *PLoS ONE* **12**, e0188921 (2017).
44. Pietzner, M. *et al.* Genetic architecture of host proteins involved in SARS-CoV-2 infection. *Nat. Commun.* **11**, 6397 (2020).
45. International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* **365**, eaav7188 (2019).
46. Zhang, Y., Liu, Q., Yang, S. & Liao, Q. CD58 Immunobiology at a Glance. *Front. Immunol.* **12**, 705260 (2021).
47. Grundtner, P. *et al.* The IL-10R1 S138G loss-of-function allele and ulcerative colitis. *Genes Immun.* **10**, 84–92 (2009).
48. Brophy, K. *et al.* Evaluation of 6 candidate genes on chromosome 11q23 for coeliac disease susceptibility: a case control study. *BMC Med. Genet.* **11**, 76 (2010).

49. Ghaith, O. A. *et al.* Lack of correlation between IL-10R1 S138G loss-of-function allele and IBD in the Lebanese population. *Inflamm. Bowel Dis.* **16**, 1819–1820 (2010).
50. Wagner, J. *et al.* Interaction of Crohn’s disease susceptibility genes in an Australian paediatric cohort. *PLoS One* **5**, e15376 (2010).
51. Saevarsdottir, S. *et al.* FLT3 stop mutation increases FLT3 ligand level and risk of autoimmune thyroid disease. *Nature* **584**, 619–623 (2020).
52. Ono, T., Hayashi, M., Sasaki, F. & Nakashima, T. RANKL biology: bone metabolism, the immune system, and beyond. *Inflamm. Regen.* **40**, 2 (2020).
53. Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* **28**, 166–174 (2019).
54. Zeng, F., Wang, Y., Kloepper, L. A., Wang, S. & Harris, R. C. ErbB4 deletion predisposes to development of metabolic syndrome in mice. *Am. J. Physiol. Endocrinol. Metab.* **315**, E583–E593 (2018).
55. de Rojas, I. *et al.* Common variants in Alzheimer’s disease and risk stratification by polygenic risk scores. *Nat. Commun.* **12**, 3417 (2021).
56. Ferkingstad, E. *et al.* Genome-wide association meta-analysis yields 20 loci associated with gallstone disease. *Nat. Commun.* **9**, 5101 (2018).