nature

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41586-018-0175-2

In the format provided by the authors and unedited.

Genomic atlas of the human plasma proteome

Benjamin B. Sun[1,22], Joseph C. Maranville[2,20,22], James E. Peters[1,3,22], David Stacey[1], James R. Staley[1], James Blackshaw[1], Stephen Burgess[1,4], Tao Jiang[1], Ellie Paige[1,5], Praveen Surendran[1], Clare Oliver-Williams[1,6], Mihir A. Kamat[1], Bram P. Prins[1], Sheri K. Wilcox[7], Erik S. Zimmerman[7], An Chi[2], Narinder Bansal[1,8], Sarah L. Spain[9], Angela M. Wood[1], Nicholas W. Morrell[3,10], John R. Bradley[11], Nebojsa Janjic[7], David J. Roberts[12,13], Willem H. Ouwehand[3,14,15,16,17], John A. Todd[18], Nicole Soranzo[3,14,16,17], Karsten Suhre[19], Dirk S. Paul[1], Caroline S. Fox[2], Robert M. Plenge[2,20], John Danesh[1,3,16,17]*, Heiko Runz[2,21,23] & Adam S. Butterworth[1,17,23]*

[1]MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [2]MRL, Merck & Co., Inc., Kenilworth, NJ, USA. [3]British Heart Foundation Cambridge Centre of Excellence, Division of Cardiovascular Medicine, Addenbrooke's Hospital, Cambridge, UK. [4]MRC Biostatistics Unit, University of Cambridge, Cambridge, UK. [5]National Centre for Epidemiology and Population Health, The Australian National University, Canberra, Australian Capital Territory, Australia. [6]Homerton College, Cambridge, UK. [7]SomaLogic Inc, Boulder, CO, USA. [8]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. [9]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. [10]Division of Respiratory Medicine, Department of Medicine, University of Cambridge, Cambridge, UK. [11]NIHR Cambridge Biomedical Research Centre/BioResource, Cambridge University Hospitals, Cambridge, UK. [12]National Health Service (NHS) Blood and Transplant and Radcliffe Department of Medicine, NIHR Oxford Biomedical Research Centre, University of Oxford, John Radcliffe Hospital, Oxford, UK. [13]BRC Haematology Theme and Department of Haematology, Churchill Hospital, Oxford, UK. [14]Department of Haematology, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK. [15]National Health Service (NHS) Blood and Transplant, Cambridge Biomedical Campus, Cambridge, UK. [16]Department of Human Genetics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. [17]NIHR Blood and Transplant Research Unit in Donor Health and Genomics, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [18]JDRF/Wellcome Trust Diabetes and Inflammation Laboratory, Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, NIHR Oxford Biomedical Research Centre, University of Oxford, Oxford, UK. [19]Department of Physiology and Biophysics, Weill Cornell Medicine–Qatar, Doha, Qatar. [20]Present address: Celgene Inc., Cambridge, MA, USA. [21]Present address: Biogen Inc., Cambridge, MA, USA. [22]These authors contributed equally: Benjamin B. Sun, Joseph C. Maranville, James E. Peters. [23]These authors jointly supervised this work: Heiko Runz, Adam S. Butterworth. *e-mail: jd292@medschl.cam.ac.uk; asb38@medschl.cam.ac.uk

Supplementary Information

Funding

This research was supported as follows. The Cardiovascular Epidemiology Unit at the University of Cambridge: UK MRC (G0800270), BHF (SP/09/002), UK NIHR Cambridge Biomedical Research Centre, ERC (268834), and European Commission Framework Programme 7 (HEALTH-F2-2012-279233); B.B.S.: Cambridge School of Clinical Medicine MB-PhD programme and MRC/Sackler Prize PhD Studentship (MR/K50127X/1); J.E.P.: a BHF Cambridge Centre of Excellence Research Fellowship [RE/13/6/30180] and a UK Research Innovation Fellowship (MR/S004068/1). P.S.: a Rutherford Fund Fellowship (MR/S003746/1); D.S.P. and D.S.: the Wellcome Trust (105602/Z/14/Z); N.S.: the Wellcome Trust (WT098051 and WT091310), the EU FP7 (EPIGENESYS 257082 and BLUEPRINT HEALTH-F5-2011-282510); J.A.T: the Wellcome Trust (091157) and JDRF (9-2011-253); K.S.: the Biomedical Research Program at Weill Cornell Medicine in Qatar via the Qatar Foundation; J.D.: BHF Professor, European Research Council Senior Investigator, and NIHR Senior Investigator; the INTERVAL study: NHSBT (11-01-GEN) and the NIHR-BTRU in Donor Health and Genomics (NIHR BTRU-2014-10024) at the University of Cambridge in partnership with NHSBT. Data analysis was partly supported by the Cambridge Substantive Site of Health Data Research UK. This study was partially funded by Merck Sharp & Dohme Corp. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health of England, or NHSBT.

Supplementary Table Legends

Supplementary Table 1. Replication of age and sex associations reported previously using the SOMAscan platform.
Reported associations are from Menni et al., 2014 and Ngo et al., 2016. Estimates adjusting for other factors (BMI and eGFR) are also listed for comparison.
BMI = body-mass index; eGFR = estimated glomerular filtration rate.

Supplementary Table 2. Significant ($p<1x10^{-5}$) associations between protein levels and age, sex, BMI and eGFR.
n=3,301 individuals. P-values from linear regression (reported p-values are unadjusted for multiple testing). SOMAmer ID: Identifier of the SOMAmer; UniProt: Uniprot ID; Adjusted R^2: Protein level variance explained (adjusted R^2) by age, sex, BMI and eGFR; Subcohort passed: Whether the protein passed QC in subcohorts 1, 2 or both. BMI = body-mass index; eGFR = estimated glomerular filtration rate.

Supplementary Table 3. Cross-reactivity testing of 920 SOMAmers against homologous proteins.
Table lists each of the target proteins (and the genes encoding those proteins) as well as the related proteins tested for cross-reactivity. Cross-reactivity is classified as 'No binding observed', 'Binding at least 10-fold weaker than target' or 'Comparable binding observed'.

Supplementary Table 4. Summary statistics for 1,927 sentinel variant-protein associations.
Locus ID: unique identifier for the locus in the format CHR_locusnumber; Chromosome (Chr) and position (Pos) are in build GRCh37/hg19. Region start and region end are the boundaries for the locus as defined in Methods; EAF: effect allele frequency; INFO: Imputation information score; cis: variant within 1Mb of the gene(s) encoding the protein/protein complex, trans: variant >1Mb from the gene(s) encoding the protein/protein complex; Mapped gene: cis variants are mapped to the gene encoding the protein, trans variants are mapped based on nearest gene using Ensembl annotations; beta: per-allele effect on protein levels in standard deviation units with respect to the effect allele (EA); Previously reported: same locus-protein association reported previously in the literature (Supplementary Table 20); Replicates: NA = not assayed in the Olink validation study, Yes = p-value for replication<$3.1x10^{-4}$, No = p-value for replication>$3.1x10^{-4}$; Uncorrelated with PAV (r^2>0.1): absence of protein altering variant(s) in LD at r^2>0.1 with the sentinel variant, 1 = yes, 0 = no. To be conservative, for PAVs with MAF>5% in ExAC not present in the imputed genetic data or if the gene is not found in ExAC, we assign the value of 0.
**: 2 loci (also highlighted in Supplementary Table 5) that were no longer considered trans after conditional analysis of the top variant for that protein in cis.*

Supplementary Table 5. Summary of conditional analysis results for significant pQTLs.
n=3,301 individuals. P-values from GCTA conditional analysis of summary statistics from linear regression. Reported p-values are nominal p-values, unadjusted for multiple correction. The "Joint Model" column shows the association of the variant in the joint multi-variant model (see Methods). Yellow cells in the "Joint Model P" column highlight variants with genome-wide significant associations in univariable analyses ($p=1.5x10^{-11}$) which are close to genome-wide significance in the joint model. Orange cells show the two associations in Supplementary Table 4 that were no longer significant in the joint model (and thus no longer considered trans

in Supplementary Table 4) after including the top variant for that protein in cis. "Univariable beta/SE/p" columns show the univariable associations for the Conditional variant. Betas represent the per-allele joint model effect estimate in standard deviations with respect to the Conditional variant's effect allele (EA).

Supplementary Table 6. Replication of 163 pQTLs using the Olink platform.
n=3,301 individuals for SomaLogic assay, and n=4,998 for Olink assays. Reported P-values are nominal p-values from linear regression.

Supplementary Table 7. Summary statistics for *cis* pQTLs after adjustment for protein altering variants (PAVs).
"Univariable beta/SE/p": the univariable associations of the sentinel variant prior to PAV adjustment (as in Supplementary Table 4); "No. of PAVs": Number of PAVs in the gene(s) encoding the protein; "PAV-adjusted beta/SE/p": the association estimates for the variant after adjusting for all PAVs present in the genetic data in gene(s) encoding the protein, estimates are set to NA (null) for sentinel variants where at least one PAV is in very strong LD (r^2>0.9). Variants with a PAV-adjusted p<5x10^{-8} are highlighted in green. "No. PAVs not present in imputed data": number of protein altering variants (MAF>5%) present in ExAC data not available for testing in the post-QC imputed genetic data.

Supplementary Table 8. *Cis* pQTL overlap with *cis* eQTL data (p<1x10^{-5}).
Sentinel pQTL variant or its strongest proxy (r^2>0.8) were cross-referenced for eQTLs for the same gene as the gene(s) encoding the protein at p<1x10^{-5}. Overlapping eQTL are reported in the format: **(Number of entries, 1 per cell type per study): rsID, LD r^2, eQTL p-value (coloc PP4 - GTEx tissues) [cell-type, study, PMID]**, *with entries separated by ";".*

Supplementary Table 9. Enrichment analysis for regulatory features in ENCODE and ROADMAP.
Results ordered and shaded by fold enrichment values. Fold enrichment values in bold are statistically significant at p<5x10^{-5} (0.05/998 tests). TFBS = transcription factor binding sites; TSS = transcription start site.

Supplementary Table 10. Table of eQTL studies used to compare overlap of pQTLs with eQTLs.

Supplementary Table 11. Enrichment of *cis* pQTLs for *cis* eQTLs using GTEx Consortium data.
Post-PAV: analyses restricted to pQTLs that remain significant at p<5x10^{-8} after conditioning on protein-altering variants.

Supplementary Table 12. Bioinformatic annotation of candidate genes at *trans* pQTLs using "bottom up" and "top down" approaches from ProGeM.

Supplementary Table 13. Examples of *trans* pQTLs with known relevant biological links.

Supplementary Table 14. pQTL overlap with disease-associated loci from GWAS (p<5x10^{-8}).
List of overlapping disease GWAS hits from publicly available sources (via PhenoScanner) curated to retain only one entry per disease. Non-disease phenotypes such as anthropometric traits, intermediate biomarkers and lipids were excluded. Where the sentinel protein-

associated variant was not available, the strongest proxy (r^2>0.8) was used. Overlapping variants are reported in the format: ***Disease trait (year [Pubmed ID]).*** *The posterior probability of colocalisation (PP4) of disease association and pQTL is provided for each SOMAmer, as is the evidence against aptamer-binding effects.*

Supplementary Table 15. Protein counts for drug/pQTL/GWAS overlap by development stage and *cis/trans* status.

Supplementary Table 16.

a) **Licensed drugs where the drug target overlaps a pQTL and a GWAS signal matching the indication.**
   ** bold indicates match with current drug indication.*
   *** ustekinumab has been used for ankylosing spondylitis in clinical trials.*
   # the genetic signal lies in IL12. IL-23 comprises IL-23p19 and IL-12/23p40 subunits, therefore this is a cis signal.
b) **Drug targets where *cis* pQTL overlaps a GWAS signal with a similar indication.**
   **GPIBA example identified based on manual review of cis pQTL for drug targets that overlap GWAS hits, rather than matching MedDRA terms.*
c) **Potential drug targets for therapeutic inhibition.**
   Here we list proteins whose plasma levels are increased by disease risk alleles. Disease loci with pleiotropic effects on proteins were manually excluded to provide a list of targets that are more likely to be specific.

Supplementary Table 17. Baseline characteristics of study participants.
**Age and BMI: p-values from two-sided t-test for difference between subcohorts. Gender, smoking and alcohol use: p-values from two-sided Chi-squared test for difference between subcohorts.*

Supplementary Table 18. List of the 3,622 proteins assayed on the SOMAscan platform.

Supplementary Table 19. Previous pQTL GWAS.

Supplementary Table 20. Previously reported pQTLs (p<5x10^{-8}) with proteins overlapping those tested identified through literature search and the NHGRI-EBI GWAS Catalog.

Supplementary Table 21. Summary of replication evidence for previously reported pQTLs.

Supplementary Note

Supplementary Note Table 1. Numerical breakdown of the number of associations identified.

	Total number	***Cis* only**	***Trans* only**	**Both**
Sentinel variant to protein associations	1927	549 (28%)	1378 (72%)	N/A
Genomic regions with pQTLs	764	502 (66%)	228 (30%)	34 (4%)
Proteins with pQTLs	1478	374 (25%)	925 (63%)	179 (12%)

1) Reliability of protein measurements using the SOMAscan platform

We used several distinct analytical and experimental approaches to evaluate the reliability of the protein measurements and the protein-genotype associations that we identified. These complementary approaches provide evidence for the reliability of protein measurements made using the SOMAscan assay, and are set out below.

a) Baseline reproducibility of plasma protein measurements

To assess the concordance of repeat SOMAscan measurements on the same sample, we assayed 40 replicate pooled plasma samples from INTERVAL participants (Methods). Measurements in replicate samples were highly consistent; the median coefficient of variation (CV) across all proteins was 0.064 (interquartile range 0.049-0.092) and 96% of proteins had a CV<0.2 (Extended Data Figure 3a).

b) Longitudinal stability of protein levels

We compared contemporaneous measurements from baseline samples and samples taken two years later from 60 participants. This analysis showed temporal consistency in protein levels within individuals, with 64% of proteins having a Spearman's correlation above 0.5 (median

0.56) between the two time-points, comparable to some clinically useful biomarkers such as C-reactive protein and major lipid fractions (Extended Data Figure 3b).

c) Replication of associations of proteins with non-genetic factors

Of 45 proteins previously associated with age using the SOMAscan platform[1-2], we replicated 41 (91%) associations at $p<1.1x10^{-3}$, whilst for 40 proteins previously associated with sex[1], we replicated 38 (95%) at $p<1.2x10^{-3}$ (Supplementary Table 1). We also replicated established associations between proteins and other participant phenotypes, such as estimated glomerular filtration rate (eGFR), a measure of renal function, with cystatin C ($p=1.7x10^{-6}$) and beta-2-microglobulin ($p=4.2x10^{-6}$)[3]; and body-mass index (BMI) with leptin ($p=1.1x10^{-441}$), insulin ($p=1.3x10^{-25}$), and ghrelin ($p=1.7x10^{-17}$) (Supplementary Table 2). Replication of these previous findings provides valuable 'positive control' associations, indicating that these proteins are being correctly targeted by the relevant SOMAmers. In addition, our results provide evidence that protein quantification is consistently reflecting relative protein abundance across studies, even those conducted some years apart using different versions of the SOMAscan platform.

d) Replication of previously reported genetic associations

We identified 119 published genetic association studies of plasma or serum protein levels, measured using a wide variety of assays. Of these, 58 publications included proteins that we analysed here (Supplementary Table 19). These studies identified pQTLs for 367 proteins, of which 284 were analysed in our study (Supplementary Table 20). To avoid issues arising from differences in p-value thresholds between studies, we applied a filter of $p<5x10^{-8}$ across all studies to identify significant associations. Of a total of 432 pQTLs that had been previously reported with $p<5x10^{-8}$, we replicated 258 (59.7%) at $p<1x10^{-4}$ in our data, with a higher

proportion of *cis* (83.8%) than *trans* (39.6%) replicating at this threshold (Supplementary Table 20). 81.4% (210) of the pQTLs that replicated at $p<1x10^{-4}$ also reached our Bonferroni-corrected genome-wide significance threshold ($p<1.5x10^{-11}$).

e) Identification of biologically plausible pQTLs

The 554 *cis* pQTLs that we identified demonstrate that the SOMAmer reagents (hereafter 'SOMAmers') for those proteins are binding the correct targets, since it is extremely unlikely that the one region of the genome that contains the protein-encoding gene would be significantly associated ($p<1.5x10^{-11}$) with levels of the protein by chance.

Similarly, identification of *trans* pQTLs with strong plausibility due to prior biological knowledge can help to validate the protein assays. A few compelling examples are presented in Supplementary Table 13, including an intronic variant (rs7787942) in *C1GALT1* that was associated for the first time with levels of C1GALT1-specific chaperone 1 (C1GALT1C1) ($p=1.0x10^{-28}$), consistent with suggestions that C1GALT1C1 is a probable molecular chaperone for C1GalT1[4]. Another such example is our identification of rs855791 in the *TMPRSS6* gene region as a pQTL for transferrin receptor protein 1 (TR1) is unsurprising given the well-established roles of matriptase-2 (the protein product of *TMPRSS6*) and TR1 in iron homeostasis.

2) External evidence for specific *cis* pQTLs at protein-altering variants

Associations between protein levels and genetic variants in LD with non-synonymous polymorphisms that affect protein structure could be driven by differential aptamer binding to alternative protein isoforms rather than true quantitative differences in protein abundance. Although it is difficult to distinguish these two possibilities bioinformatically, in specific

instances external biological evidence supports the presence of a true abundance pQTL, as illustrated by the well-characterised example of a *cis* pQTL for IL6R. We identified an association of rs4129267 with IL-6R levels (Supplementary Table 4). This variant is in strong LD (r^2>0.95) with rs8192284 (p.Asp358Ala), which has been previously associated with plasma IL-6R levels measured using other assay methods[5]. rs8192284 has been shown to influence plasma IL6R by altering shedding of the receptor from the cell surface[6]. IL-6R is cleaved from the cell surface by proteases ADAM10 and ADAM17, and the cleavage site lies between residues 357 and 358[7]. The p.Asp358Ala variant makes the IL-6R protein more susceptible to protease cleavage[8], therefore influencing shedding of the receptor from the cell surface. This example demonstrates how pQTLs at protein-altering variants can reflect true differences in plasma protein abundance rather than just differential aptamer binding to alternative protein isoforms.

3) Validation of high-abundance proteins using mass spectroscopy

In order to identify the SOMAmers that were used in the SOMAscan assay, individual purified proteins from a variety of sources were used for *in vitro* selections in optimized buffer conditions[9]. To test the ability of SOMAmers identified under such optimized conditions to recognize endogenous proteins in a complex biological fluid such as plasma, we tested a subset of 140 SOMAmers to the most abundant plasma proteins for their ability to pull-down cognate proteins and confirmed their identity by mass spectroscopic analysis. More specifically, for each characterization sample, a single SOMAmer sequence functionalized with a photocleavable 5' biotin moiety (identical to that used in the SOMAscan assay) was used to enrich its protein target from human plasma. SOMAmers were immobilized on Streptavidin-coated agarose beads and incubated with human plasma diluted to 50% in protein binding buffer (40mM HEPES, 120 mM NaCl, 5mM MgCl2, 5mM KCl, pH 7.5) for 3 hours at 28°C.

SOMAmer reagent-protein complexes were then incubated with 10 mM dextran sulfate for 10 minutes at 25°C. The complexes were then washed with protein binding buffer three times for 10 minutes at 25°C and then eluted by photocleavage of the o-Nitrobenzylether linker. Samples were submitted to MS Bioworks (Ann Arbor, MI, USA; www.msbioworks.com) for trypsin digest and peptide identification by LC-MS/MS. Data were analyzed by comparing the numbers of spectral counts corresponding to each unique protein identified within a sample and across a set of samples.

Of the 140 most abundant proteins on the SOMAscan platform, we were able to confirm SOMAmer-mediated, specific enrichment of 123 protein targets. For 10 proteins, we were not able to identify specific enrichment of any proteins in plasma, presumably because these proteins were present at levels below the threshold needed for detection by peptide tandem mass spectrometry. Therefore, for 123 out of 130 proteins (95%), enrichment was detected with the intended analytes. For seven proteins, we identified enrichment of a protein other than the intended target. Upon further investigation, we found that the SOMAmer reagents for these seven proteins were actually selected against product- or process-related impurities in the initial protein preparation used in selections. In four of these instances, the protein associated with each SOMAmer was re-named to reflect the identified enriched protein (three to immunoglobulin G and one to complement factor H; the immunoglobulin G binding is a consequence of the use of Fc-fusion recombinant proteins in selections, for which SomaLogic now have a specificity screen, and complement factor H appears to have been an impurity in the C1s preparation that has higher propensity to bind nucleic acids than C1s). The enriched proteins for the remaining three SOMAmer reagents (low-density lipoprotein receptor-related protein 1B, CD27 antigen, protein SET) could not be unambiguously identified and so identities were not changed.

Overall, these data show that that the large majority of SOMAmers that measure high abundance plasma proteins are specifically targeting the intended protein. This is supported by the large number of *cis* pQTLs and biologically plausible *trans* pQTLs, which also corroborate the correct identity of the targeted protein.

4) Specific pull-down of cognate proteins

Based on analysis of 16 published co-crystal structures of aptamers bound to their protein targets, which include three SOMAmer-protein complexes, it is clear that aptamers engage well-defined epitopes on proteins through shape and functional group complementarity, which is the source of both their high-affinity and specificity for a given epitope[10]. Because there are many proteins that share structural and functional features, it is possible that the conformational epitope to which a SOMAmer binds is also present on other proteins with similarity to the target protein used to select the SOMAmer. To assess the potential for a SOMAmer to bind other proteins likely to contain shared epitopes with the intended target protein, we used publicly available databases of known human protein sequences and sequence alignment tools (e.g., BLAST) to identify those "relevant relative" proteins that share significant homology with proteins used to select the SOMAmers. Any proteins with significant homology to the SOMAmer target protein (defined here as proteins with greater than 40% sequence identity with the target protein or members of the same protein family) were obtained for direct experimental testing.

For this purpose, we tested 920 proteins representing a subset of 1305 proteins from the previous version of SOMAscan that met the following criteria: 1) confirmatory pull-down results with cognate proteins and 2) availability of related protein(s) in the same protein family

or meeting the minimum identity threshold in pure enough form to allow unambiguous specificity testing. When available, we tested cross-reactivity to more than one related protein. We then set up our specificity testing in two stages.

In the first stage, we performed affinity capture experiments similar to immunoprecipitation in binding buffer (40 mM HEPES, 120 mM NaCl, 5mM $MgCl_2$, 5 mM KCl, pH 7.5) at 100 nM concentration of the related protein using 10 pmol of SOMAmer immobilized on streptavidin-coated agarose beads (133 µl of a 7.5% bead slurry, Pierce High Capacity Streptavidin Agarose, Thermo P/N 20357). Binding incubations were performed for 3 hours at 28°C. After binding incubation, bead-immobilized SOMAmer-protein complexes were washed six times with 200 µl of binding buffer to remove unbound protein. The remaining protein was labeled with 50 nM NHS-AlexaFluor647 (Thermo, P/N A37573) for 20 minutes at 25°C, followed by elution and analysis by SDS-PAGE.

For 674 of the 920 SOMAmers tested (73%), we did not detect binding to any related proteins that were tested other than the cognate target. Examples of such experiments for two SOMAmers mentioned in the main text – targeting IL1RL2 and GP1BA – are shown in Extended Data Figure 2, with results for all of the 920 SOMAmers tested in this manner shown in Supplementary Table 3.

In the second stage of testing, for the remaining 246 SOMAmers (27% of 920) which did pull down at least one of the related proteins, we performed solution binding affinity measurements using a filter binding assay[9]. Of these 246 SOMAmers, 120 (13% of 920) bound with at least 10-fold weaker affinity, and 126 (14% of 920) bound with comparable affinity (defined as within 10-fold of that observed for the cognate target) to at least one related target

(Supplementary Table 3). Of these 126 SOMAmers, 54 (6% of 920) exhibited cross-reactivity to at least one product of the same gene with a common epitope (for example, proteins representing proteolytically-processed forms like factor X and Xa, complement factor C5 and C5a, pro-thrombin and thrombin, or splice variants, like vascular endothelial growth factor 121 and 165 isoforms), or shared subunits in a multi-subunit complex (cyclin-dependent kinase 1/cyclin B1 complex, in which the SOMAmer binds to the cyclin B1 subunit).

To assess whether the likelihood of cross-reactivity among pairs of proteins correlates with sequence identity, we calculated the percent amino acid identity for each pair of proteins among the 246 SOMAmers that pulled down at least one related protein, using the accession numbers in the ALIGN program on the UniProt website[11]. Many of these 246 SOMAmers were tested against multiple proteins (Supplementary Table 3), resulting in 499 pairwise tests. In aggregate, the amino acid sequence identity tended to be greater for those pairs that exhibited cross-reactivity: 48±17% (mean ± standard deviation) for 249 pairs that exhibited no cross-reactivity (no positive pull-down results), 62±20% for 147 pairs with >10-fold lower affinity, but positive pull-down results, and 70±18% for 103 pairs with similar affinity. Median values for the three groups (45%, 62% and 70%, respectively) were similar to the mean values. The observed difference in amino acid identity between all pairwise comparisons among the three groups were highly statistically significant (two-tailed p-values$\leq$0.002, Mann-Whitney test). Therefore, the likelihood of binding to a shared epitope on a related protein correlated with higher sequence identity, although the distributions were broad, as expected, because the epitopes for aptamers in general have variable sizes and are often discontiguous[10].

These cross-reactivity results, taken together with pull-down mass spectrometric analysis of abundant proteins described in Supplementary Note Section 3, demonstrate that the large majority of SOMAmers are highly specific for their cognate protein targets.

5) Does cross-reactivity give rise to artefactual *trans* pQTL associations?

Although the overall proportion of SOMAmers cross-reacting to another protein with similar affinity to the intended target was modest (14%), we considered the possibility that such cross-binding SOMAmers could nevertheless substantially impact our results by giving rise to artefactual *trans* pQTLs. To address this, we compared the prevalence of cross-reactivity (binding another protein with similar affinity to the target protein) in SOMAmers with a *trans* pQTL versus those without a *trans* pQTL. This revealed no enrichment (*p*-value = 0.432 by Fisher's exact test; Supplementary Note Table 2).

Supplementary Note Table 2. Contingency table of 920 SOMAmers tested for cross-reactivity according to the presence or absence of a *trans* pQTL.

	No cross-reactivity	**Cross-reactivity**
SOMAmers without *trans* pQTLs	606 (86.9%)	91 (13.1%)
SOMAmers with *trans* pQTLs	189 (84.8%)	34 (15.2%)

A particular concern was that there might be cross-reactivity between a *trans* pQTL target protein and a protein encoded by a gene close to the pQTL variant. In this scenario a true *cis* pQTL might appear as an artefactual *trans* association due to cross-reactivity. To investigate this possibility, we first took the set of target proteins that exhibit cross-reactivity (with either similar affinity or at least 10-fold weaker affinity) for which we had found *trans* pQTLs. Next,

for each of those target proteins, we identified the positions of the genes encoding proteins that were cross-reactive with these targets. Finally, we checked whether these genes were located near the sentinel *trans* pQTL variant(s) for the relevant target protein.

This analysis revealed only one instance where the gene encoding a cross-reactive protein overlapped the relevant *trans* pQTL. The SOMAmer in this case (PRKCB.5475.10.3) bound to the cross-reacting protein (Protein kinase C gamma type, encoded by *PRKCG*) with 10-fold weaker affinity than to its intended target, another protein kinase C family member (Protein kinase C beta type, encoded by *PRKCB*). These results therefore provide reassurance that cross-reactivity is not systematically driving the *trans* pQTLs that we observed.

6) Genetic and proteomic data can inform physical interactions between proteins *in vivo*

Here we describe two examples where pQTL signals reflect altered abundance of protein complexes. These examples serve both to highlight that pQTL signals can provide evidence of physical interaction between proteins and to illustrate potential pitfalls in the interpretation of pQTL signals. Importantly, pQTL signals in complex biological matrices such as plasma may reflect genetically determined differences in the free versus bound forms of a protein.

a) Proteinase-3 and alpha-1-antitrypsin

We identified an association between the rs28929474 at *SERPINA1* and one of two SOMAmers targeting proteinase-3 (PR3) (Figure 2, Figure 4a). Both SOMAmers targeting PR3 (PRTN3.13720.95.3 and PRTN3.3514.49.2) had a *cis* signal, indicating that they were binding their intended target (Figure 4b). However, the *trans* association at *SERPINA1* was specific to PRTN3.3514.49.2. *SERPINA1* encodes alpha-1-antitrypsin (A1AT), a protease inhibitor known to bind PR3. rs28929474:T (the 'Z-allele') is a missense variant which results in an

abnormal form of A1AT that accumulates intracellularly and thus defective secretion of A1AT into the circulation. The Z allele was associated with *lower* levels of PRTN3.3514.49.2, raising that possibility that the *trans* pQTL signal reflected a reduction in the abundance of the PR3:A1AT complex, and that SOMAmer PRTN3.3514.49.2 was preferentially binding PR3 complexed to A1AT over the free form of PR3.

To investigate this further, we assayed the relative affinity of these SOMAmers for the free and complexed states of these two proteins. We first tested the ability of each SOMAmer reagent to enrich either purified protein in its free form or as a PR3:A1AT covalent complex in binding buffer (SB-17: 40 mM HEPES, 100 mM NaCl, 5 mM KCl, 5 mM $MgCl_2$, 1 mM EDTA, pH 7.5 0.05% Tween) containing 20 µM Z-block. Each individual SOMAmer containing a biotin and a photocleavable linker at the 5' end was immobilized on Streptavidin-agarose beads in a manner similar to the multiplexed SOMAmer-bead complexes used in the SOMAscan assay[9]. These SOMAmer-bead complexes were then incubated with either free PR3 at 100 nM (Athens Research and Technology, Athens, GA, USA), free A1T at 100 nM (R&D Systems, Minneapolis, MN, USA), or PR3:A1AT complex formed by incubating PR3 with a three-fold molar excess of A1AT (100 nM PR3, 300 nM A1AT). After incubation, the solution was removed and the SOMAmer-protein complexes were incubated with 10 mM dextran sulfate in binding buffer for 5 minutes. The complexes were then washed and the SOMAmer-bound proteins labeled with NHS-AlexaFluor647. The SOMAmer-protein complexes were then eluted by photocleavage (12 min irradiation with black light, 15-watt, 16-inch, 350 nm light bulb manufactured by Osram Sylvania (Wilmington, MA, USA), at a distance of 6 cm from the sample), separated by SDS-PAGE and visualized by fluorescence imaging.

We found that SOMAmer PRTN3.3514.49.2 (which produced both *cis* and *trans* signals), enriched the PR3:A1AT complex to a greater degree than free PR3, whereas SOMAamer PRTN3.13720.95.3 (which produced only *cis* signal) enriched free PR3 to a greater degree than the PR3:A1AT complex (Extended Data Figure 8a). Both SOMAmers pulled down similar amounts of the PR3:A1AT complex. Importantly, neither SOMAmer pulled down free A1AT at 100 nM protein.

Determination of binding affinity of SOMAmers PRTN3.3514.49.2 and PRTN3.13720.95.3 for free and complexed forms of PR3.

We next determined the binding affinity in solution of each SOMAmer reagent for either free PR3 or the PR3:A1AT complex (formed by adding 3-fold excess of A1AT over PR3). Limiting amounts of ^{32}P-radiolabeled SOMAmer (<10pM) were incubated with a dilution series of either biotinylated free PR3 or biotinylated PR3 in a complex with A1AT. We found that the affinity of PRTN3.3514.49.2 was approximately 5-fold higher for the PR3:A1AT complex (K_d=3x10^{-9} M) than for free PR3 (K_d=2x10^{-8} M). In contrast, the affinity of PRTN3.13720.95.3 was approximately 6-fold higher for free PR3 (K_d=6x10^{-11} M) than for the PR3:A1AT complex (K_d=4x10^{-10} M). SOMAmer PRTN3.13720.95.3 exhibited more than 300-fold higher affinity for free PR3 compared with SOMAmer PRTN3.3514.49.2 (Extended Data Figure 8b).

Competition binding experiment

To test the possibility that both of the PR3-specific SOMAmers bind to a common epitope on the protein, we performed a competition binding experiment. A limiting amount of radiolabeled SOMAmer PRTN3.13720.95.3 (<10pM) was incubated with 1 nM PR3 in the absence or presence of unlabeled SOMAmer PRTN3.3514.49.2 or PRTN3.13720.95.3 over a concentration range from 100 pM to 3 μM. Unlabeled PRTN3.3514.49.2 inhibited the binding

of radiolabeled SOMAmer PRTN3.13720.95.3 to PR3, indicating that the two SOMAmers bind to a common epitope on the protein. Unlabeled PRTN3.13720.95.3, which was used as a control, also inhibited the binding of the labeled PRTN3.13720.95.3 more effectively than PRTN3.3514.49.2, as expected in view of the difference in solution affinities of these two SOMAmers for PR3 (Extended Data Figure 8c).

The two PR3 SOMAmers are present in the same dilution (1%) in the SOMAscan experiment, both at the concentration of 0.5 nM, so they compete for the binding of limiting amounts of free PR3, which is expected to be present at the concentration range of 0.05-0.08 nM (1% of its typical concentration range in plasma)[12]. Since SOMAmer PRTN3.13720.95.3 binds to PR3 with more than 300-fold higher affinity (K_d=6x10^{-11} M) compared with SOMAmer PRTN3.3514.49.2 (K_d=2x10^{-8} M), any free PR3 will be essentially entirely bound to SOMAmer PRTN3.13720.95.3.

The situation is different with the PR3:A1AT complex, where the difference in affinities between the two SOMAmers is much smaller (K_d=3x10^{-9} M for PRTN3.13720.95.3 and K_d=0.4x10^{-9} M for PRTN3.13720.95.3) (Extended Data Figure 8b). Therefore the PR3:A1AT complex is expected to be distributed between the two SOMAmers, albeit unequally, with more of it bound to SOMAmer PRTN3.13720.95.3. We highlight that the A1AT SOMAmer is present in a different SOMAscan dilution (0.005%) because of its higher abundance in plasma, so there cannot be any interference between the PR3 SOMAmers and the A1AT SOMAmer.

Taken together, these results provide a clear explanation for the observed pQTL associations: SOMAmer PRTN3.3514.49.2 measures only the PR3:A1AT complex, whereas SOMAmer PRTN3.13720.95.3 measures both the free and complexed forms of PR3. In the presence of

the Z allele, reduced secretion of A1AT leads to reduced abundance of the PR3:A1AT complex (Figure 4c), and this is reflected in the read-out from SOMAmer PRTN3.3514.49.2, producing a *trans* pQTL association signal at *SERPINA1*. Conversely, SOMAmer PRTN3.13720.95.3, detects both free and bound forms of PR3. In the presence of the Z allele, the reduced read-out from this SOMAmer due to lower levels of the PR3:A1AT complex is offset by increased detection of free PR3 (which is exclusively measured by this SOMAmer because of the competitive nature of binding of the two SOMAmers and the difference in their relative affinities). Therefore no association at *SERPINA1* is observed for SOMAmer PRTN3.13720.95.3.

In summary, this example highlights that (i) there may be distinct genetic signals for free versus bound forms of a proteins and (ii) the mechanism underlying distant-acting ("*trans*") pQTLs in plasma may be altered availability of binding proteins, rather than true *trans*-acting regulatory elements.

b) WFIKKN2, GDF11 and GDF8

We identified a common allele (rs11079936:C) near *WFIKKN2* that is associated in *cis* with lower levels of plasma WFIKKN2 (p=6.9x10^{-136}) and also in *trans* with lower binding of the SOMAmer targeting GDF11/8 (p=7.9x10^{-12}, Supplementary Table 4, Extended Data Figure 4). Due to close homology of GDF8 (myostatin) and GDF11, the SOMAmer (like many binding assays[13-14]) is unable to distinguish between them.

Because insufficient myostatin has been shown to result in excessive muscle growth[15], there is interest in better understanding the regulation of myostatin since therapeutic inhibition might be used to treat conditions characterised by muscle weakness, such as muscular dystrophy[16].

The *trans* signal for GDF11/8 was completely abrogated when we conditioned on WFIKKN2 levels (p=0.7), while in contrast the *cis* association remained significant after adjustment for GDF11/8 levels (p=7.2x10^{-113}), suggesting that WFIKKN2 may regulate GDF11 and/or GDF8. This observation is supported by *in vitro* evidence demonstrating that WFIKKN2 has high affinity for both GDF8 and GDF11[17] and that it inhibits the biological activity of GDF8[18].

To investigate the *trans* pQTL for GDF11/8 further, we used an alternative assay (proximity extension assay; Olink Bioscience, Uppsala, Sweden[19]) that is specific for GDF8 and not GDF11. We found that rs11079936 was not associated with GDF8 measuring using the Olink platform (p>0.05). We were unable to assess association with GDF11 since Olink do not provide an assay for this.

To better understand the measurements made by the GDF11/8 SOMAmer, we performed pull-down enrichment experiments from human plasma using the GDF11/8 SOMAmer followed by tandem mass spectrometry analysis to identify the SOMAmer-bound proteins as described in Supplementary Note Section 3. We identified unique peptides corresponding to GDF-8, GDF-11 and the known GDF-11/8 binding protein WFIKKN2 associated with the GDF11/8 SOMAmer. Using protein and peptide identification probability thresholds of 85%, we identified four spectra corresponding to four unique peptides for GDF-8 (including one peptide in the pro form of GDF-8), one spectrum corresponding to a unique peptide for GDF-11, and one spectrum corresponding to a peptide that is common to both GDF-11 and GDF-8 (Supplementary Note Table 3). We also identified two spectra corresponding to two unique peptides in WFIKKN2. In pull-down experiments performed with control sequences, which included irrelevant sequence 4666-212_3, or SOMAmers selected against other proteins (3352-80_3/carbonic anhydrase 6 or 3343-1_4/aminoacylase-1), we did not identify any GDF-11,

GDF-8 or WFIKKN2 peptides. The active forms of GDF-11 and GDF-8 share 90% amino acid identity and greater than 95% amino acid similarity, and both bind the GDF-11/8 SOMAmer with the same affinity. These results confirm that the SOMAmer captures both of these proteins in the plasma pull-down sample.

Supplementary Note Table 3. Summary of peptides observed in GDF-11/8 SOMAmer pulldown from pooled human plasma.

Protein	**Peptide Sequence**	**Peptide Identification Probability**	**Observed m/z**
GDF-8	(K)QPESNLGIEIK(A)	100%	614.33
GDF-8	(K)MSPINMLYFNGK(E)	100%	707.85
GDF-8	(K)YPHTHLVHQANPR(G)	98%	523.94
GDF-8	(K)IPAMVV(C)	94%	458.75
GDF-11	(K)QQIIYGK(I)	90%	425.25
GDF-11/8	(R)GSAGPCCTPTK(M)	100%	568.25
WFIKKN2	(R)CYMDAEACSK(G)	100%	617.73
WFIKKN2	(R)VSELTEEPDSGR(A)	100%	659.81

In our analysis rs11079936:C leads to lower levels of WFIKKN2 SOMAmer binding and lower levels of GDF11/8 SOMAmer binding. This would be consistent with the C allele leading to lower WFIKKN2 and thus to lower levels of the WFIKKN2:GDF8 complex (but without any effect on total GDF8 *per se*, explaining the lack of association using the Olink assay). This hypothesis is supported by the identification of WFIKNN2 in the pull-downs of the GDF11/8 SOMAmer described above. However, we are unable to formally exclude the possibility that the *trans* genetic signal is acting through altered abundance of the homologous protein GDF11.

These two examples highlight that proteins in plasma do not exist in isolation. Biological activity is regulated through multiple mechanisms including binding in complexes and cleavage of inactive precursors, and cannot always be interpreted as simply proportional to total abundance.

7) *Cis* pQTL for plasma proteinase-3 levels at an ANCA-associated vasculitis (AAV) risk locus

AAV background: clinical and serological phenotypes

AAV is an umbrella term for a group of immune-mediated diseases characterised by vasculitis of small- and medium-sized blood vessels. AAV comprises two principal syndromes, Granulomatosis with Polyangiitis (GPA, formerly known as Wegener's granulomatosis) and microscopic polyangiitis (MPA). Both GPA and MPA can result in organ- or life-threatening disease, and pauci-immune necrotising glomerulonephritis is a hallmark of both conditions. GPA is distinguished from MPA by the presence of granulomatous inflammation of the respiratory tract and the ear, nose and throat. Anti-neutrophil cytoplasmic antibodies (ANCA) are detectable in the serum of over 90% of patients with GPA and MPA. The targets of ANCA are the neutrophil proteases, proteinase-3 (PR3) and myeloperoxidase (MPO). Approximately two thirds of patients with a clinico-pathological diagnosis of GPA have ANCA specific to PR3, whilst ANCA to MPO occur in the minority. In MPA the converse is true, with ANCA against MPO in ~60%.

Previously reported associations of the *PRTN3* and *SERPINA1* loci in GWAS of AAV

The first genome-wide association study of AAV was performed by the European Vasculitis Genetics Consortium (EVGC)[20]. Genotyping was performed using the Affymetrix SNP 6.0 genotyping array, and association testing was performed at 612,676 variants passing quality

control. This study revealed distinct genetic associations in patient subgroups defined by ANCA specificity. These genetic signals mirrored ANCA specificity more closely than they did the clinical labels of GPA or MPA. PR3 antibody positive vasculitis (hereafter referred to as PR3 positive vasculitis) showed genome-wide associations with *HLA-DP* and *SERPINA1* whereas MPO antibody positive vasculitis was associated with *HLA-DQ*. Haplotype analysis of the *SERPINA1* locus indicated that the association with PR3 positive vasculitis was conferred by either rs28929474:T (the "Z allele") or a variant in high LD with it, rather than the tag SNP rs7151526.

A previous candidate gene study had suggested an association of GPA with the promoter region of *PRTN3* (which encodes PR3)[21]. The Affymetrix SNP 6.0 genotyping array used in the EVGC GWAS does not include probes for SNPs in the *PRTN3* locus; therefore, the EVGC study[20] used a TaqMan assay to genotype rs62132295, a variant in the promoter region of *PRTN3*. rs62132295 showed an association with PR3 positive vasculitis (p=2.6x10^{-7}), but this signal was diminished when all AAV cases (i.e. both PR3 positive and MPO positive cases) were analysed, indicating an association specific to PR3 positive vasculitis. The EVGC study was unable to assess associations of other variants in the *PRTN3* locus due to the lack of coverage outlined above.

A subsequent GWAS by the Vasculitis Clinical Research Consortium (VCRC)[22] used the Affymetrix Axiom Biobank array, and replicated the associations reported by the EVGC. In particular, this study confirmed the association with the *SERPINA1* Z allele, and replicated at genome-wide significance the association at the *PRTN3* locus (Supplementary Note Table 4). As in the European study, the associations at *SERPINA1* and *PRTN3* were stronger in the PR3

positive vasculitis subgroup than in the combined AAV cohort, indicating the specificity of these genetic associations to the PR3 positive subgroup.

The tag SNP reported by the VCRC at the *PRTN3* locus was rs62132293, in high LD with rs62132295 (r^2=0.94 in 1000 Genomes phase 3 European-ancestry individuals). Of note, in the VCRC study, an imputed variant (rs138303849) had a more significant association with AAV as a whole in their discovery cohort than did rs62132295, although the summary statistics for the association of this variant with the PR3 positive subgroup were not reported.

Associations of previously reported disease-associated variants with plasma PR3 abundance

The associations of the AAV-associated variants in the *PRTN3* region with plasma PR3 levels in our data are reported in Supplementary Note Table 4. We found that the disease risk allele increases plasma PR3 (for both the tag SNP reported by the EVGC and for that reported by the VCRC). Of note, rs138303849, the SNP most strongly associated with AAV in the VCRC study, had a stronger association with plasma PR3 levels (both smaller *p*-value and larger estimated effect size) than either the directly genotyped tag SNP reported in the VCRC study (rs62132293) or the SNP reported in the EVGC study (rs62132295).

In our pQTL study, the variant most strongly associated with plasma PR3 (in both the samples assayed using two distinct SOMAmers and in the largely non-overlapping group of individuals assayed using Olink) was rs10425544, which lies in the promoter region of *PRTN3* (Figure 4b). In addition, we identified conditionally independent PR3 pQTL signals at rs351111, rs7254911 and rs6510982. The linkage disequilibrium between these pQTL variants and the reported disease-associated SNPs is shown in Supplementary Note Table 5. Our lead pQTL variant rs10425544 has D' of 1.0 with rs138303849 (the variant most strongly associated with AAV

in the VCRC study) and D' of 0.83 with rs62132293 (the most strongly associated of the directly genotyped variants in the VCRC study), although the correlation assessed by r^2 was weaker (0.139 and 0.128, respectively). However, the conditionally independent pQTL variant, rs7254911, is in high LD with the reported disease-associated variants by both metrics (D'=0.995, r^2=0.97 to rs138303849, see Supplementary Note Table 5 for LD with the other reported vasculitis tag SNPs). Our data suggest that future fine-mapping and functional studies in PR3 positive vasculitis should prioritise examination of both rs10425544 and rs7254911.

Supplementary Note Table 4. Associations of AAV-associated variants with plasma PR3 levels.

AAV GWASs						pQTL data from this study			
Study (first author, year, PMID, reference)	**rsID**	**Typed or imputed?**	***p*-value PR3+ AAV**	***p*-value all AAV**	**Risk allele**	***p*-value pQTL (SomaLogic) PRTN3.13720.95.3**	***p*-value pQTL (SomaLogic) PRTN3.3514.49.2**	***p*-value pQTL (Olink)**	**Effect of risk allele on plasma PR3**
Lyons 2012 22808956 [20]	rs62132295	typed	2.6×10^{-7}	6.6×10^{-4}	A	2.0×10^{-13}	4.0×10^{-8}	2.4×10^{-40}	↑
Merkel 2017 28029757 [22]	rs62132293	typed	3.6×10^{-13} combined 7.9×10^{-10} discovery	8.6×10^{-11} combined 5.5×10^{-8} discovery	G	3.5×10^{-15}	3.1×10^{-9}	2.0×10^{-41}	↑
Merkel 2017 28029757 [22]	rs138303849	imputed	not reported	2.8×10^{-10} discovery not assessed in replication	not reported	6.5×10^{-27}	1.9×10^{-17}	7.8×10^{-53}	Unable to assess

Supplementary Note Table 5. LD between vasculitis-associated variants (from Lyons *et al*[20] and Merkel *et al*[22]) and PR3 pQTL variants. LD calculations based on 1000 Genomes phase 3 European-ancestry individuals.

	rs10425544 (lead pQTL)	**rs138303849** (Merkel *et al*, strongest association with AAV)	**rs7254911** (conditionally independent pQTL)	**rs62132293** (Merkel *et al* most strongly associated directly genotyped variant)	**rs62132295** (Lyons *et al*)	**rs351111** (conditionally independent pQTL)	**rs6510982** (conditionally independent pQTL)
rs138303849 (Merkel *et al*, strongest association with AAV)	D' 1.000 r^2 0.139						
rs7254911 (conditionally independent pQTL)	D' 1.000 r^2 0.142	D' 0.995 r^2 0.970					
rs62132293 (Merkel *et al* most strongly associated directly genotyped variant)	D' 0.834 r^2 0.128	D' 0.995 r^2 0.747	D' 1.000 r^2 0.771				
rs62132295 (Lyons *et al*)	D' 0.819 r^2 0.120	D' 0.967 r^2 0.727	D' 0.947 r^2 0.710	D' 0.982 r^2 0.938			
rs351111 (conditionally independent pQTL)	D' 0.763 r^2 0.146	D' 0.945 r^2 0.221	D' 0.946 r^2 0.226	D' 0.947 r^2 0.294	D' 0.962 r^2 0.295		
rs6510982 (conditionally independent pQTL)	D' 0.752 r^2 0.428	D' 0.833 r^2 0.073	D' 0.836 r^2 0.075	D' 0.836 r^2 0.097	D' 0.847 r^2 0.097	D' 0.955 r^2 0.173	

Supplementary References

1. Menni, C. *et al.* Circulating proteomic signatures of chronological age. *J Gerontol A Biol Sci Med Sci* **70**, 809-16 (2014).

2. Ngo, D. *et al.* Aptamer-based proteomic profiling reveals novel candidate biomarkers and pathways in cardiovascular disease. *Circulation* **134**, 270-285 (2016).

3. Jovanović, D., Krstivojević, P., Obradović, I., Durdević, V. & Dukanović, L. Serum cystatin C and beta2-microglobulin as markers of glomerular filtration rate. *Ren. Fail.* **25**, 123–33 (2003).

4. Ju, T. & Cummings, R. D. A unique molecular chaperone Cosmc required for activity of the mammalian core 1 beta 3-galactosyltransferase. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16613–8 (2002).

5. Melzer, D. *et al.* A genome-wide association study identifies protein quantitative trait loci (pQTLs). *PLoS Genet* **4**, e1000072 (2008).

6. Ferreira, R. C. et al. Functional IL6R 358Ala allele impairs classical IL-6 receptor signaling and influences risk of diverse inflammatory diseases. *PLoS Genet* **9**, e1003444 (2013).

7. Müllberg, J. et al. A metalloprotease inhibitor blocks shedding of the IL-6 receptor and the p60 TNF receptor. *J. Immunol.* **155**, 5198–205 (1995).

8. Garbers, C. et al. The interleukin-6 receptor Asp358Ala single nucleotide polymorphism rs2228145 confers increased proteolytic conversion rates by ADAM proteases. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1842**, 1485–1494 (2014).

9. Gold, L. *et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* **5**, e15004 (2010).

10. Gelinas, A. D., Davies, D. R. & Janjic, N. Embracing proteins: structural themes in aptamer–protein complexes. *Curr. Opin. Struct. Biol.* **36**, 122–132 (2016).

11. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169 (2017).

12. Abdgawad, M., Hellmark, T., Gunnarsson, L., Westman, K. W. A. & Segelmark, M. Increased neutrophil membrane expression and plasma level of proteinase 3 in systemic vasculitis are not a consequence of the -564 A/G promotor polymorphism. *Clin. Exp. Immunol.* **145**, 63–70 (2006).

13. Egerman, M. A. et al. GDF11 increases with age and inhibits skeletal muscle regeneration. *Cell Metab.* **22**, 164–174 (2015).

14. Smith, S. C. et al. GDF11 does not rescue aging-related pathological hypertrophy. *Circ. Res.* **117**, 926–932 (2015).

15. Schuelke, M. et al. Myostatin mutation associated with gross muscle hypertrophy in a child. *N. Engl. J. Med*. **350**, 2682–2688 (2004).

16. Malik, V., Rodino-Klapac, L. R. & Mendell, J. R. Emerging drugs for Duchenne muscular dystrophy. *Expert Opin. Emerg. Drugs* **17**, 261–277 (2012).

17. Kondás, K., Szláma, G., Trexler, M. & Patthy, L. Both WFIKKN1 and WFIKKN2 have high affinity for growth and differentiation factors 8 and 11. *J. Biol. Chem*. **283**, 23677–84 (2008).

18. Hill, J. J., Qiu, Y., Hewick, R. M. & Wolfman, N. M. Regulation of myostatin in vivo by growth and differentiation factor-associated serum protein-1: a novel protein with protease inhibitor and follistatin domains. *Mol. Endocrinol*. **17**, 1144–1154 (2003).

19. 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).

20. Lyons, P. A. et al. Genetically distinct subsets within ANCA-associated vasculitis. *N. Engl. J. Med*. **367**, 214–223 (2012).

21. Gencik, M., Meller, S., Borgmann, S. & Fricke, H. Proteinase 3 gene polymorphisms and Wegener's granulomatosis. *Kidney Int*. **58**, 2473–2477 (2000).

22. Merkel, P. A. *et al.* Identification of functional and expression polymorphisms associated with risk for anti-neutrophil cytoplasmic autoantibody-associated vasculitis. *Arthritis Rheumatol*. **69**, 1054-1066 (2016).