

Mapping the proteo-genomic convergence of human diseases

Pietzner, Maik; Wheeler, Eleanor; Carrasco-Zanini, Julia; Cortes, Adrian; Koprulu, Mine; Worheide, Maria; Oerton, Erin; Cook, James; Stewart, Isobel; Kerrison, Nicola; Luan, Jian'an; Raffler, Johannes; Arnold, Matthias; Arlt, Wiebke; O'Rahilly, Stephen; Kastenmüller, Gabi; Gamazon, Eric R; Hingorani, Aroon D; Scott, Robert; Wareham, Nicholas J

DOI:

[10.1126/science.abj1541](https://doi.org/10.1126/science.abj1541)

License:

Other (please specify with Rights Statement)

Document Version

Peer reviewed version

Citation for published version (Harvard):

Pietzner, M, Wheeler, E, Carrasco-Zanini, J, Cortes, A, Koprulu, M, Worheide, M, Oerton, E, Cook, J, Stewart, I, Kerrison, N, Luan, J, Raffler, J, Arnold, M, Arlt, W, O'Rahilly, S, Kastenmüller, G, Gamazon, ER, Hingorani, AD, Scott, R, Wareham, NJ & Langenberg, C 2021, 'Mapping the proteo-genomic convergence of human diseases', *Science*, vol. 374, no. 6569, eabj1541. <https://doi.org/10.1126/science.abj1541>

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:

This is the author's version of the work. It is posted here by permission of the AAAS for personal use, not for redistribution. The definitive version was published in *Science* on 14/10/2021, DOI: 10.1126/science.abj1541

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

1 **Title: Mapping the proteo-genomic convergence of human diseases**

2
3 **Authors:** Maik Pietzner^{1,2*}, Eleanor Wheeler^{1*}, Julia Carrasco-Zanini¹, Adrian Cortes³, Mine Koprulu¹,
4 Maria A. Wörheide⁴, Erin Oerton¹, James Cook¹, Isobel D. Stewart¹, Nicola D. Kerrison¹, Jian'an Luan¹,
5 Johannes Raffler^{4,5}, Matthias Arnold^{4,6}, Wiebke Arlt⁷, Stephen O'Rahilly⁸, Gabi Kastenmüller^{4,9}, Eric R.
6 Gamazon^{10,11}, Aroon D. Hingorani^{12,13,14}, Robert A. Scott³, Nicholas J. Wareham^{1,13}, Claudia
7 Langenberg^{1,2,13#}

8
9 **Affiliations:**

- 10 1. MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge School of Clinical
11 Medicine, Cambridge CB2 0QQ, UK
12 2. Computational Medicine, Berlin Institute of Health (BIH) at Charité – Universitätsmedizin Berlin,
13 10117 Berlin, Germany
14 3. GlaxoSmithKline, Stevenage SG1 2NY, UK
15 4. Institute of Computational Biology, Helmholtz Zentrum München, German Research Center for
16 Environmental Health, 85764 Neuherberg, Germany
17 5. Institut für Digitale Medizin, Universitätsklinikum Augsburg, 86156 Augsburg, Germany
18 6. Department of Psychiatry and Behavioural Sciences, Duke University, Durham, NC 27710, USA
19 7. Institute of Metabolism and Systems Research, University of Birmingham, Birmingham B15 2TT,
20 United Kingdom
21 8. MRC Metabolic Diseases Unit, Wellcome Trust-Medical Research Council Institute of Metabolic
22 Science, University of Cambridge, Cambridge CB2 0QQ, UK
23 9. German Centre for Diabetes Research (DZD), 85764 Neuherberg, Germany
24 10. Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN 37203, USA
25 11. Clare Hall, University of Cambridge, Cambridge CB3 9AL, United Kingdom
26 12. UCL British Heart Foundation Research Accelerator, Institute of Cardiovascular Science, University
27 College London, WC1E 6BT, UK.
28 13. Health Data Research UK, Gibbs Building, 215 Euston Road, London NW1 2BE, UK
29 14. Institute of Health Informatics, University College London, 222 Euston Road, London NW1 2DA, UK

30
31 *these authors contributed equally

32
33 #Correspondence to Dr. Claudia Langenberg (claudia.langenberg@mrc-epid.cam.ac.uk)

36 **ABSTRACT:**

37 Characterization of the genetic regulation of proteins is essential for understanding disease etiology
38 and developing therapies. We identified 10,674 genetic associations for 3,892 plasma proteins to
39 create a cis-anchored gene-protein-disease map of 1,859 connections that highlights strong cross-
40 disease biological convergence. This proteo-genomic map provides a framework to 1) connect
41 etiologically related diseases, 2) provide biological context for emerging disorders, and 3) integrate
42 different biological domains to establish mechanisms for known gene-disease link. Our results
43 establish the value of cis-protein variants for annotation of likely causal disease genes at GWAS loci,
44 addressing a major barrier for experimental validation and clinical translation of genetic discoveries,
45 and identify proteo-genomic connections within and between diseases.

46

47 **One Sentence summary:** A genetically anchored map of protein – disease links identifies shared
48 etiology across diverse diseases and possible therapeutic directions.

49

50

51

52 MAIN TEXT

53 Proteins are the central layer of information transfer from the genome to the phenome and
54 recent studies have started to elucidate how natural sequence variation in the human genome impacts
55 on protein concentrations measured from readily available biofluids such as blood (1–6). Investigation
56 of the clinical consequences of these so-called protein-quantitative trait loci (pQTLs) can help to better
57 understand disease mechanisms and provide insights into the shared genetic architecture across
58 diseases within a translational framework that puts humans as the model organisms at the center (2,
59 4). This is now pursued at scale by pharmaceutical companies for the discovery of drug targets or
60 repurposing opportunities (7, 8). Earlier studies have started to characterize the genetic architecture
61 of proteins using bespoke panels (3, 6, 9) or larger proteomic platforms (1, 2, 4, 5), and have
62 demonstrated how this can provide insight into the pathogenesis of specific diseases. There has been
63 less attention on: a) providing a framework to assess the protein specificity of genetic variation
64 residing outside (trans) the protein encoding gene, b) understanding the clinical relevance of pQTLs
65 for proteins detected in plasma but known to not be actively secreted (7), c) classifying thousands of
66 proteins based on their genetic architecture as explained by merely cis variants, specific trans variants,
67 or unspecific trans variants, d) demonstrating the specific utility of pQTLs for the prioritization of
68 candidate genes at established risk loci, and e) systematically mapping shared gene-protein-disease
69 signals to uncover connections among thousands of considered diseases and phenotypes.

70 Profiling thousands of proteins circulating in blood at population-scale is currently only possible using
71 large libraries of affinity reagents, namely antibodies or alternatively short oligonucleotides, called
72 aptamers, since gold standard methods such as mass spectrometry lack throughput. We have
73 previously provided a detailed comparison of 871 overlapping proteins measured in 485 individuals
74 (10) of the two most comprehensive platforms, the aptamer-based SomaScan v4 assay and the
75 antibody-based Olink proximity extension assay. We demonstrated that the majority of pQTLs are
76 consistent across platforms (64%), in line with smaller scale efforts (4), but highlighted the need to
77 triangulate pQTLs with gene expression and phenotypic information to derive tangible biological
78 hypotheses. Here we present a genome-proteome-wide association study targeting 4,775 distinct
79 proteins measured from plasma samples of 10,708 generally healthy European-descent individuals
80 who were participants in the Fenland study (Table S1) (11). We identified 10,674 variant – protein
81 associations and developed a framework to systematically identify protein- and pathway-specific
82 pQTLs augmenting current ontology-based classifications in a data-driven manner. We show that half
83 of all pQTLs close to the protein-encoding gene, cis-pQTLs, colocalize with gene expression or splicing
84 QTLs in various tissues allowing to derive functional insights within tissues by integrating genetics with

85 plasma proteomics. We demonstrate the specific ability of cis-pQTLs to prioritize candidate causal
86 genes at established genetic risk loci. By means of phenome-wide colocalization screens we generate
87 a proteo-genomic map of human health covering 1,859 gene-protein-trait triplets providing insights
88 into the shared etiology across diseases and the identification of pathophysiological pathways through
89 cross-domain integration.

90 RESULTS

91 Genetic associations for protein targets

92 We performed a genome-proteome-wide association analysis by testing 10.2 million
93 genotyped or imputed autosomal and X-chromosomal genetic variants with a minor allele frequency
94 (MAF) >1% among 10,708 participants in the Fenland study targeting 4,775 distinct proteins (12). We
95 identified 2,584 genomic regions (1,543 within ± 500 kb of the protein-encoding genes, cis) associated
96 with at least one of 3,892 protein targets at $p < 1.004 \times 10^{-11}$. 1,097 of these regions covered variants
97 that have not been reported to be associated with plasma proteins so far (1–6, 9) ($r^2 < 0.1$), of which
98 64% (867 out of 1,356 pQTLs) available in (4) replicated ($p < 0.05$, directionally consistent). Further, 61%
99 (488 out of 797, Table S2) of pQTLs replicated using the complementary Olink technique (12), with a
100 higher proportion for variants in cis (81.2%) compared to trans (44.2%). Most regions (79.3%, $n = 2,050$)
101 were associated with a single protein target, but we observed substantial pleiotropy (≥ 2 protein
102 targets) at the remaining regions, including up to five (16.1%, $n = 418$), 6–20 (3.4%, $n = 88$), or 21–50
103 (0.7%, $n = 19$) associated protein targets, and with eight regions (*CFH*, *ARF4-ARHGEF3*, *C4A-CFB*, *BCHE*,
104 *VTN*, *CFD*, *ABO*, *GCKR*) associated with 59 to 1,539 protein targets (Fig. 1). The 194 pleiotropic regions
105 harboring a cis-pQTL identified master regulators of the plasma proteome, including
106 glycosyltransferases such as the histo-blood group ABO system transferase (*ABO*), key metabolic
107 enzymes like glucokinase regulatory protein (*GCKR*), or lipid mediators such as apolipoprotein E,
108 establishing a network-like structure of the circulating proteome (1).

109 Out of the 3,892 protein targets, 26.8% ($n = 1,046$) had pQTLs in cis and trans, 13.4% ($n = 523$) in cis only,
110 and 59.6% ($n = 2,323$) in trans only, among a total of 8,328 sentinel variant-protein target associations
111 (Fig. 1 and Tables S2 and S3). We identified another 2,346 secondary pQTLs at those loci using an
112 adapted stepwise conditional analysis (median: 1, range: 1 - 13) indicating widespread allelic
113 heterogeneity in cis (68.8%) and trans (31.2%). The majority of the 5,442 distinct variants were located
114 in introns (~44%) or were in high LD ($r^2 > 0.6$) with a missense variant (~21%), with similar distributions
115 across cis- and trans-pQTLs (Fig. S1). We observed 663 cis-pQTLs with direct consequences for the
116 structure of the protein target (protein-altering variants, PAVs), including important substructures,
117 such as disulfide bonds (4.2%), α -helices (3.1%), and β -strands (2.6%) (Fig. S1). Such variants are

118 predicted to affect correct folding of protein targets, including diminished secretion or reduced half-
119 life in the bloodstream, rather than expression of the protein-encoding gene (13). For example, we
120 observed an enrichment of PAVs among actively secreted proteins (14) (39.6% vs 33.7%, $p=0.04$, X^2 -
121 test) possibly indicating modulation of common posttranslational modifications, such as glycosylation.

122 An integrated classification system for pathway-specific pQTLs

123 We integrated a data-driven protein network with ontology mapping (GO terms, **Fig. 2A-B and**
124 **S2**) to distinguish pathway-specific pQTLs from those exerting effects on multiple unrelated targets
125 (12, 15). We successfully assigned 40.8% ($n=1,790$ in cis, $n=423$ in trans) of the 5,442 genetic variants
126 as protein-specific and 5.9% ($n=236$ in cis, $n=86$ in trans) as pathway-specific based on converging
127 evidence from the network and ontology mapping, and another 16.5% ($n=498$ in cis, $n=402$ in trans)
128 to be likely pathway-specific based on either source. In total, 1,802 protein targets had at least one
129 (likely) specific pQTL in cis ($n=1,385$) or trans ($n=417$). We classified 648 variants that would have been
130 missed by ontology mapping as protein community-specific through our data-driven network
131 approach. One example is rs738408 (*PNPLA3*), a non-alcoholic fatty liver disease variant (16) which
132 was associated with 22 out of 70 aptamers from the same protein community (**Fig. 2C**). *PNPLA3*
133 encodes patatin-like phospholipase domain-containing protein 3 (*PNPLA3*), and rs738408 tags the
134 missense variant rs738409 (I148M) rendering *PNPLA3* resistant to ubiquitylation-mediated
135 degradation and resulting in subsequent accumulation on hepatic lipid droplets causing fatty liver
136 disease (17). The associated protein targets included multiple metabolic and detoxification enzymes
137 highly expressed in the liver, such as alcohol dehydrogenases, arginosuccinate lyase, bile salt
138 sulfotransferase, or aminoacylase-1. Our results support the hypothesis that those might only appear
139 in plasma of otherwise healthy individuals as a result of lipid overload-induced lysis of hepatocytes.
140 The putative liver damage-specific effect, anchored on the *PNPLA3* trans-pQTL, makes those protein
141 targets potential biomarker candidates compared to tissue unspecific proteins currently used to
142 identify fatty liver disease or liver injury in the clinic (18).

143 Contribution of cis and trans genetic architecture

144 We observed three major categories of protein targets based on the contribution of genetic
145 variation to plasma concentrations (**Fig. 2D and S3, and Table S3**). For about a third ($n=1,249$) of the
146 protein targets, genetic variance was mainly explained by one or more cis-pQTLs, whereas for 7.2%
147 ($n=282$) protein- or pathway-specific trans-pQTLs accounted for most of the genetic variation, leaving
148 two-thirds ($n=2,361$) mainly explained by unspecific trans-pQTLs (12). Overall, we observed a median
149 genetic contribution of 2.7% (IQR: 1.0% - 7.6%) reaching values above 70% for proteins like vitronectin
150 (rs704, MAF=47.3%) or sialic acid-binding Ig-like lectin 9 (rs2075803, MAF=44.1%) which were often

151 driven by only a single common cis-pQTL. PAVs, affecting the binding epitope of the protein target,
152 are the likely explanation for such strong and isolated genetic effects. While more than two-thirds of
153 the protein targets with at least one cis-pQTL were unrelated to PAVs, we provide evidence that 158
154 (32.9%) of the protein targets linked to a PAV ($r^2 > 0.6$) shared a genetic signal with at least one disease
155 or risk factor (see below). This suggests that conformation and hence function of the protein target,
156 rather than plasma abundance of the protein target, might be more relevant as mediators of
157 downstream phenotypic consequences and that aptamers are able to detect such probably
158 dysfunctional proteins.

159 Our approach to identify protein-/pathway-specific trans-pQTLs allowed us to uncover biological
160 relevant information, which was otherwise hidden by strong and unspecific trans-pQTLs that possibly
161 interfere with the measurement technique rather than the biology of the protein target. For example,
162 rs704, a missense variant within *VTN* associated with a higher fraction of single chain vitronectin with
163 altered binding properties (19, 20), explained 72% of the variance in MICOS complex subunit MIC10
164 (MOS1), far outperforming the contribution of the specific trans-pQTL rs398041972 (0.7%).
165 Rs398041972 resides about 1 Mb upstream of *TMEM11*, encoding transmembrane protein 11, a
166 physical interaction partner of MOS1 as part of the MICOS complex (21). In general, we observed that
167 the median contribution of specific trans-pQTLs to the variance in plasma concentrations was 1.1%
168 (IQR: 0.6%-2.6%) across 687 protein targets, reaching values as high as 38.3% for catenin β -1 via two
169 trans-pQTLs (rs1392446 and rs35024584) within the same region for which we prioritized *CDH6* as a
170 candidate causal gene. *CDH6* encodes cadherin 6, which physically interacts with catenin β -1 (22). We
171 systematically tested for an enrichment of putative protein interaction partners among the 20 closest
172 genes at each specific trans locus and observed a 1.53-fold enrichment (Chi-square test p-
173 value = 1.8×10^{-10}) of first- and second-degree neighbors from the STRING network (23), highlighting the
174 ability of our classification system to identify biologically meaningful trans-pQTLs.

175 **Shared genetic architecture with gene expression and splicing**

176 We integrated plasma pQTL results with both gene expression and splicing QTL data (eQTL
177 and sQTL) from the GTEx version 8 release (24) using statistical colocalization (posterior probability
178 (PP) > 80%) for all 1,584 protein targets with at least one cis-pQTL (12). There was strong evidence
179 that half (50.1%) of these had a shared signal with gene expression in at least one and a median of 4.5
180 tissues (IQR: 2-12; **Fig. 3A**), vastly expanding previous knowledge of gene expression contribution
181 across tissues (4, 9). The majority of cis-pQTLs (n=584, 73.4%) showed plasma protein and gene
182 expression effects in the same direction in all tissues (**Fig. 3A**), but 26.6% (n=212) showed evidence of
183 at least one pair with opposite effects, including 108 where the protein effect was opposite of the

184 direction observed for gene expression across all tissues with evidence for colocalization. For example,
185 the A-allele of the lead cis-pQTL rs2295621 for immunoglobulin superfamily member 8 (*IGSF8*) was
186 inversely associated with plasma abundance of the protein target ($\beta=-0.19$, $p<1.65\times 10^{-32}$) but
187 positively associated with expression of the corresponding mRNA across 33 tissues (**Table S4**).
188 Uncoupling of gene and protein expression, even within the same cell, is a frequently described
189 phenomenon, and possible mechanisms include differential translation, protein degradation,
190 contextual confounders, such as time and developmental state, or protein-level buffering (25). For
191 145 protein targets, we identified strong evidence of a tissue-specific contribution to plasma
192 abundances based on a single tissue strongly outweighing all others (**Fig. 3A and Table S4**). These
193 included known tissue-specific examples such as protein C in liver tissue, but also less obvious ones,
194 such as hepatitis A virus cellular receptor 1 (or TIM-1), an entry receptor for multiple human viruses,
195 for which the cis-pQTL and cis-eQTL specifically colocalized in tissue from the transverse colon. **To**
196 **maximize power for the most closely aligned tissue compartment, whole blood, we integrated gene**
197 **expression data from the eQTLGen consortium (26), which confirmed 140 cis-eQTL/pQTL pairs and**
198 **revealed another 38 cis-eQTL/pQTL pairs not seen in the GTEx resource, including immune cell-specific**
199 **mediators of the inflammatory response such as leukocyte immunoglobulin-like receptor subfamily A**
200 **member 3 (Table S4).**

201 To obtain insights beyond the average readout across all transcript species, we examined alternative
202 splicing as a source of protein target variation (12). One-fifth (20.1%) of cis-signals were shared with a
203 cis-sQTL in at least one tissue (median: 6 tissues, IQR: 2-15) (**Fig. 3B**), and 84 of these were not seen
204 with eQTL data, suggesting that the pQTL-relevant transcript isoform was masked from the bulk of
205 assayed transcripts. In contrast to the eQTL colocalization, we did not observe an overall pattern of
206 aligning effect directions (**Fig. 3B**). This might be best explained by the intron-usage quantification of
207 splicing events within GTEx version 8, which does not allow straightforward mapping of the eventually
208 transcribed isoforms, and the expression of an alternative protein isoform with less affinity to the
209 SOMAmer reagent. The latter may have accounted for the 90 protein target examples where the
210 colocalizing cis-sQTL explained more than 10% of the variance in plasma concentrations (**Table S4**) and
211 emphasizes the ability of splicing QTLs to determine the underlying sources of variation in plasma
212 abundances of protein targets. In summary, our results demonstrate that proteins measured in plasma
213 can be used as proxies for tissue processes when anchored on a shared genetic variation with tissue-
214 specific gene expression or alternative splicing data.

215 **cis-pQTLs enable identification of candidate causal genes at GWAS loci**

216 We used the inherent biological specificity of cis-pQTLs to systematically identify candidate
217 causal genes for genome-wide significant variants reported in the GWAS catalog ($p<5\times 10^{-8}$; download:

218 25/01/2021) by assessing 558 cis-regions for which the pQTL was in strong LD ($r^2>0.8$) with at least
219 one variant for 537 collated traits and diseases (**Fig. 4** and **Table S5**) (12). For a quarter of these
220 (24.6%), we annotated a gene different from the reported or mapped gene, and for another 79 cis-
221 regions (14.2%), our predicted causal gene was reported as part of a longer list of potential causal
222 genes.

223 Among the genes we identified are candidates with strong biological plausibility, such as *AGRP*,
224 encoding Agouti-related protein, a neuropeptide involved in appetite regulation (27), suggesting a
225 possible mechanism for measures of body fat distribution associated at this locus. Another example
226 was *NSF*, encoding N-ethylmaleimide-sensitive factor (NSF), which may be involved in the fusion of
227 vesicles with membranes, enabling the release of neurotransmitters into the extracellular space (28);
228 a locus that was identified for Parkinson's disease (**Table S5**).

229 We further assigned *PRSS8* as a candidate causal gene at the *KAT8* locus for Alzheimer's disease (AD),
230 supported by strong LD ($r^2=0.96$) and a high posterior probability of a shared genetic signal (98%)
231 between the lead cis-pQTL (rs368991827, MAF=27.8%) and the common *KAT8* intronic variant
232 (rs59735493) that has been reported for AD (**Fig. S4**). *PRSS8* codes for prostaticin, and we estimated a
233 13% reduction in AD risk (odds ratio: 0.87; 95%-CI: 0.82-0.91, $p=3.8\times 10^{-8}$) for each 1 s.d. higher
234 normalized plasma abundance of prostaticin. The locus has been identified by multiple GWAS efforts
235 (29), yet prioritization strategies have failed to provide conclusive evidence (30). Prostaticin is a serine
236 protease highly expressed in epithelial tissue, which regulates sodium channels (31) and represses
237 TLR4-mediated inflammation in human and mouse models of inflammatory bowel disease (32), a
238 mechanism which might also be relevant to TLR4-mediated neuroinflammation in AD (33).

239 We observed multiple examples in which our cis-pQTL mapping identified biologically plausible
240 candidates that were not implicated by cis-eQTL mapping (**Fig. 4**). For example, we assigned *RSPO1* as
241 a candidate causal gene at the eQTL-supported *CDCA8* locus for endometrial cancer (34). The
242 intergenic variant rs113998067 is the lead signal for endometrial cancer and was a secondary cis-pQTL
243 for R-spondin-1, encoded by *RSPO1*. Statistical colocalization confirmed a highly likely shared signal
244 (PP=98.2%) (**Fig. S5**). Accordingly, we estimated a 91% increased risk for endometrial cancer per 1 s.d.
245 higher plasma abundance of R-spondin-1 (odds ratio: 1.91, 95%-CI: (1.52-2.41), $p\text{-value}=3.6\times 10^{-8}$). R-
246 spondin-1 is a secreted activator protein which acts as an agonist for the canonical Wnt signaling
247 pathway (35), playing a regulatory role as an adult stem cell growth factor. Work in mouse models
248 (36), however, suggests that R-spondin-1 upregulates estrogen receptor alpha independent of Wnt/ β -
249 catenin signaling and might therefore amplify estrogen-mediated endometrial cancer risk (36). We
250 note that the effect estimate for rs113998067 did not differ by sex ($p=0.12$), and knockout models in

251 male and female mice have shown abnormal development of testis and ovary, respectively (37, 38),
252 possibly indicating a wider impact on diseases of reproductive tissues.

253 **A map of proteo-genomic connections across the phenome**

254 We systematically assessed sharedness of gene-protein-disease triplets through phenome-
255 wide colocalization of cis-pQTL regions (12) to identify and create a genetically anchored map of
256 proteins involved in the etiology of common complex diseases, which could represent potential
257 druggable targets. We identified 1,859 gene-protein-trait triplets (network edges, **Fig. 5 and S6**)
258 comprising 412 protein targets and 506 curated traits (**Fig. S7 and Table S6**). The mapping of these
259 shared gene-protein-phenome connections highlights a large number of insights, as discussed below,
260 while confirming previously established connections for known pleiotropic loci (for example GCKR
261 (n=197 traits), alpha-1-antitrypsin (n=79 traits), or apolipoprotein A-V (n=64 traits)) and established
262 disease genes (for example proto-oncogene tyrosine protein kinase receptor RET (*RET*) and
263 Hirschsprung's disease (39) or C-C motif chemokine 21 (*CCL21*) and rheumatoid arthritis (40)).

264 The map highlights ten diseases for which we identified five or more colocalizing cis-pQTLs, including
265 coronary artery disease (n=12), hyperlipidemia indicated by lipid-lowering medication (n=8),
266 ulcerative colitis (n=7), Alzheimer's disease (n=6), and type 2 diabetes (n=5). Statistical power was
267 greatest for the detection of shared genetic architecture for traits for which measures were available
268 in the largest number of people, in line with a median of 2 colocalizing cis-pQTLs (IQR: 2 - 4, maximum
269 32 for mean platelet volume) for blood cell parameters and biomarkers available in large-scale
270 biobanks. For 104 out of 191 curated phenotypes with at least 3 colocalizing protein targets, we
271 observed significant enrichment of pathways (false discovery rate (q-value) < 5%; **Table S7**). These
272 reflected known biology of the corresponding clinical entities, such as 'wound healing' for platelet
273 count, 'skeletal system development' for height, 'cholesterol metabolism' for coronary artery disease,
274 or 'response to virus' for Crohn's disease, as well as yet less understood ones such as 'toll-like receptor
275 signaling' for hypothyroidism, for which two of the genes (*IRF3* and *TLR3*) have already been shown to
276 confer virus-induced disease onset in mouse models (41).

277 The proteo-genomic map provides a new framework to 1) connect etiologically related diseases, 2)
278 provide biological context for new or emerging disorders, such as COVID-19, and 3) integrate
279 information from different biological domains to establish mechanisms for known gene-disease links.
280 For each of these scenarios, we provide selected examples below to highlight the scientific
281 opportunities arising from this map and the related open resource platform (www.omicscience.org).

282

283 **Potential candidate genes for COVID-19 outcomes**

284 We integrated GWAS summary statistics in our map for four different outcome definitions
285 related to COVID-19 (42), that differed substantially in the number of included cases (5,101 – 38,984),
286 and observed that results were sensitive to the choice of outcome. We replicated *ABO* and *OAS1* as
287 two candidate causal genes (43) (**Fig. S8**) with both showing consistent evidence across outcomes,
288 ranging from susceptibility to COVID-19 to severe cases requiring hospitalization. The lead cis-pQTL
289 for *BGAT* (rs576125, MAF=33.5%, within *ABO*) also colocalized with pulmonary embolism (**Fig. 5**), a
290 common complication of severe COVID-19 (44), which might be attributable to altered abundances of
291 proteins involved in the coagulation cascade (15). We further observed suggestive evidence for *NSF*
292 (for the risk of COVID-19 hospitalization) and *BCAT2* (for severe COVID-19) which each shared a genetic
293 signal with only one of these four outcomes, and therefore requiring external validation of their
294 possible role in COVID-19 or associated pathologies.

295 **Integrating multiple OMICs layers elucidates a disease mechanism for gallstones**

296 We identified a shared signal at *SULT2A1*, a known gallstone locus (45), between bile salt
297 sulfotransferase (*SULT2A1*) and risk of cholelithiasis (odds ratio per 1 s.d. higher normalized protein
298 abundances: 2.12, 95%-CI: 1.66 – 2.70, p-value= 2.1×10^{-37}) and cholecystectomy (odds ratio: 2.09, 95%-
299 CI: 1.86 – 2.34, p-value= 7.8×10^{-38}). We next used multi-trait colocalization (46) and further identified
300 that mRNA expression of *SULT2A1* in the liver, plasma concentrations of multiple sulfated steroids
301 (47), including sulfate conjugates of androgen and pregnenolone metabolites, and bile acids shared
302 the same signal with high posterior probability (PP=99%) largely explained (63%) by rs212100, a
303 variant in high LD ($r^2 = 0.90$) with the lead cis-pQTL at this locus (**Fig. 6A and Fig. S9**). The consistent
304 positive effect directions across all physiological entities, and in particular sulfated steroids and
305 primary bile acid metabolites, clearly favor higher *SULT2A1* activity as a mode of action. The
306 concurrent inverse association with lower plasma concentrations of the secondary bile acid
307 glycholithocholate indicates diminished formation of lithocholic acid, an essential detergent to
308 solubilize fats, including cholesterol (48). Our vertical integration of diverse biology entities indicates
309 a supersaturated bile that promotes cholesterol crystallization and gallstone formation as a causal
310 mechanism at a locus for which the mode of action has only been vaguely hypothesized (45).

311 **Convergence of soft tissue disorders through *FBLN3***

312 A protein target connected to a very large number (n=37) of diseases and phenotypes was
313 *FBLN3* (extracellular matrix glycoprotein encoded by *EFEMP1*), showing gene-protein convergence of
314 diverse connective tissue disorders as well as gene expression of *EFEMP1* in subcutaneous adipose
315 tissue, with high confidence in the lead cis-pQTL (rs3791679, MAF=33.6%) being the causal variant in

316 multi-trait colocalization (Fig. 6B and Fig. S10). The locus has previously been reported but not
317 connected across separate GWASs conducted for height (49), optic disc area (50), carpal tunnel
318 syndrome (51), inguinal hernia (52), and lung function (53). *Efemp1* knock-out mice display abnormal
319 elastic fiber morphology, develop different types of hernias, and have smaller body size and lower
320 body fat (54), in line with the human spectrum of clinical features. FBLN3 is part of the extracellular
321 matrix and widely expressed but its function is incompletely understood (55). We provide insights
322 about its role in the etiology of a large number of connective tissue disorders, including a potential
323 explanation for the established link between carpal tunnel syndrome and shorter stature (51).
324 Mutations in *EFEMP1* cause a rare eye disease called Doyme honeycomb retinal dystrophy (DHRD) (56),
325 characterized by visual disturbances and drusenoid deposits due to accumulating intracellular FBLN3.
326 We observed sharedness of the common signal at this protein locus with vision-related phenotypes,
327 including use of contact lenses (myopia) and decreased optic disc area, a risk factor for open-angle
328 glaucoma (50), with lower protein concentrations associated with greater risk, as also observed in
329 patients with DHRD.

330 Differences of cis-pQTLs by sex and age

331 We systematically tested differences in the genetic associations of all protein targets included
332 in the proteogenomic map (N=412) by age or sex. We identified a total of 14 protein targets that
333 showed evidence for significant ($p < 5.9 \times 10^{-5}$) effect modification of the cis-pQTL by sex (N=10) or age
334 (N=8), including four common to both (Table S8). This included biological plausible candidates, such
335 as annexin II, where the cis-pQTL showed a stronger effect in women, albeit with a strong significant
336 effect in either sex (women: $\beta = -0.86$, $p\text{-value} < 1.7 \times 10^{-467}$; men: $\beta = -0.64$, $p\text{-value} < 2.5 \times 10^{-231}$). This
337 finding is in line with evidence of isoform expression of the protein-encoding gene *ANXA2* in male and
338 female reproductive tissues, including prostate (PP=81.9%) and vagina (PP=87.4%) and a possible role
339 of the locus in puberty timing (57, 58).

340 We noted that most of the identified cis-pQTLs showed age- and sex-differential and not dimorphic
341 effects (59) and were linked to missense variation (inhibin C, vitronectin, Siglec 9, GCKR, SOD3, CPA4,
342 and PILRA) or alternative splicing events (annexin II, BGAT, and CO8G) with very strong overall effects,
343 enabling the detection of even small effect differences between strata more easily (60). In general,
344 our results are concordant with the few sex-specific effects of molecular QTLs reported so far (61, 62)
345 and show that systematic efforts for both molecular QTLs and disease GWAS are needed to better
346 understand the mechanisms underlying such differences. Crucially, investigating the relevance of
347 these genetic differences for phenotypic expression depends on the availability of sex-specific GWAS
348 results across the human phenome.

349 **Druggable targets and repurposing opportunities**

350 We systematically identified druggable proteins in the proteo-genomic map by linking the
351 protein-encoding gene to the druggable genome (63) and identified 60 protein targets linked to at
352 least one phenotype, including 22 protein targets linked to a disease (**Table S9**). We replicate
353 established examples, such as the IL-6 receptor for rheumatoid arthritis or thrombin for deep venous
354 thrombosis (**Fig. 5**). We also identified 31 candidates with potential repurposing opportunities for 1 to
355 8 diseases (total of 32 different indications), following a search and prioritization strategy in Open
356 targets (64).

357 **Webserver**

358 To enable customized and in-depth exploration of high-priority protein targets, that is, those
359 with at least one cis-pQTL, we created an interactive online resource
360 (www.omiscience.org/apps/pgwas). The webserver provides intuitive representations of genetic
361 findings and enables the look-up of summary statistics for individual SNPs, genes, and whole genomic
362 regions across all protein targets. To interactively assess specificity and identify pleiotropic cis-pQTLs
363 that present strong trans-like association profiles, we generated an interactive heatmap of genetic
364 associations of all cis-pQTLs across all high-priority candidate proteins. We further provided detailed
365 annotations of the protein targets, including links to external databases, such as UniProt or Reactome,
366 information on currently available drugs, characterization of associated SNPs, as well as results from
367 our colocalization analysis with eQTLs, sQTLs, and disease phenotypes. An interactive version of the
368 proteo-genomic map allows a deep dive into proteins or phenotypes of particular interest to explore
369 cross-disease connections within subnetworks.

370 **DISCUSSION**

371 The promise of proteomic technologies and their integration with genomic data lies in their
372 application to rare and common human diseases. While previous studies started to exploit the
373 phenotypic consequences of pQTLs, they have mainly focused on identifying and describing the
374 genetic architecture of proteins measured by specific platforms (1–6, 9). We performed a systematic
375 integration of the phenome and created a proteo-genomic map of human health that identifies many
376 potential causal disease genes and highlights genetically driven connections across diverse human
377 conditions. The traditional classification of diseases relies on the aggregation of symptoms commonly
378 presenting together and, with the exception of Mendelian disorders, is rarely based on shared etiology
379 (65). Our network anchors the convergence of diseases in their shared genetic etiology, as shown for
380 FBLN3, providing mechanistic understanding and a starting point for the identification of treatment
381 strategies targeting underlying genetic causes.

382 Uncertainty in assigning causal genes and variants remains a major limitation for experimental
383 validation and clinical translation of results from the plethora of hypothesis-free genetic association
384 studies. We show how cis-pQTLs identify causal candidate genes at established disease risk loci,
385 including COVID-19, providing immediate hypotheses for experimental follow-up for a large number
386 of disease genes.

387 The uncertain specificity of genetic variation affecting protein content outside of the protein-encoding
388 region, trans-pQTLs, restricts the discovery of *de novo* biological insights in protein regulation and
389 instrumentation of such variants for genetic prediction, such as with polygenic scores. We show how
390 data-driven network clustering augments ontology-based classification approaches and identifies
391 biologically plausible examples, such as for *PNLPA3* and a community of liver-derived protein targets.

392 Genetic variation found for proteins circulating in blood raises the question of transferability to
393 disease-relevant tissue processes. We demonstrate that for about half of the protein targets with a
394 cis-pQTL, this can be linked to gene expression in various tissues and provide examples, such as for
395 *SULT2A1*, that illustrate how multi-domain integration can identify tissue-specific mechanism. In its
396 most simple form such cis-pQTLs determine the basal rate of protein production within cells and are
397 more or less constantly released into plasma due to natural cell turnover (66). Integration of genetic
398 information allowed us to separate out such enclosed effects from other mechanisms leading to
399 higher cell turn over or leakage, such as for *SULT2A1* and the liver-specific effect of the *PNPLA3* variant.

400 While this provides a tangible strategy to point to relevant tissues, overlapping data for tissue-specific
401 gene and protein expression is required to quantify the contribution of various tissues to the plasma
402 proteome.

403 To accelerate use and translational potential of our findings, we generated an open access interactive
404 web resource that enables the scientific community to easily and rapidly capitalize on these results
405 for future research across clinical specialties. We demonstrate for multiple examples how this
406 resource can be used to put gene-phenotype findings into a systems biological context.

407 While our study is distinguished by its comprehensive discovery and characterization of pQTLs in cis
408 and trans along with a systematic integration of the phenome, it does have limitations. Firstly, the
409 nature of the technology used to measure protein concentrations is designed to maximize discovery
410 by generating a large library of affinity reagents, which rely on a preserved shape of the target protein
411 and hence might miss genetic effects specific to a particular isoform of the protein (10). The semi-
412 quantitative nature of the assay makes risk estimates based on Mendelian randomization studies
413 challenging. A thorough discussion of assay differences can be found in our previous work, and we
414 observed consistent cis-pQTLs for examples highlighted, including *RSPO1*, *SULT2A1*, and *FBLN3*, as

415 measured with Olink. Secondly, our study cohort consisted of predominantly healthy middle-aged
416 participants of European-decent and replication of our results in ethnically diverse populations is
417 warranted, in particular for the discovery of drug targets. Further work would also be required to
418 investigate possible modifying effects of phenotypic characteristics on gene – protein associations,
419 such as by sex, age, or behavioral factors. Thirdly, our study concentrated on the common spectrum
420 of variation in the genome. Investigation of rare variation is likely to identify pQTLs with larger effect
421 sizes and possibly more severe phenotypic consequences. Finally, our proteo-genomic map is limited
422 to publicly available GWAS summary statistics and inclusion of further data for additional phenotypes,
423 in particular cancers, and understudied diseases, will provide additional insights.

424

425 MATERIALS and METHODS

426 Detailed materials and methods are provided in the supplementary materials (12). We performed a
427 genome-proteome-wide association study among 10,708 participants of European-decent in the
428 Fenland study (**Table S1**) on 10.2 million genetic variants and plasma abundances of 4,775 distinct
429 protein targets measured in plasma using established workflows (15). Protein targets were measured
430 using the SomaScan v4 assay employing 4,979 single-stranded oligonucleotides (aptamers) with
431 specific binding affinities to 4,775 unique protein targets (67, 68). We used the term ‘protein target’
432 to refer to proteins targeted by at least one aptamer. We define significant genetic variant – protein
433 target associations (pQTLs) at a stringent Bonferroni-threshold ($p < 1.004 \times 10^{-11}$) and performed
434 approximate conditional analysis to detect secondary signals for each genomic region identified by
435 distance-based clumping of association statistics. We defined cross-aptamer regions using a combined
436 approach of multi-trait colocalization (46) and LD-clumping. We classified pQTLs as protein- or
437 pathway-specific by assessing pQTL-specificity across the entire proteome ($p < 5 \times 10^{-8}$) while testing
438 whether associated protein targets were captured by a common GO term or a protein community in
439 a data-driven protein network. We computed the variance explained in plasma abundances of protein
440 targets by cis- (within ± 500 kb of the protein-encoding gene) or trans-pQTLs according to different
441 specificity categories using linear regression models. We used statistical colocalization (69) to test for
442 a shared genetic signal between expression or alternative splicing of the protein-encoding gene and
443 the cis-pQTL in one out of at least 49 tissues of the GTEx v8 project (24). We systematically cross-
444 referenced established genetic risk loci for common complex diseases and phenotypes with pQTLs by
445 identifying cis-pQTLs or strong proxies ($r^2 > 0.8$) in the GWAS catalog (<https://www.ebi.ac.uk/gwas/>).
446 We finally performed phenome-wide colocalization screens at 1,548 protein-encoding loci using
447 publicly available (70) as well as in-house curated genome-wide association statistics for thousands of
448 phenotypes. We applied stringent priors and conservative filters to derive high confidence protein –
449 phenotype links. We used basic functions of R (v.3.6.0), the R package *igraph*, and the BioRender web
450 application (<https://biorender.com/>) to create figures. The Fenland study was approved by the
451 National Health Service (NHS) Health Research Authority Research Ethics Committee (NRES
452 Committee – East of England Cambridge Central, ref. 04/Q0108/19), and all participants provided
453 written informed consent.

454

455 **SUPPLEMENTARY MATERIALS**

456 Materials and methods

457 **Tables S1-S9**

458 **Fig. S1-S10**

459 **References (66 - 80)**

460

461 REFERENCES AND NOTES

- 462 1. V. Emilsson, M. Ilkov, J. R. Lamb, N. Finkel, E. F. Gudmundsson, R. Pitts, H. Hoover, V.
463 Gudmundsdottir, S. R. Horman, T. Aspelund, L. Shu, V. Trifonov, S. Sigurdsson, A. Manolescu,
464 J. Zhu, Ö. Olafsson, J. Jakobsdottir, S. A. Lesley, J. To, J. Zhang, T. B. Harris, L. J. Launer, B.
465 Zhang, G. Eiriksdottir, X. Yang, A. P. Orth, L. L. Jennings, V. Gudnason, Co-regulatory networks
466 of human serum proteins link genetics to disease. *Science*. **361**, 769–773 (2018).
- 467 2. K. Suhre, M. Arnold, A. M. Bhagwat, R. J. Cotton, R. Engelke, J. Raffler, H. Sarwath, G. Thareja,
468 A. Wahl, R. K. Delisle, L. Gold, M. Pezer, G. Lauc, M. A. E. D. Selim, D. O. Mook-Kanamori, E. K.
469 Al-Dous, Y. A. Mohamoud, J. Malek, K. Strauch, H. Grallert, A. Peters, G. Kastenmüller, C.
470 Gieger, J. Graumann, Connecting genetic risk to disease end points through the human blood
471 plasma proteome. *Nat. Commun.* **8** (2017), doi:10.1038/ncomms14357.
- 472 3. L. Folkersen, E. Fauman, M. Sabater-Lleal, R. J. Strawbridge, M. Frånberg, B. Sennblad, D.
473 Baldassarre, F. Veglia, S. E. Humphries, R. Rauramaa, U. de Faire, A. J. Smit, P. Giral, S. Kurl, E.
474 Mannarino, S. Enroth, Å. Johansson, S. B. Enroth, S. Gustafsson, L. Lind, C. Lindgren, A. P.
475 Morris, V. Giedraitis, A. Silveira, A. Franco-Cereceda, E. Tremoli, IMPROVE study group, U.
476 Gyllensten, E. Ingelsson, S. Brunak, P. Eriksson, D. Ziemek, A. Hamsten, A. Mälarstig, Mapping
477 of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. *PLoS Genet.* **13**,
478 e1006706 (2017).
- 479 4. B. B. Sun, J. C. Maranville, J. E. Peters, D. Stacey, J. R. Staley, J. Blackshaw, S. Burgess, T. Jiang,
480 E. Paige, P. Surendran, C. Oliver-Williams, M. A. Kamat, B. P. Prins, S. K. Wilcox, E. S.
481 Zimmerman, A. Chi, N. Bansal, S. L. Spain, A. M. Wood, N. W. Morrell, J. R. Bradley, N. Janjic,
482 D. J. Roberts, W. H. Ouwehand, J. A. Todd, N. Soranzo, K. Suhre, D. S. Paul, C. S. Fox, R. M.
483 Plenge, J. Danesh, H. Runz, A. S. Butterworth, Genomic atlas of the human plasma proteome.
484 *Nature*. **558**, 73–79 (2018).
- 485 5. C. Yao, G. Chen, C. Song, J. Keefe, M. Mendelson, T. Huan, B. B. Sun, A. Laser, J. C. Maranville,
486 H. Wu, J. E. Ho, P. Courchesne, A. Lyass, M. G. Larson, C. Gieger, J. Graumann, A. D. Johnson,
487 J. Danesh, H. Runz, S.-J. J. Hwang, C. Liu, A. S. Butterworth, K. Suhre, D. Levy, Genome-wide
488 mapping of plasma protein QTLs identifies putatively causal genes and pathways for
489 cardiovascular disease. *Nat. Commun.* **9**, 3268 (2018).
- 490 6. A. Gilly, Y.-C. Park, G. Png, A. Barysenka, I. Fischer, T. Bjørnland, L. Southam, D. Suveges, S.
491 Neumeyer, N. W. Rayner, E. Tsafantakis, M. Karaleftheri, G. Dedoussis, E. Zeggini, Whole-
492 genome sequencing analysis of the cardiometabolic proteome. *Nat. Commun.* **11**, 6336
493 (2020).
- 494 7. K. Suhre, M. I. McCarthy, J. M. Schwenk, Genetics meets proteomics: perspectives for large
495 population-based studies. *Nat. Rev. Genet.*, 1–19 (2020).
- 496 8. J. Zheng, V. Haberland, D. Baird, V. Walker, P. C. Haycock, M. R. Hurle, A. Gutteridge, P. Erola,
497 Y. Liu, S. Luo, J. Robinson, T. G. Richardson, J. R. Staley, B. Elsworth, S. Burgess, B. B. Sun, J.
498 Danesh, H. Runz, J. C. Maranville, H. M. Martin, J. Yarmolinsky, C. Laurin, M. V. Holmes, J. Z.
499 Liu, K. Estrada, R. Santos, L. McCarthy, D. Waterworth, M. R. Nelson, G. D. Smith, A. S.
500 Butterworth, G. Hemani, R. A. Scott, T. R. Gaunt, Phenome-wide Mendelian randomization
501 mapping the influence of the plasma proteome on complex diseases. *Nat. Genet.* **52**, 1122–
502 1131 (2020).
- 503 9. L. Folkersen, S. Gustafsson, Q. Wang, D. H. Hansen, Å. K. Hedman, A. Schork, K. Page, D. V.
504 Zhernakova, Y. Wu, J. Peters, N. Eriksson, S. E. Bergen, T. S. Boutin, A. D. Bretherick, S. Enroth,
505 A. Kalnapenkis, J. R. Gådin, B. E. Suur, Y. Chen, L. Matic, J. D. Gale, J. Lee, W. Zhang, A. Quazi,
506 M. Ala-Korpela, S. H. Choi, A. Claringbould, J. Danesh, G. Davey Smith, F. de Masi, S. Elmståhl,
507 G. Engström, E. Fauman, C. Fernandez, L. Franke, P. W. Franks, V. Giedraitis, C. Haley, A.

- 508 Hamsten, A. Ingason, Å. Johansson, P. K. Joshi, L. Lind, C. M. Lindgren, S. Lubitz, T. Palmer, E.
509 Macdonald-Dunlop, M. Magnusson, O. Melander, K. Michaelsson, A. P. Morris, R. Mägi, M.
510 W. Nagle, P. M. Nilsson, J. Nilsson, M. Orho-Melander, O. Polasek, B. Prins, E. Pålsson, T. Qi,
511 M. Sjögren, J. Sundström, P. Surendran, U. Vösa, T. Werge, R. Wernersson, H. J. Westra, J.
512 Yang, A. Zhernakova, J. Ärnlöv, J. Fu, J. G. Smith, T. Esko, C. Hayward, U. Gyllensten, M.
513 Landen, A. Siegbahn, J. F. Wilson, L. Wallentin, A. S. Butterworth, M. V. Holmes, E. Ingelsson,
514 A. Mälarstig, Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931
515 individuals. *Nat. Metab.* **2**, 1135–1148 (2020).
- 516 10. M. Pietzner, E. Wheeler, J. Carrasco-Zanini, N. D. Kerrison, E. Oerton, M. Koprulu, J. Luan, A.
517 D. Hingorani, S. A. Williams, N. J. Wareham, C. Langenberg, *bioRxiv*, in press,
518 doi:10.1101/2021.03.18.435919.
- 519 11. T. Lindsay, K. Westgate, K. Wijndaele, S. Hollidge, N. Kerrison, N. Forouhi, S. Griffin, N.
520 Wareham, S. Brage, Descriptive epidemiology of physical activity energy expenditure in UK
521 adults (The Fenland study). *Int. J. Behav. Nutr. Phys. Act.* **16**, 126 (2019).
- 522 12. Associated code is available on GitHub. Extended Material and Methods are available as
523 Supplemental Material (2021), , doi:10.5281/zenodo.5385532.
- 524 13. M. Narayan, Disulfide bonds: protein folding and subcellular protein trafficking. *FEBS J.* **279**,
525 2272–82 (2012).
- 526 14. M. Uhlén, M. J. Karlsson, A. Hober, A. Svensson, J. Scheffel, D. Kotol, W. Zhong, A. Tebani, L.
527 Strandberg, F. Edfors, E. Sjöstedt, J. Mulder, A. Mardinoglu, A. Berling, S. Ekblad, M.
528 Dannemeyer, S. Kanje, J. Rockberg, M. Lundqvist, M. Malm, A. Volk, P. Nilsson, A. Månberg, T.
529 Dodig-crnkovic, E. Pin, M. Zwahlen, P. Oksvold, K. Von Feilitzen, R. S. Häussler, M. Hong, C.
530 Lindskog, F. Ponten, B. Katona, J. Vuu, E. Lindström, J. Nielsen, J. Robinson, B. Ayoglu, D.
531 Mahdessian, D. Sullivan, P. Thul, F. Danielsson, C. Stadler, E. Lundberg, G. Bergström, A.
532 Gummesson, B. G. Voldborg, H. Tegel, S. Hober, B. Forsström, J. M. Schwenk, L. Fagerberg, Å.
533 Sivertsson, The human secretome. **0274**, 1–9 (2019).
- 534 15. M. Pietzner, E. Wheeler, J. Carrasco-Zanini, J. Raffler, N. D. Kerrison, E. Oerton, V. P. W.
535 Auyeung, J. Luan, C. Finan, J. P. Casas, R. Ostroff, S. A. Williams, G. Kastenmüller, M. Ralser, E.
536 R. Gamazon, N. J. Wareham, A. D. Hingorani, C. Langenberg, Genetic architecture of host
537 proteins involved in SARS-CoV-2 infection. *Nat. Commun.* **11**, 6397 (2020).
- 538 16. M. Eslam, L. Valenti, S. Romeo, Genetics and epigenetics of NAFLD and NASH: Clinical impact.
539 *J. Hepatol.* **68**, 268–279 (2018).
- 540 17. S. BasuRay, Y. Wang, E. Smagris, J. C. Cohen, H. H. Hobbs, Accumulation of PNPLA3 on lipid
541 droplets is the basis of associated hepatic steatosis. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 9521–
542 9526 (2019).
- 543 18. P. N. Newsome, R. Cramb, S. M. Davison, J. F. Dillon, M. Foulerton, E. M. Godfrey, R. Hall, U.
544 Harrower, M. Hudson, A. Langford, A. Mackie, R. Mitchell-Thain, K. Sennett, N. C. Sheron, J.
545 Verne, M. Walmsley, A. Yeoman, Guidelines on the management of abnormal liver blood
546 tests. *Gut.* **67**, 6–19 (2018).
- 547 19. D. M. Tollefsen, C. J. Weigel, M. H. Kabeer, The presence of methionine or threonine at
548 position 381 in vitronectin is correlated with proteolytic cleavage at arginine 379. *J. Biol.*
549 *Chem.* **265**, 9778–9781 (1990).
- 550 20. D. I. Leavesley, A. S. Kashyap, T. Croll, M. Sivaramakrishnan, A. Shokoohmand, B. G. Hollier, Z.
551 Upton, Vitronectin—master controller or micromanager? *IUBMB Life.* **65**, 807–18 (2013).
- 552 21. V. Guarani, E. M. McNeill, J. A. Paulo, E. L. Huttlin, F. Fröhlich, S. P. Gygi, D. Van Vactor, J. W.
553 Harper, QIL1 is a novel mitochondrial protein required for MICOS complex stability and

- 554 cristae morphology. *Elife*. **4** (2015), doi:10.7554/eLife.06265.
- 555 22. P. B. Maguire, T. Donlon, M. Parsons, K. Wynne, E. Dillon, F. Ní Áinle, P. B. Szklanna,
556 Proteomic Analysis Reveals a Strong Association of β -Catenin With Cadherin Adherens
557 Junctions in Resting Human Platelets. *Proteomics*. **18** (2018), doi:10.1002/pmic.201700419.
- 558 23. D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T.
559 Doncheva, J. H. Morris, P. Bork, L. J. Jensen, C. von Mering, STRING v11: protein-protein
560 association networks with increased coverage, supporting functional discovery in genome-
561 wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
- 562 24. GTEx Consortium, The GTEx Consortium atlas of genetic regulatory effects across human
563 tissues. *Science*. **369**, 1318–1330 (2020).
- 564 25. C. Buccitelli, M. Selbach, mRNAs, proteins and the emerging principles of gene expression
565 control. *Nat. Rev. Genet.* **21**, 630–644 (2020).
- 566 26. U. Vösa, A. Claringbould, H.-J. Westra, M. Jan Bonder, P. Deelen, B. Zeng, H. Kirsten, A. Saha,
567 R. Kreuzhuber, S. Kasela, N. Pervjakova, I. Alvaes, M.-J. Fave, M. Agbessi, M. Christiansen, R.
568 Jansen, I. Seppälä, L. Tong, A. Teumer, K. Schramm, G. Hemani, J. Verlouw, H. Yaghootkar, R.
569 Sönmez, A. Brown, V. Kukushkina, A. Kalnapenkis, S. Rüeger, E. Porcu, J. Kronberg-Guzman, J.
570 Kettunen, J. Powell, B. Lee, F. Zhang, W. Arindrarto, F. Beutner, B. Consortium, H. Brugge, Q.
571 Consortium, J. Dmitreva, M. Elansary, B. P. Fairfax, M. Georges, B. T. Heijmans, M. Kähönen,
572 Y. Kim, J. C. Knight, P. Kovacs, K. Krohn, S. Li, M. Loeffler, U. M. Marigorta, H. Mei, Y.
573 Momozawa, M. Müller-Nurasyid, M. Nauck, M. Nivard, B. Penninx, J. Pritchard, O. Raitakari,
574 O. Rotzchke, E. P. Slagboom, C. DA Stehouwer, M. Stumvoll, P. Sullivan, J. Thiery, A. Tönjes, J.
575 van Dongen, M. van Iterson, J. Veldink, U. Völker, C. Wijmenga, M. Swertz, A. Andiappan, G.
576 W. Montgomery, S. Ripatti, M. Perola, Z. Kutalik, P. Awadalla, L. Milani, W. Ouwehand, K.
577 Downes, O. Stegle, A. Battle, J. Yang, P. M. Visscher, M. Scholz, G. Gibson, T. Esko, L. Franke,
578 M. J. Bonder, P. Deelen, B. Zeng, H. Kirsten, A. Saha, R. Kreuzhuber, S. Kasela, N. Pervjakova, I.
579 Alvaes, M.-J. Fave, M. Agbessi, M. Christiansen, R. Jansen, I. Seppälä, L. Tong, A. Teumer, K.
580 Schramm, G. Hemani, J. Verlouw, H. Yaghootkar, R. Sönmez, A. Brown, V. Kukushkina, A.
581 Kalnapenkis, S. Rüeger, E. Porcu, J. Kronberg-Guzman, J. Kettunen, J. Powell, B. Lee, F. Zhang,
582 W. Arindrarto, F. Beutner, H. Brugge, J. Dmitreva, M. Elansary, B. P. Fairfax, M. Georges, B. T.
583 Heijmans, M. Kähönen, Y. Kim, J. C. Knight, P. Kovacs, K. Krohn, S. Li, M. Loeffler, U. M.
584 Marigorta, H. Mei, Y. Momozawa, M. Müller-Nurasyid, M. Nauck, M. Nivard, B. Penninx, J.
585 Pritchard, O. Raitakari, O. Rotzchke, E. P. Slagboom, C. DA Stehouwer, M. Stumvoll, P.
586 Sullivan, P. Hoen, J. Thiery, A. Tönjes, J. van Dongen, M. van Iterson, J. Veldink, U. Völker, C.
587 Wijmenga, M. Swertz, A. Andiappan, G. W. Montgomery, S. Ripatti, M. Perola, Z. Kutalik, E.
588 Dermitzakis, S. Bergmann, T. Frayling, J. van Meurs, H. Prokisch, H. Ahsan, B. Pierce, T.
589 Lehtimäki, D. Boomsma, B. Psaty, S. Gharib, P. Awadalla, L. Milani, W. Ouwehand, K. Downes,
590 O. Stegle, A. Battle, J. Yang, P. M. Visscher, M. Scholz, G. Gibson, T. Esko, L. Franke, Unraveling
591 the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv*. **18**,
592 447367 (2018).
- 593 27. S. M. Sternson, D. Atasoy, Agouti-related protein neuron circuits that regulate appetite.
594 *Neuroendocrinology*. **100**, 95–102 (2014).
- 595 28. R. W. Baker, F. M. Hughson, Chaperoning SNARE assembly and disassembly. *Nat. Rev. Mol.*
596 *Cell Biol.* **17**, 465–79 (2016).
- 597 29. I. E. Jansen, J. E. Savage, K. Watanabe, J. Bryois, D. M. Williams, S. Steinberg, J. Sealock, I. K.
598 Karlsson, S. Hägg, L. Athanasiu, N. Voyle, P. Proitsi, A. Witoelar, S. Stringer, D. Aarsland, I. S.
599 Almdahl, F. Andersen, S. Bergh, F. Bettella, S. Bjornsson, A. Brækhus, G. Bråthen, C. de Leeuw,
600 R. S. Desikan, S. Djurovic, L. Dumitrescu, T. Fladby, T. J. Hohman, P. V. Jonsson, S. J. Kiddle, A.
601 Rongve, I. Saltvedt, S. B. Sando, G. Selbæk, M. Shoai, N. G. Skene, J. Snaedal, E. Stordal, I. D.

- 602 Ulstein, Y. Wang, L. R. White, J. Hardy, J. Hjerling-Leffler, P. F. Sullivan, W. M. van der Flier, R.
603 Dobson, L. K. Davis, H. Stefansson, K. Stefansson, N. L. Pedersen, S. Ripke, O. A. Andreassen,
604 D. Posthuma, Genome-wide meta-analysis identifies new loci and functional pathways
605 influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).
- 606 30. J. Schwartzentruber, S. Cooper, J. Z. Liu, I. Barrio-Hernandez, E. Bello, N. Kumasaka, A. M. H.
607 Young, R. J. M. Franklin, T. Johnson, K. Estrada, D. J. Gaffney, P. Beltrao, A. Bassett, Genome-
608 wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's
609 disease risk genes. *Nat. Genet.* (2021), doi:10.1038/s41588-020-00776-w.
- 610 31. S. Aggarwal, P. K. Dabla, S. Arora, Proxistin: An Epithelial Sodium Channel Regulator. *J.*
611 *biomarkers.* **2013**, 179864 (2013).
- 612 32. Y. Sugitani, A. Nishida, O. Inatomi, M. Ohno, T. Imai, M. Kawahara, K. Kitamura, A. Andoh,
613 Sodium absorption stimulator proxistin (PRSS8) has an anti-inflammatory effect via
614 downregulation of TLR4 signaling in inflammatory bowel disease. *J. Gastroenterol.* **55**, 408–
615 417 (2020).
- 616 33. M. Calvo-Rodríguez, C. García-Rodríguez, C. Villalobos, L. Núñez, Role of Toll Like Receptor 4
617 in Alzheimer's Disease. *Front. Immunol.* **11**, 1588 (2020).
- 618 34. T. A. O'Mara, D. M. Glubb, F. Amant, D. Annibaldi, K. Ashton, J. Attia, P. L. Auer, M. W.
619 Beckmann, A. Black, M. K. Bolla, H. Brauch, H. Brenner, L. Brinton, D. D. Buchanan, B.
620 Burwinkel, J. Chang-Claude, S. J. Chanock, C. Chen, M. M. Chen, T. H. T. Cheng, C. L. Clarke, M.
621 Clendenning, L. S. Cook, F. J. Couch, A. Cox, M. Crous-Bous, K. Czene, F. Day, J. Dennis, J.
622 Depreeuw, J. A. Doherty, T. Dörk, S. C. Dowdy, M. Dürst, A. B. Ekici, P. A. Fasching, B. L.
623 Fridley, C. M. Friedenreich, L. Fritschi, J. Fung, M. García-Closas, M. M. Gaudet, G. G. Giles, E.
624 L. Goode, M. Gorman, C. A. Haiman, P. Hall, S. E. Hankison, C. S. Healey, A. Hein, P.
625 Hillemanns, S. Hodgson, E. A. Hoivik, E. G. Holliday, J. L. Hopper, D. J. Hunter, A. Jones, C.
626 Krakstad, V. N. Kristensen, D. Lambrechts, L. Le Marchand, X. Liang, A. Lindblom, J. Lissowska,
627 J. Long, L. Lu, A. M. Magliocco, L. Martin, M. McEvoy, A. Meindl, K. Michailidou, R. L. Milne,
628 M. Mints, G. W. Montgomery, R. Nassir, H. Olsson, I. Orlov, G. Otton, C. Palles, J. R. B. Perry,
629 J. Peto, L. Pooler, J. Prescott, T. Proietto, T. R. Rebbeck, H. A. Risch, P. A. W. Rogers, M.
630 Rübner, I. Runnebaum, C. Sacerdote, G. E. Sarto, F. Schumacher, R. J. Scott, V. W. Setiawan,
631 M. Shah, X. Sheng, X.-O. Shu, M. C. Southey, A. J. Swerdlow, E. Tham, J. Trovik, C. Turman, J.
632 P. Tyrer, C. Vachon, D. VanDen Berg, A. Vanderstichele, Z. Wang, P. M. Webb, N. Wentzensen,
633 H. M. J. Werner, S. J. Winham, A. Wolk, L. Xia, Y.-B. Xiang, H. P. Yang, H. Yu, W. Zheng, P. D. P.
634 Pharoah, A. M. Dunning, P. Kraft, I. De Vivo, I. Tomlinson, D. F. Easton, A. B. Spurdle, D. J.
635 Thompson, Identification of nine new susceptibility loci for endometrial cancer. *Nat.*
636 *Commun.* **9**, 3166 (2018).
- 637 35. M. E. Binnerts, K.-A. Kim, J. M. Bright, S. M. Patel, K. Tran, M. Zhou, J. M. Leung, Y. Liu, W. E.
638 Lomas, M. Dixon, S. A. Hazell, M. Wagle, W.-S. Nie, N. Tomasevic, J. Williams, X. Zhan, M. D.
639 Levy, W. D. Funk, A. Abo, R-Spondin1 regulates Wnt signaling by inhibiting internalization of
640 LRP6. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 14700–5 (2007).
- 641 36. A. Geng, T. Wu, C. Cai, W. Song, J. Wang, Q. C. Yu, Y. A. Zeng, A novel function of R-spondin1
642 in regulating estrogen receptor expression independent of Wnt/ β -catenin signaling. *Elife.* **9**
643 (2020), doi:10.7554/eLife.56434.
- 644 37. A.-A. Chassot, S. T. Bradford, A. Auguste, E. P. Gregoire, E. Pailhoux, D. G. de Rooij, A. Schedl,
645 M.-C. Chaboissier, WNT4 and RSPO1 together are required for cell proliferation in the early
646 mouse gonad. *Development.* **139**, 4461–72 (2012).
- 647 38. A.-A. Chassot, F. Ranc, E. P. Gregoire, H. L. Roepers-Gajadien, M. M. Taketo, G. Camerino, D.
648 G. de Rooij, A. Schedl, M.-C. Chaboissier, Activation of beta-catenin signaling by Rspo1

- 649 controls differentiation of the mammalian ovary. *Hum. Mol. Genet.* **17**, 1264–77 (2008).
- 650 39. P. Edery, S. Lyonnet, L. M. Mulligan, A. Pelet, E. Dow, L. Abel, S. Holder, C. Nihoul-Fékété, B. A.
651 Ponder, A. Munnich, Mutations of the RET proto-oncogene in Hirschsprung’s disease. *Nature.*
652 **367**, 378–80 (1994).
- 653 40. E. A. Stahl, S. Raychaudhuri, E. F. Remmers, G. Xie, S. Eyre, B. P. Thomson, Y. Li, F. A. S.
654 Kurreeman, A. Zhernakova, A. Hinks, C. Guiducci, R. Chen, L. Alfredsson, C. I. Amos, K. G.
655 Ardlie, A. Barton, J. Bowes, E. Brouwer, N. P. Burtt, J. J. Catanese, J. Coblyn, M. J. H. Coenen,
656 K. H. Costenbader, L. A. Criswell, J. B. A. Crusius, J. Cui, P. I. W. De Bakker, P. L. De Jager, B.
657 Ding, P. Emery, E. Flynn, P. Harrison, L. J. Hocking, T. W. J. Huizinga, D. L. Kastner, X. Ke, A. T.
658 Lee, X. Liu, P. Martin, A. W. Morgan, L. Padyukov, M. D. Posthumus, T. R. D. J. Radstake, D. M.
659 Reid, M. Seielstad, M. F. Seldin, N. A. Shadick, S. Steer, P. P. Tak, W. Thomson, A. H. M. Van
660 Der Helm-Van Mil, I. E. Van Der Horst-Bruinsma, C. E. Van Der Schoot, P. L. C. M. Van Riel, M.
661 E. Weinblatt, A. G. Wilson, G. J. Wolbink, B. P. Wordsworth, C. Wijmenga, E. W. Karlson, R. E.
662 M. Toes, N. De Vries, A. B. Begovich, J. Worthington, K. A. Siminovitch, P. K. Gregersen, L.
663 Klareskog, R. M. Plenge, Genome-wide association study meta-analysis identifies seven new
664 rheumatoid arthritis risk loci. *Nat. Genet.* **42**, 508–514 (2010).
- 665 41. N. Harii, C. J. Lewis, V. Vasko, K. McCall, U. Benavides-Peralta, X. Sun, M. D. Ringel, M. Saji, C.
666 Giuliani, G. Napolitano, D. J. Goetz, L. D. Kohn, Thyrocytes express a functional toll-like
667 receptor 3: overexpression can be induced by viral infection and reversed by
668 phenylmethimazole and is associated with Hashimoto’s autoimmune thyroiditis. *Mol.*
669 *Endocrinol.* **19**, 1231–50 (2005).
- 670 42. C.-19 H. G. Initiative, Mapping the human genetic architecture of COVID-19. *Nature* (2021),
671 doi:10.1038/s41586-021-03767-x.
- 672 43. S. Zhou, G. Butler-Laporte, T. Nakanishi, D. R. Morrison, J. Afilalo, M. Afilalo, L. Laurent, M.
673 Pietzner, N. Kerrison, K. Zhao, E. Brunet-Ratnasingham, D. Henry, N. Kimchi, Z. Afrasiabi, N.
674 Rezk, M. Bouab, L. Petitjean, C. Guzman, X. Xue, C. Tselios, B. Vulesevic, O. Adeleye, T.
675 Abdullah, N. Almamlouk, Y. Chen, M. Chassé, M. Durand, C. Paterson, J. Normark, R. Frithiof,
676 M. Lipcsey, M. Hultström, C. M. T. Greenwood, H. Zeberg, C. Langenberg, E. Thysell, M.
677 Pollak, V. Mooser, V. Forgetta, D. E. Kaufmann, J. B. Richards, A Neanderthal OAS1 isoform
678 protects individuals of European ancestry against COVID-19 susceptibility and severity. *Nat.*
679 *Med.* **27**, 659–667 (2021).
- 680 44. M. B. Whyte, P. A. Kelly, E. Gonzalez, R. Arya, L. N. Roberts, Pulmonary embolism in
681 hospitalised patients with COVID-19. *Thromb. Res.* **195**, 95–99 (2020).
- 682 45. A. D. Joshi, C. Andersson, S. Buch, S. Stender, R. Noordam, L.-C. Weng, P. E. Weeke, P. L. Auer,
683 B. Boehm, C. Chen, H. Choi, G. Curhan, J. C. Denny, I. De Vivo, J. D. Eicher, D. Ellinghaus, A. R.
684 Folsom, C. Fuchs, M. Gala, J. Haessler, A. Hofman, F. Hu, D. J. Hunter, H. L. A. Janssen, J. H.
685 Kang, C. Kooperberg, P. Kraft, W. Kratzer, W. Lieb, P. L. Lutsey, S. Darwish Murad, B. G.
686 Nordestgaard, L. R. Pasquale, A. P. Reiner, P. M. Ridker, E. Rimm, L. M. Rose, C. M. Shaffer, C.
687 Schafmayer, R. M. Tamimi, A. G. Uitterlinden, U. Völker, H. Völzke, Y. Wakabayashi, J. L.
688 Wiggs, J. Zhu, D. M. Roden, B. H. Stricker, W. Tang, A. Teumer, J. Hampe, A. Tybjærg-Hansen,
689 D. I. Chasman, A. T. Chan, A. D. Johnson, Four Susceptibility Loci for Gallstone Disease
690 Identified in a Meta-analysis of Genome-Wide Association Studies. *Gastroenterology.* **151**,
691 351-363.e28 (2016).
- 692 46. C. N. Foley, J. R. Staley, P. G. Breen, B. B. Sun, P. D. W. Kirk, S. Burgess, J. M. M. Howson, A
693 fast and efficient colocalization algorithm for identifying shared genetic risk factors across
694 multiple traits. *Nat. Commun.* **12**, 764 (2021).
- 695 47. S.-Y. Y. Shin, E. B. Fauman, A.-K. K. Petersen, J. Krumsiek, R. Santos, J. Huang, M. Arnold, I.

696 Erte, V. Forgetta, T.-P. P. Yang, K. Walter, C. Menni, L. Chen, L. Vasquez, A. M. Valdes, C. L.
697 Hyde, V. Wang, D. Ziemek, P. Roberts, L. Xi, E. Grundberg, M. Waldenberger, J. B. Richards, R.
698 P. Mohny, M. V. Milburn, S. L. John, J. Trimmer, F. J. Theis, J. P. Overington, K. Suhre, M. J.
699 Brosnan, C. Gieger, G. Kastenmüller, T. D. Spector, N. Soranzo, N. Soranzo, An atlas of genetic
700 influences on human blood metabolites. *Nat. Genet.* **46**, 543–550 (2014).

701 48. F. Lammert, K. Gurusamy, C. W. Ko, J.-F. Miquel, N. Méndez-Sánchez, P. Portincasa, K. J. van
702 Erpecum, C. J. van Laarhoven, D. Q.-H. Wang, Gallstones. *Nat. Rev. Dis. Prim.* **2**, 16024 (2016).

703 49. A. R. Wood, T. Esko, J. Yang, S. Vedantam, T. H. Pers, S. Gustafsson, A. Y. Chu, K. Estrada, J.
704 Luan, Z. Kutalik, N. Amin, M. L. Buchkovich, D. C. Croteau-Chonka, F. R. Day, Y. Duan, T. Fall, R.
705 Fehrmann, T. Ferreira, A. U. Jackson, J. Karjalainen, K. S. Lo, A. E. Locke, R. Mägi, E. Mihailov,
706 E. Porcu, J. C. Randall, A. Scherag, A. A. E. Vinkhuyzen, H.-J. Westra, T. W. Winkler, T.
707 Workalemahu, J. H. Zhao, D. Absher, E. Albrecht, D. Anderson, J. Baron, M. Beekman, A.
708 Demirkan, G. B. Ehret, B. Feenstra, M. F. Feitosa, K. Fischer, R. M. Fraser, A. Goel, J. Gong, A.
709 E. Justice, S. Kanoni, M. E. Kleber, K. Kristiansson, U. Lim, V. Lotay, J. C. Lui, M. Mangino, I.
710 Mateo Leach, C. Medina-Gomez, M. A. Nalls, D. R. Nyholt, C. D. Palmer, D. Pasko, S.
711 Pechlivanis, I. Prokopenko, J. S. Ried, S. Ripke, D. Shungin, A. Stancáková, R. J. Strawbridge, Y.
712 J. Sung, T. Tanaka, A. Teumer, S. Trompet, S. W. van der Laan, J. van Setten, J. V Van Vliet-
713 Ostaptchouk, Z. Wang, L. Yengo, W. Zhang, U. Afzal, J. Arnlöv, G. M. Arscott, S. Bandinelli, A.
714 Barrett, C. Bellis, A. J. Bennett, C. Berne, M. Blüher, J. L. Bolton, Y. Böttcher, H. A. Boyd, M.
715 Bruinenberg, B. M. Buckley, S. Buyske, I. H. Caspersen, P. S. Chines, R. Clarke, S. Claudi-
716 Boehm, M. Cooper, E. W. Daw, P. A. De Jong, J. Deelen, G. Delgado, J. C. Denny, R.
717 Dhonukshe-Rutten, M. Dimitriou, A. S. F. Doney, M. Dörr, N. Eklund, E. Eury, L. Folkersen, M.
718 E. Garcia, F. Geller, V. Giedraitis, A. S. Go, H. Grallert, T. B. Grammer, J. Gräßler, H. Grönberg,
719 L. C. P. G. M. de Groot, C. J. Groves, J. Haessler, P. Hall, T. Haller, G. Hallmans, A. Hannemann,
720 C. A. Hartman, M. Hassinen, C. Hayward, N. L. Heard-Costa, Q. Helmer, G. Hemani, A. K.
721 Henders, H. L. Hillege, M. A. Hlatky, W. Hoffmann, P. Hoffmann, O. Holmen, J. J. Houwing-
722 Duistermaat, T. Illig, A. Isaacs, A. L. James, J. Jeff, B. Johansen, Å. Johansson, J. Jolley, T.
723 Juliusdottir, J. Juntila, A. N. Kho, L. Kinnunen, N. Klopp, T. Kocher, W. Kratzer, P. Lichtner, L.
724 Lind, J. Lindström, S. Lobbens, M. Lorentzon, Y. Lu, V. Lyssenko, P. K. E. Magnusson, A.
725 Mahajan, M. Maillard, W. L. McArdle, C. A. McKenzie, S. McLachlan, P. J. McLaren, C. Menni,
726 S. Merger, L. Milani, A. Moayyeri, K. L. Monda, M. A. Morken, G. Müller, M. Müller-Nurasyid,
727 A. W. Musk, N. Narisu, M. Nauck, I. M. Nolte, M. M. Nöthen, L. Oozageer, S. Pilz, N. W.
728 Rayner, F. Renstrom, N. R. Robertson, L. M. Rose, R. Roussel, S. Sanna, H. Scharnagl, S.
729 Scholtens, F. R. Schumacher, H. Schunkert, R. A. Scott, J. Sehmi, T. Seufferlein, J. Shi, K.
730 Silventoinen, J. H. Smit, A. V. Smith, J. Smolonska, A. V Stanton, K. Stirrups, D. J. Stott, H. M.
731 Stringham, J. Sundström, M. A. Swertz, A.-C. Syvänen, B. O. Tayo, G. Thorleifsson, J. P. Tyrer,
732 S. van Dijk, N. M. van Schoor, N. van der Velde, D. van Heemst, F. V. A. van Oort, S. H.
733 Vermeulen, N. Verweij, J. M. Vonk, L. L. Waite, M. Waldenberger, R. Wennauer, L. R. Wilkens,
734 C. Willenborg, T. Wilsgaard, M. K. Wojczynski, A. Wong, A. F. Wright, Q. Zhang, D. Arveiler, S.
735 J. L. Bakker, J. Beilby, R. N. Bergman, S. Bergmann, R. Biffar, J. Blangero, D. I. Boomsma, S. R.
736 Bornstein, P. Bovet, P. Brambilla, M. J. Brown, H. Campbell, M. J. Caulfield, A. Chakravarti, R.
737 Collins, F. S. Collins, D. C. Crawford, L. A. Cupples, J. Danesh, U. de Faire, H. M. den Ruijter, R.
738 Erbel, J. Erdmann, J. G. Eriksson, M. Farrall, E. Ferrannini, J. Ferrières, I. Ford, N. G. Forouhi, T.
739 Forrester, R. T. Gansevoort, P. V Gejman, C. Gieger, A. Golay, O. Gottesman, V. Gudnason, U.
740 Gyllensten, D. W. Haas, A. S. Hall, T. B. Harris, A. T. Hattersley, A. C. Heath, C. Hengstenberg,
741 A. A. Hicks, L. A. Hindorff, A. D. Hingorani, A. Hofman, G. K. Hovingh, S. E. Humphries, S. C.
742 Hunt, E. Hypponen, K. B. Jacobs, M.-R. Jarvelin, P. Jousilahti, A. M. Jula, J. Kaprio, J. J. P.
743 Kastelein, M. Kayser, F. Kee, S. M. Keinänen-Kiukaanniemi, L. A. Kiemeny, J. S. Kooner, C.
744 Kooperberg, S. Koskinen, P. Kovacs, A. T. Kraja, M. Kumari, J. Kuusisto, T. A. Lakka, C.
745 Langenberg, L. Le Marchand, T. Lehtimäki, S. Lupoli, P. A. F. Madden, S. Männistö, P.
746 Manunta, A. Marette, T. C. Matise, B. McKnight, T. Meitinger, F. L. Moll, G. W. Montgomery,

- 747 A. D. Morris, A. P. Morris, J. C. Murray, M. Nelis, C. Ohlsson, A. J. Oldehinkel, K. K. Ong, W. H.
748 Ouwehand, G. Pasterkamp, A. Peters, P. P. Pramstaller, J. F. Price, L. Qi, O. T. Raitakari, T.
749 Rankinen, D. C. Rao, T. K. Rice, M. Ritchie, I. Rudan, V. Salomaa, N. J. Samani, J. Saramies, M.
750 A. Sarzynski, P. E. H. Schwarz, S. Seibert, P. Sever, A. R. Shuldiner, J. Sinisalo, V.
751 Steinthorsdottir, R. P. Stolk, J.-C. Tardif, A. Tönjes, A. Tremblay, E. Tremoli, J. Virtamo, M.-C.
752 Vohl, Electronic Medical Records and Genomics (eMEMERGE) Consortium, MIGen
753 Consortium, PAGEGE Consortium, LifeLines Cohort Study, P. Amouyel, F. W. Asselbergs, T. L.
754 Assimes, M. Bochud, B. O. Boehm, E. Boerwinkle, E. P. Bottinger, C. Bouchard, S. Cauchi, J. C.
755 Chambers, S. J. Chanock, R. S. Cooper, P. I. W. de Bakker, G. Dedoussis, L. Ferrucci, P. W.
756 Franks, P. Froguel, L. C. Groop, C. A. Haiman, A. Hamsten, M. G. Hayes, J. Hui, D. J. Hunter, K.
757 Hveem, J. W. Jukema, R. C. Kaplan, M. Kivimaki, D. Kuh, M. Laakso, Y. Liu, N. G. Martin, W.
758 März, M. Melbye, S. Moebus, P. B. Munroe, I. Njølstad, B. A. Oostra, C. N. A. Palmer, N. L.
759 Pedersen, M. Perola, L. Pérusse, U. Peters, J. E. Powell, C. Power, T. Quertermous, R.
760 Rauramaa, E. Reinmaa, P. M. Ridker, F. Rivadeneira, J. I. Rotter, T. E. Saaristo, D. Saleheen, D.
761 Schlessinger, P. E. Slagboom, H. Snieder, T. D. Spector, K. Strauch, M. Stumvoll, J. Tuomilehto,
762 M. Uusitupa, P. van der Harst, H. Völzke, M. Walker, N. J. Wareham, H. Watkins, H.-E.
763 Wichmann, J. F. Wilson, P. Zanen, P. Deloukas, I. M. Heid, C. M. Lindgren, K. L. Mohlke, E. K.
764 Speliotes, U. Thorsteinsdottir, I. Barroso, C. S. Fox, K. E. North, D. P. Strachan, J. S. Beckmann,
765 S. I. Berndt, M. Boehnke, I. B. Borecki, M. I. McCarthy, A. Metspalu, K. Stefansson, A. G.
766 Uitterlinden, C. M. van Duijn, L. Franke, C. J. Willer, A. L. Price, G. Lettre, R. J. F. Loos, M. N.
767 Weedon, E. Ingelsson, J. R. O’Connell, G. R. Abecasis, D. I. Chasman, M. E. Goddard, P. M.
768 Visscher, J. N. Hirschhorn, T. M. Frayling, Defining the role of common variation in the
769 genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–86 (2014).
- 770 50. H. Springelkamp, A. I. Iglesias, A. Mishra, R. Höhn, R. Wojciechowski, A. P. Khawaja, A. Nag, Y.
771 X. Wang, J. J. Wang, G. Cuellar-Partida, J. Gibson, J. N. C. Bailey, E. N. Vithana, P. Gharahkhani,
772 T. Boutin, W. D. Ramdas, T. Zeller, R. N. Luben, E. Yonova-Doing, A. C. Viswanathan, S. Yazar,
773 A. J. Cree, J. L. Haines, J. Y. Koh, E. Souzeau, J. F. Wilson, N. Amin, C. Müller, C. Venturini, L. S.
774 Kearns, J. H. Kang, NEIGHBORHOOD Consortium, Y. C. Tham, T. Zhou, E. M. van Leeuwen, S.
775 Nickels, P. Sanfilippo, J. Liao, H. van der Linde, W. Zhao, L. M. E. van Koolwijk, L. Zheng, F.
776 Rivadeneira, M. Baskaran, S. J. van der Lee, S. Perera, P. T. V. M. de Jong, B. A. Oostra, A. G.
777 Uitterlinden, Q. Fan, A. Hofman, E.-S. Tai, J. R. Vingerling, X. Sim, R. C. W. Wolfs, Y. Y. Teo, H.
778 G. Lemij, C. C. Khor, R. Willemsen, K. J. Lackner, T. Aung, N. M. Jansonius, G. Montgomery, P.
779 S. Wild, T. L. Young, K. P. Burdon, P. G. Hysi, L. R. Pasquale, T. Y. Wong, C. C. W. Klaver, A. W.
780 Hewitt, J. B. Jonas, P. Mitchell, A. J. Lotery, P. J. Foster, V. Vitart, N. Pfeiffer, J. E. Craig, D. A.
781 Mackey, C. J. Hammond, J. L. Wiggs, C.-Y. Cheng, C. M. van Duijn, S. MacGregor, New insights
782 into the genetics of primary open-angle glaucoma based on meta-analyses of intraocular
783 pressure and optic disc characteristics. *Hum. Mol. Genet.* **26**, 438–453 (2017).
- 784 51. A. Wiberg, M. Ng, A. B. Schmid, R. W. Smillie, G. Baskozos, M. V Holmes, K. Künnapuu, R.
785 Mägi, D. L. Bennett, D. Furniss, A genome-wide association analysis identifies 16 novel
786 susceptibility loci for carpal tunnel syndrome. *Nat. Commun.* **10**, 1030 (2019).
- 787 52. E. Jorgenson, N. Makki, L. Shen, D. C. Chen, C. Tian, W. L. Eckalbar, D. Hinds, N. Ahituv, A.
788 Avins, A genome-wide association study identifies four novel susceptibility loci underlying
789 inguinal hernia. *Nat. Commun.* **6**, 10130 (2015).
- 790 53. N. Shrine, A. L. Guyatt, A. M. Erzurumluoglu, V. E. Jackson, B. D. Hobbs, C. A. Melbourne, C.
791 Batini, K. A. Fawcett, K. Song, P. Sakornsakolpat, X. Li, R. Boxall, N. F. Reeve, M. Obeidat, J. H.
792 Zhao, M. Wielscher, S. Weiss, K. A. Kentistou, J. P. Cook, B. B. Sun, J. Zhou, J. Hui, S. Karrasch,
793 M. Imboden, S. E. Harris, J. Marten, S. Enroth, S. M. Kerr, I. Surakka, V. Vitart, T. Lehtimäki, R.
794 J. Allen, P. S. Bakke, T. H. Beaty, E. R. Bleecker, Y. Bossé, C. A. Brandsma, Z. Chen, J. D. Crapo,
795 J. Danesh, D. L. DeMeo, F. Dudbridge, R. Ewert, C. Gieger, A. Gulsvik, A. L. Hansell, K. Hao, J. D.
796 Hoffman, J. E. Hokanson, G. Homuth, P. K. Joshi, P. Joubert, C. Langenberg, X. Li, L. Li, K. Lin, L.

- 797 Lind, N. Locantore, J. Luan, A. Mahajan, J. C. Maranville, A. Murray, D. C. Nickle, R. Packer, M.
798 M. Parker, M. L. Paynton, D. J. Porteous, D. Prokopenko, D. Qiao, R. Rawal, H. Runz, I. Sayers,
799 D. D. Sin, B. H. Smith, M. Soler Artigas, D. Sparrow, R. Tal-Singer, P. R. H. J. Timmers, M. Van
800 den Berge, J. C. Whittaker, P. G. Woodruff, L. M. Yerges-Armstrong, O. G. Troyanskaya, O. T.
801 Raitakari, M. Kähönen, O. Polašek, U. Gyllensten, I. Rudan, I. J. Deary, N. M. Probst-Hensch, H.
802 Schulz, A. L. James, J. F. Wilson, B. Stubbe, E. Zeggini, M. R. Jarvelin, N. Wareham, E. K.
803 Silverman, C. Hayward, A. P. Morris, A. S. Butterworth, R. A. Scott, R. G. Walters, D. A.
804 Meyers, M. H. Cho, D. P. Strachan, I. P. Hall, M. D. Tobin, L. V. Wain, New genetic signals for
805 lung function highlight pathways and chronic obstructive pulmonary disease associations
806 across multiple ancestries. *Nat. Genet.* **51**, 481–493 (2019).
- 807 54. P. J. McLaughlin, B. Bakall, J. Choi, Z. Liu, T. Sasaki, E. C. Davis, A. D. Marmorstein, L. Y.
808 Marmorstein, Lack of fibulin-3 causes early aging and herniation, but not macular
809 degeneration in mice. *Hum. Mol. Genet.* **16**, 3059–70 (2007).
- 810 55. I. Livingstone, V. N. Uversky, D. Furniss, A. Wiberg, The Pathophysiological Significance of
811 Fibulin-3. *Biomolecules.* **10** (2020), doi:10.3390/biom10091294.
- 812 56. L. Y. Marmorstein, F. L. Munier, Y. Arsenijevic, D. F. Schorderet, P. J. McLaughlin, D. Chung, E.
813 Traboulsi, A. D. Marmorstein, Aberrant accumulation of EFEMP1 underlies drusen formation
814 in Malattia Leventinese and age-related macular degeneration. *Proc. Natl. Acad. Sci. U. S. A.*
815 **99**, 13067–72 (2002).
- 816 57. B. Hollis, F. R. Day, A. S. Busch, D. J. Thompson, A. L. G. Soares, P. R. H. J. Timmers, A. Kwong,
817 D. F. Easton, P. K. Joshi, N. J. Timpson, PRACTICAL Consortium, 23andMe Research Team, K. K.
818 Ong, J. R. B. Perry, Genomic analysis of male puberty timing highlights shared genetic basis
819 with hair colour and lifespan. *Nat. Commun.* **11**, 1536 (2020).
- 820 58. F. R. Day, D. J. Thompson, H. Helgason, D. I. Chasman, H. Finucane, P. Sulem, K. S. Ruth, S.
821 Whalen, A. K. Sarkar, E. Albrecht, E. Altmaier, M. Amini, C. M. Barbieri, T. Boutin, A. Campbell,
822 E. Demerath, A. Giri, C. He, J. J. Hottenga, R. Karlsson, I. Kolcic, P.-R. Loh, K. L. Lunetta, M.
823 Mangino, B. Marco, G. McMahon, S. E. Medland, I. M. Nolte, R. Noordam, T. Nutile, L.
824 Paternoster, N. Perjakova, E. Porcu, L. M. Rose, K. E. Schraut, A. V Segrè, A. V Smith, L. Stolk,
825 A. Teumer, I. L. Andrusis, S. Bandinelli, M. W. Beckmann, J. Benitez, S. Bergmann, M. Bochud,
826 E. Boerwinkle, S. E. Bojesen, M. K. Bolla, J. S. Brand, H. Brauch, H. Brenner, L. Broer, T.
827 Brüning, J. E. Buring, H. Campbell, E. Catamo, S. Chanock, G. Chenevix-Trench, T. Corre, F. J.
828 Couch, D. L. Cousminer, A. Cox, L. Crisponi, K. Czene, G. Davey Smith, E. J. C N de Geus, R. de
829 Mutsert, I. De Vivo, J. Dennis, P. Devilee, I. dos-Santos-Silva, A. M. Dunning, J. G. Eriksson, P.
830 A. Fasching, L. Fernández-Rhodes, L. Ferrucci, D. Flesch-Janys, L. Franke, M. Gabrielson, I.
831 Gandin, G. G. Giles, H. Grallert, D. F. Gudbjartsson, P. Guénel, P. Hall, E. Hallberg, U. Hamann,
832 T. B. Harris, C. A. Hartman, G. Heiss, M. J. Hoening, J. L. Hopper, F. Hu, D. J. Hunter, M. Arfan
833 Ikram, H. Kyung Im, M.-R. Jarvelin, P. K. Joshi, D. Karasik, M. Kellis, Z. Kutalik, G. LaChance, D.
834 Lambrechts, C. Langenberg, L. J. Launer, J. S. E Laven, S. Lenarduzzi, J. Li, P. A. Lind, S.
835 Lindstrom, Y. Liu, an Luan, R. Mägi, A. Mannermaa, H. Mbarek, M. I. McCarthy, C. Meisinger,
836 T. Meitinger, C. Menni, A. Metspalu, K. Michailidou, L. Milani, R. L. Milne, G. W. Montgomery,
837 A. M. Mulligan, M. A. Nalls, P. Navarro, H. Nevanlinna, D. R. Nyholt, A. J. Oldehinkel, T. A. O, S.
838 Padmanabhan, A. Palotie, N. Pedersen, A. Peters, J. Peto, P. D. P Pharoah, A. Pouta, P. Radice,
839 I. Rahman, S. M. Ring, A. Robino, F. R. Rosendaal, I. Rudan, R. Rueedi, D. Ruggiero, C. F. Sala,
840 M. K. Schmidt, R. A. Scott, M. Shah, R. Sorice, M. C. Southey, U. Sovio, M. Stampfer, M. Steri,
841 K. Strauch, T. Tanaka, E. Tikkanen, N. J. Timpson, M. Traglia, T. Truong, J. P. Tyrer, A. G.
842 Uitterlinden, D. R. Velez Edwards, V. Vitart, U. Völker, P. Vollenweider, Q. Wang, E. Widen, K.
843 Willems van Dijk, G. Willemsen, R. Winqvist, B. H. R Wolffenbuttel, J. Hua Zhao, M.
844 Zoledziewska, M. Zygunt, B. Z. Alizadeh, D. I. Boomsma, M. Ciullo, F. Cucca, T. Esko, N.
845 Franceschini, C. Gieger, V. Gudnason, C. Hayward, P. Kraft, D. A. Lawlor, P. K. E Magnusson, N.

- 846 G. Martin, D. O. Mook-Kanamori, E. A. Nohr, O. Polasek, D. Porteous, A. L. Price, P. M. Ridker,
847 H. Snieder, T. D. Spector, D. Stöckl, D. Toniolo, S. Ulivi, J. A. Visser, H. Völzke, N. J. Wareham, J.
848 F. Wilson, A. B. Spurdle, U. Thorsteindottir, K. S. Pollard, D. F. Easton, J. Y. Tung, J. Chang-
849 Claude, D. Hinds, A. Murray, J. M. Murabito, K. Stefansson, K. K. Ong, J. R. B Perry, Genomic
850 analyses identify hundreds of variants associated with age at menarche and support a role for
851 puberty timing in cancer risk. *Nat. Publ. Gr.* **49** (2017), doi:10.1038/ng.3841.
- 852 59. E. A. Khramtsova, L. K. Davis, B. E. Stranger, The role of sex in the genomics of human
853 complex traits. *Nat. Rev. Genet.* **20**, 173–190 (2019).
- 854 60. H. Aschard, A perspective on interaction effects in genetic association studies. *Genet.*
855 *Epidemiol.* **40**, 678–688 (2016).
- 856 61. M. Oliva, M. Muñoz-Aguirre, S. Kim-Hellmuth, V. Wucher, A. D. H. Gewirtz, D. J. Cotter, P.
857 Parsana, S. Kasela, B. Balliu, A. Viñuela, S. E. Castel, P. Mohammadi, F. Aguet, Y. Zou, E. A.
858 Khramtsova, A. D. Skol, D. Garrido-Martín, F. Reverter, A. Brown, P. Evans, E. R. Gamazon, A.
859 Payne, R. Bonazzola, A. N. Barbeira, A. R. Hamel, A. Martinez-Perez, J. M. Soria, GTEx
860 Consortium, B. L. Pierce, M. Stephens, E. Eskin, E. T. Dermitzakis, A. V Segrè, H. K. Im, B. E.
861 Engelhardt, K. G. Ardlie, S. B. Montgomery, A. J. Battle, T. Lappalainen, R. Guigó, B. E.
862 Stranger, The impact of sex on gene expression across human tissues. *Science.* **369** (2020),
863 doi:10.1126/science.aba3066.
- 864 62. K. Mittelstrass, J. S. Ried, Z. Yu, J. Krumsiek, C. Gieger, Discovery of Sexual Dimorphisms in
865 Metabolic and Genetic Biomarkers. *PLoS Genet.* **7**, 1002215 (2011).
- 866 63. C. Finan, A. Gaulton, F. A. Kruger, R. T. Lumbers, T. Shah, J. Engmann, L. Galver, R. Kelley, A.
867 Karlsson, R. Santos, J. P. Overington, A. D. Hingorani, J. P. Casas, The druggable genome and
868 support for target identification and validation in drug development. *Sci. Transl. Med.* **9**
869 (2017), doi:10.1126/scitranslmed.aag1166.
- 870 64. D. Ochoa, A. Hercules, M. Carmona, D. Suveges, A. Gonzalez-Uriarte, C. Malangone, A.
871 Miranda, L. Fumis, D. Carvalho-Silva, M. Spitzer, J. Baker, J. Ferrer, A. Raies, O.
872 Razuvayevskaya, A. Faulconbridge, E. Petsalaki, P. Mutowo, S. Machlitt-Northen, G. Peat, E.
873 McAuley, C. K. Ong, E. Mountjoy, M. Ghousaini, A. Pierleoni, E. Papa, M. Pignatelli, G.
874 Koscielny, M. Karim, J. Schwartzentruber, D. G. Hulcoop, I. Dunham, E. M. McDonagh, Open
875 Targets Platform: supporting systematic drug-target identification and prioritisation. *Nucleic*
876 *Acids Res.* **49**, D1302–D1310 (2021).
- 877 65. I. Kola, J. Bell, A call to reform the taxonomy of human disease. *Nat. Rev. Drug Discov.* **10**,
878 641–2 (2011).
- 879 66. R. Sender, R. Milo, The distribution of cellular turnover in the human body. *Nat. Med.* **27**, 45–
880 48 (2021).
- 881 67. S. A. Williams, M. Kivimaki, C. Langenberg, A. D. Hingorani, J. P. Casas, C. Bouchard, C.
882 Jonasson, M. A. Sarzynski, M. J. Shipley, L. Alexander, J. Ash, T. Bauer, J. Chadwick, G. Datta,
883 R. K. DeLisle, Y. Hagar, M. Hinterberg, R. Ostroff, S. Weiss, P. Ganz, N. J. Wareham, Plasma
884 protein patterns as comprehensive indicators of health. *Nat. Med.* **25**, 1851–1857 (2019).
- 885 68. L. Gold, D. Ayers, J. Bertino, C. Bock, A. Bock, E. N. Brody, J. Carter, A. B. Dalby, B. E. Eaton, T.
886 Fitzwater, D. Flather, A. Forbes, T. Foreman, C. Fowler, B. Gawande, M. Goss, M. Gunn, S.
887 Gupta, D. Halladay, J. Heil, J. Heilig, B. Hicke, G. Husar, N. Janjic, T. Jarvis, S. Jennings, E.
888 Katilius, T. R. Keeney, N. Kim, T. H. Koch, S. Kraemer, L. Kroiss, N. Le, D. Levine, W. Lindsey, B.
889 Lollo, W. Mayfield, M. Mehan, R. Mehler, S. K. Nelson, M. Nelson, D. Nieuwlandt, M. Nikrad,
890 U. Ochsner, R. M. Ostroff, M. Otis, T. Parker, S. Pietrasiewicz, D. I. Resnicow, J. Rohloff, G.
891 Sanders, S. Sattin, D. Schneider, B. Singer, M. Stanton, A. Sterkel, A. Stewart, S. Stratford, J. D.
892 Vaught, M. Vrkljan, J. J. Walker, M. Watrobka, S. Waugh, A. Weiss, S. K. Wilcox, A. Wolfson, S.

- 893 K. Wolk, C. Zhang, D. Zichi, Aptamer-based multiplexed proteomic technology for biomarker
894 discovery. *PLoS One*. **5** (2010), doi:10.1371/journal.pone.0015004.
- 895 69. C. Giambartolomei, D. Vukcevic, E. E. Schadt, L. Franke, A. D. Hingorani, C. Wallace, V.
896 Plagnol, Bayesian test for colocalisation between pairs of genetic association studies using
897 summary statistics. *PLoS Genet*. **10**, e1004383 (2014).
- 898 70. B. Elsworth, M. Lyon, T. Alexander, Y. Liu, P. Matthews, J. Hallett, P. Bates, T. Palmer, V.
899 Haberland, G. D. Smith, J. Zheng, P. Haycock, T. R. Gaunt, G. Hemani, *bioRxiv*, in press,
900 doi:10.1101/2020.08.10.244293.
- 901 71. S. McCarthy, S. Das, W. Kretzschmar, O. Delaneau, A. R. Wood, A. Teumer, H. M. Kang, C.
902 Fuchsberger, P. Danecek, K. Sharp, Y. Luo, C. Sidore, A. Kwong, N. Timpson, S. Koskinen, S.
903 Vrieze, L. J. Scott, H. Zhang, A. Mahajan, J. Veldink, U. Peters, C. Pato, C. M. Van Duijn, C. E.
904 Gillies, I. Gandin, M. Mezzavilla, A. Gilly, M. Cocca, M. Traglia, A. Angius, J. C. Barrett, D.
905 Boomsma, K. Branham, G. Breen, C. M. Brummett, F. Busonero, H. Campbell, A. Chan, S.
906 Chen, E. Chew, F. S. Collins, L. J. Corbin, G. D. Smith, G. Dedoussis, M. Dorr, A. E. Farmaki, L.
907 Ferrucci, L. Forer, R. M. Fraser, S. Gabriel, S. Levy, L. Groop, T. Harrison, A. Hattersley, O. L.
908 Holmen, K. Hveem, M. Kretzler, J. C. Lee, M. McGue, T. Meitinger, D. Melzer, J. L. Min, K. L.
909 Mohlke, J. B. Vincent, M. Nauck, D. Nickerson, A. Palotie, M. Pato, N. Pirastu, M. McInnis, J. B.
910 Richards, C. Sala, V. Salomaa, D. Schlessinger, S. Schoenherr, P. E. Slagboom, K. Small, T.
911 Spector, D. Stambolian, M. Tuke, J. Tuomilehto, L. H. Van Den Berg, W. Van Rheenen, U.
912 Volker, C. Wijmenga, D. Toniolo, E. Zeggini, P. Gasparini, M. G. Sampson, J. F. Wilson, T.
913 Frayling, P. I. W. De Bakker, M. A. Swertz, S. McCarroll, C. Kooperberg, A. Dekker, D. Altshuler,
914 C. Willer, W. Iacono, S. Ripatti, N. Soranzo, K. Walter, A. Swaroop, F. Cucca, C. A. Anderson, R.
915 M. Myers, M. Boehnke, M. I. McCarthy, R. Durbin, G. Abecasis, J. Marchini, A reference panel
916 of 64,976 haplotypes for genotype imputation. *Nat. Genet*. **48**, 1279–1283 (2016).
- 917 72. J. Huang, B. Howie, S. McCarthy, Y. Memari, K. Walter, J. L. Min, P. Danecek, G. Malerba, E.
918 Trabetti, H. F. Zheng, G. Gambaro, J. B. Richards, R. Durbin, N. J. Timpson, J. Marchini, N.
919 Soranzo, S. Al Turki, A. Amuzu, C. A. Anderson, R. Anney, D. Antony, M. S. Artigas, M. Ayub, S.
920 Bala, J. C. Barrett, I. Barroso, P. Beales, M. Benn, J. Bentham, S. Bhattacharya, E. Birney, D.
921 Blackwood, M. Bobrow, E. Bochukova, P. F. Bolton, R. Bounds, C. Boustred, G. Breen, M.
922 Calissano, K. Carss, J. P. Casas, J. C. Chambers, R. Charlton, K. Chatterjee, L. Chen, A. Ciampi, S.
923 Cirak, P. Clapham, G. Clement, G. Coates, M. Cocca, D. A. Collier, C. Cosgrove, T. Cox, N.
924 Craddock, L. Crooks, S. Curran, D. Curtis, A. Daly, I. N. M. Day, A. Day-Williams, G. Dedoussis,
925 T. Down, Y. Du, C. M. Van Duijn, I. Dunham, S. Edkins, R. Ekong, P. Ellis, D. M. Evans, I. S.
926 Farooqi, D. R. Fitzpatrick, P. Flicek, J. Floyd, A. R. Foley, C. S. Franklin, M. Futema, L. Gallagher,
927 P. Gasparini, T. R. Gaunt, M. Geihs, D. Geschwind, C. Greenwood, H. Griffin, D. Grozeva, X.
928 Guo, X. Guo, H. Gurling, D. Hart, A. E. Hendricks, P. Holmans, L. Huang, T. Hubbard, S. E.
929 Humphries, M. E. Hurler, P. Hysi, V. Iotchkova, A. Isaacs, D. K. Jackson, Y. Jamshidi, J. Johnson,
930 C. Joyce, K. J. Karczewski, J. Kaye, T. Keane, J. P. Kemp, K. Kennedy, A. Kent, J. Keogh, F.
931 Khawaja, M. E. Kleber, M. Van Kogelenberg, A. Kolb-Kokocinski, J. S. Kooner, G. Lachance, C.
932 Langenberg, C. Langford, D. Lawson, I. Lee, E. M. Van Leeuwen, M. Lek, R. Li, Y. Li, J. Liang, H.
933 Lin, R. Liu, J. Lönqvist, L. R. Lopes, M. Lopes, J. Luan, D. G. MacArthur, M. Mangino, G.
934 Marenne, W. März, J. Maslen, A. Matchan, I. Mathieson, P. McGuffin, A. M. McIntosh, A. G.
935 McKechnie, A. McQuillin, S. Metrustry, N. Migone, H. M. Mitchison, A. Moayyeri, J. Morris,
936 R. Morris, D. Muddyman, F. Muntoni, B. G. Nordestgaard, K. Northstone, M. C. O'Donovan, S.
937 O'Rahilly, A. Onoufriadis, K. Oualkacha, M. J. Owen, A. Palotie, K. Panoutsopoulou, V. Parker,
938 J. R. Parr, L. Paternoster, T. Paunio, F. Payne, S. J. Payne, J. R. B. Perry, O. Pietilainen, V.
939 Plagnol, R. C. Pollitt, S. Povey, M. A. Quail, L. Quaye, L. Raymond, K. Rehnström, C. K. Ridout,
940 S. Ring, G. R. S. Ritchie, N. Roberts, R. L. Robinson, D. B. Savage, P. Scambler, S. Schiffels, M.
941 Schmidts, N. Schoenmakers, R. H. Scott, R. A. Scott, R. K. Semple, E. Serra, S. I. Sharp, A. Shaw,
942 H. A. Shihab, S. Y. Shin, D. Skuse, K. S. Small, C. Smee, G. D. Smith, L. Southam, O. Spasic-

- 943 Boskovic, T. D. Spector, D. St. Clair, B. St. Pourcain, J. Stalker, E. Stevens, J. Sun, G. Surdulescu,
944 J. Suvisaari, P. Syrris, I. Tachmazidou, R. Taylor, J. Tian, M. D. Tobin, D. Toniolo, M. Traglia, A.
945 Tybjaerg-Hansen, A. M. Valdes, A. M. Vandersteen, A. Varbo, P. Vijayarangakannan, P. M.
946 Visscher, L. V. Wain, J. T. R. Walters, G. Wang, J. Wang, Y. Wang, K. Ward, E. Wheeler, P.
947 Whincup, T. Whyte, H. J. Williams, K. A. Williamson, C. Wilson, S. G. Wilson, K. Wong, C. J. Xu,
948 J. Yang, G. Zaza, E. Zeggini, F. Zhang, P. Zhang, W. Zhang, Improved imputation of low-
949 frequency and rare variants using the UK10K haplotype reference panel. *Nat. Commun.* **6**, 1–
950 9 (2015).
- 951 73. C. Bycroft, C. Freeman, D. Petkova, G. Band, L. T. Elliott, K. Sharp, A. Motyer, D. Vukcevic, O.
952 Delaneau, J. O’Connell, A. Cortes, S. Welsh, A. Young, M. Effingham, G. McVean, S. Leslie, N.
953 Allen, P. Donnelly, J. Marchini, The UK Biobank resource with deep phenotyping and genomic
954 data. *Nature*. **562**, 203–209 (2018).
- 955 74. C. J. Willer, Y. Li, G. R. Abecasis, METAL: Fast and efficient meta-analysis of genomewide
956 association scans. *Bioinformatics*. **26**, 2190–2191 (2010).
- 957 75. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, GCTA: A tool for genome-wide complex trait
958 analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
- 959 76. W. McLaren, L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie, A. Thormann, P. Flicek, F.
960 Cunningham, The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 1–14 (2016).
- 961 77. T. Berisa, J. K. Pickrell, Approximately independent linkage disequilibrium blocks in human
962 populations. *Bioinformatics*. **32**, 283–5 (2016).
- 963 78. G. Hemani, J. Zheng, B. Elsworth, K. H. Wade, V. Haberland, D. Baird, C. Laurin, S. Burgess, J.
964 Bowden, R. Langdon, V. Y. Tan, J. Yarmolinsky, H. A. Shihab, N. J. Timpson, D. M. Evans, C.
965 Relton, R. M. Martin, G. Davey Smith, T. R. Gaunt, P. C. Haycock, The MR-Base platform
966 supports systematic causal inference across the human phenome. *Elife*. **7** (2018),
967 doi:10.7554/eLife.34408.
- 968 79. L. A. Lotta, M. Pietzner, I. D. Stewart, L. B. L. Wittemans, C. Li, R. Bonelli, J. Raffler, E. K. Biggs,
969 C. Oliver-Williams, V. P. W. Auyeung, J. Luan, E. Wheeler, E. Paige, P. Surendran, G. A.
970 Michelotti, R. A. Scott, S. Burgess, V. Zuber, E. Sanderson, A. Koulman, F. Imamura, N. G.
971 Forouhi, K.-T. Khaw, MacTel Consortium, J. L. Griffin, A. M. Wood, G. Kastenmüller, J. Danesh,
972 A. S. Butterworth, F. M. Gribble, F. Reimann, M. Bahlo, E. Fauman, N. J. Wareham, C.
973 Langenberg, A cross-platform approach identifies genetic regulators of human metabolism
974 and health. *Nat. Genet.* **53**, 54–64 (2021).
- 975 80. G. Yu, L. G. Wang, Y. Han, Q. Y. He, ClusterProfiler: An R package for comparing biological
976 themes among gene clusters. *Omi. A J. Integr. Biol.* **16**, 284–287 (2012).
- 977

978 **ACKNOWLEDGEMENTS**

979 We are grateful to all the volunteers and to the General Practitioners and practice staff for assistance
980 with recruitment. We thank the Fenland Study Investigators, Fenland Study Co-ordination team and
981 the Epidemiology Field, Data and Laboratory teams. Proteomic measurements were supported and
982 governed by a collaboration agreement between the University of Cambridge and SomaLogic. This
983 research has been conducted using the UK Biobank Resource (application no. 20361 and 44448).

984 **FUNDING**

985 The Fenland Study (10.22025/2017.10.101.00001) is funded by the Medical Research Council
986 (MC_UU_12015/1). We further acknowledge support for genomics from the Medical Research Council
987 (MC_PC_13046). ERG is supported by the National Institutes of Health (NIH) Awards R35HG010718,
988 R01HG011138, R01GM140287, and NIH/NIA AG068026. MAW, MA, and GK are supported by grants
989 from the National Institute on Aging (NIA): U01 AG061359, RF1 AG057452, and RF1 AG059093). JCZ
990 is supported by a 4-year Wellcome Trust PhD Studentship and the Cambridge Trust; MK is supported
991 by a Gates Fellowship; CL, EW, MP, JL, EO, IS, NK, and NJW are funded by the Medical Research Council
992 (MC_UU_00006/1 - Aetiology and Mechanisms). This work was supported in part by the UKRI/NIHR
993 Strategic Priorities Award in Multimorbidity Research for the Multimorbidity Mechanism and
994 Therapeutics Research Collaborative (MR/V033867/1).

995 **AUTHOR CONTRIBUTIONS**

996 Conceptualization: CL, MP, EW

997 Data curation/Software: EO, NDK, JL, MAW, JR

998 Formal Analysis: MP, EW, JCZ, AC, MK, IDS, JC

999 Methodology: RAS, ERG

1000 Visualization: MP, JCZ, MK

1001 Funding acquisition: CL, NJW

1002 Project administration: CL, NJW

1003 Supervision: CL, RAS, GK

1004 Writing – original draft: MP, CL, EW, JCZ, MK

1005 Writing – review & editing: EO, JC, IDS, ADH, NDK, MA, WA, SOR, NJW

1006 **COMPETING INTERESTS**

1007 RAS and AC are current employees and/or stockholders of GlaxoSmithKline. ERG receives an
1008 honorarium from the journal Circulation Research of the American Heart Association as a member of
1009 the Editorial Board. All other authors declare that they have no competing interests.

1010 **DATA and MATERIALS AVAILABILITY**

1011 Data from the Fenland cohort can be requested by bona fide researchers for specified scientific
1012 purposes via the study website ([https://www.mrc-
1013 epid.cam.ac.uk/research/studies/fenland/information-for-researchers/](https://www.mrc-epid.cam.ac.uk/research/studies/fenland/information-for-researchers/)). Data will either be shared
1014 through an institutional data sharing agreement or arrangements will be made for analyses to be
1015 conducted remotely without the necessity for data transfer. Summary statistics can be obtained from
1016 www.omicscience.org/apps/pgwas. Publicly available summary statistics for look-up and
1017 colocalisation of pQTLs were obtained from <https://gwas.mrcieu.ac.uk/> and
1018 <https://www.ebi.ac.uk/gwas/>. Associated code and scripts for the analysis is available on GitHub
1019 (https://github.com/MRC-Epid/pGWAS_discovery) and has been permanently archived using Zenodo
1020 (12).
1021

1022 **FIGURE LEGENDS**

1023

1024 **Fig. 1 Regional sentinel genetic variants associated ($p < 1.004 \times 10^{-11}$) with at least one protein target**
1025 **in up to 10,708 participants in the Fenland Study.** The lower panel maps the genomic locations of the
1026 genetic variants against the genomic locations of the protein-encoding genes. Genetic variants in close
1027 proximity to the protein-encoding gene (± 500 kb) are highlighted in pink (cis-pQTLs) and all others are
1028 shown in blue. Darker colors indicate lower p-values. The upper panel shows the number of associated
1029 protein targets for each genomic region, with dot sizes on top giving the number of approximately
1030 independent genetic variants ($r^2 < 0.1$), such that larger dots refer to more genetic variants in the
1031 region.

1032

1033 **Fig. 2 Classification of protein quantitative trait loci (pQTLs, cis and trans) and subsequent partition**
1034 **of the explained variance in plasma abundances of protein targets** A) Bar chart of pQTL classification
1035 based on GO term mapping (blue) or community mapping in a protein network derived by Gaussian
1036 graphical modeling (GGM; orange) of associated protein targets. Darker colors indicate cis-pQTLs and
1037 lighter colors trans-pQTLs. B) Data-driven protein network colored according to 191 identified protein
1038 communities. (C) a community-specific pQTL (*PNPLA3*) that was not captured by GO term mapping.
1039 Gene annotation as reported in the Materials and Methods. D) Absolute (upper panel) and relative
1040 (lower panel) explained variance in plasma abundances of protein targets by identified pQTLs. Coloring
1041 indicates contribution of the lead cis-pQTL (dark purple), secondary cis-pQTLs (purple), protein- or
1042 pathway-specific trans-pQTLs (pink), and unspecific trans-pQTLs (yellow). The inset displays the overall
1043 distribution of explained variance by each of the four categories. The variance explained was
1044 computed using linear regression models. A graphical display of effect size distributions can be found
1045 in Fig. S3.

1046

1047 **Fig. 3 Integration of gene and splicing quantitative trait loci (eQTLs and sQTLs).** A) Protein targets
1048 ordered by the number of tissues for which at least one of the cis-pQTLs was also a cis-eQTL as
1049 determined by statistical colocalization (posterior probability $> 80\%$ for a shared signal). Protein targets
1050 for which the eQTL showed evidence for a tissue-specific effect are indicated with black lines
1051 underneath. B) Same as A) but considering cis-sQTLs.

1052

1053 **Fig. 4 Causal gene assignment for associations reported in the GWAS catalog using identified cis-**
1054 **pQTLs.** Each panel displays the number of loci that have been reported in the GWAS catalog for a
1055 curated trait and were identified as protein quantitative trait in close proximity (± 500 kb) to the
1056 protein-encoding gene (cis-pQTL) in the current study. Mapping of GWAS loci and cis-pQTLs was done
1057 by linkage disequilibrium of reported variants ($r^2 > 0.8$). The upper panel displays the number of GWAS
1058 loci for which cis-pQTLs provided candidate causal genes. The middle panel displays the number of
1059 GWAS loci for which cis-pQTLs refined the list of candidate causal genes at the locus. The lower panel
1060 displays the number of GWAS loci with confirmative evidence from cis-pQTLs for already assigned
1061 candidate causal genes. Examples where gene prioritization was facilitated through pQTL but not gene
1062 expression QTL evidence are highlighted by a border around the box. Colors represent broad trait
1063 categories.

1064

1065 **Fig. 5 Network representation of phenome-wide colocalization analysis for protein-encoding loci.**

1066 The entire network is composed of 412 protein targets (squares) and 506 phenotypes (circles) as
1067 nodes, which are connected (n=1,859 edges) if there is evidence of a shared genetic signal (posterior
1068 probability >80%) and is shown in Fig. S6. This figure is restricted to connections between proteins and
1069 binary endpoints, mainly diseases, to increase visibility and show shared etiology among the clinical
1070 most relevant outcomes. Only protein targets and phenotypes with at least one connection are
1071 included. Effect directions are indicated by the line type aligned with the allele associated with higher
1072 amounts of the protein target (solid – positive, dashed – inverse association with the phenotype).
1073 Colors indicate categories of phenotypes. An interactive version of the figure can be found at
1074 www.omicscience.org/apps/pgwas.

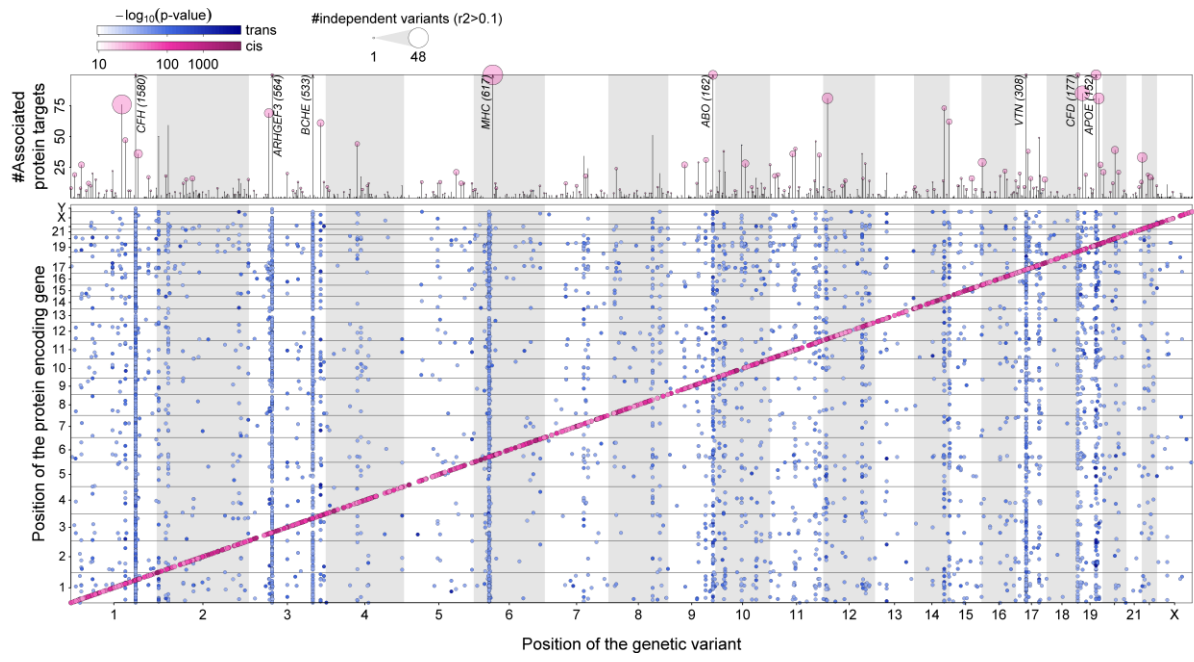
1075

1076 **Fig. 6 Selected phenotypic examples from the proteogenomic map. A** Plot visualizing convergence of

1077 genetic variants at the *SULT2A1* locus in relation to the LD with the candidate gene variant identified
1078 by multi-trait colocalization. Z-scores from GWAS for each annotated trait have been scaled by the
1079 absolute maximum, and dot size is proportional to the LD. Colors indicate the direction of effect
1080 aligned to the protein-increasing allele (red – positive, blue - inverse) The scheme on the right depicts
1081 the suggest mode of action by which higher SULT2A1 activity translates to higher risk of gallstones. **B**
1082 Same as A but for phenotypes colocalizing at the *EFEMP1* locus. The scheme on the right depicts a
1083 proposed mechanisms by which altered secretion of FBLN3 leads to the observed phenotypes. Stacked
1084 regional association plots for A and B can be found in Figs. S9 and S10.

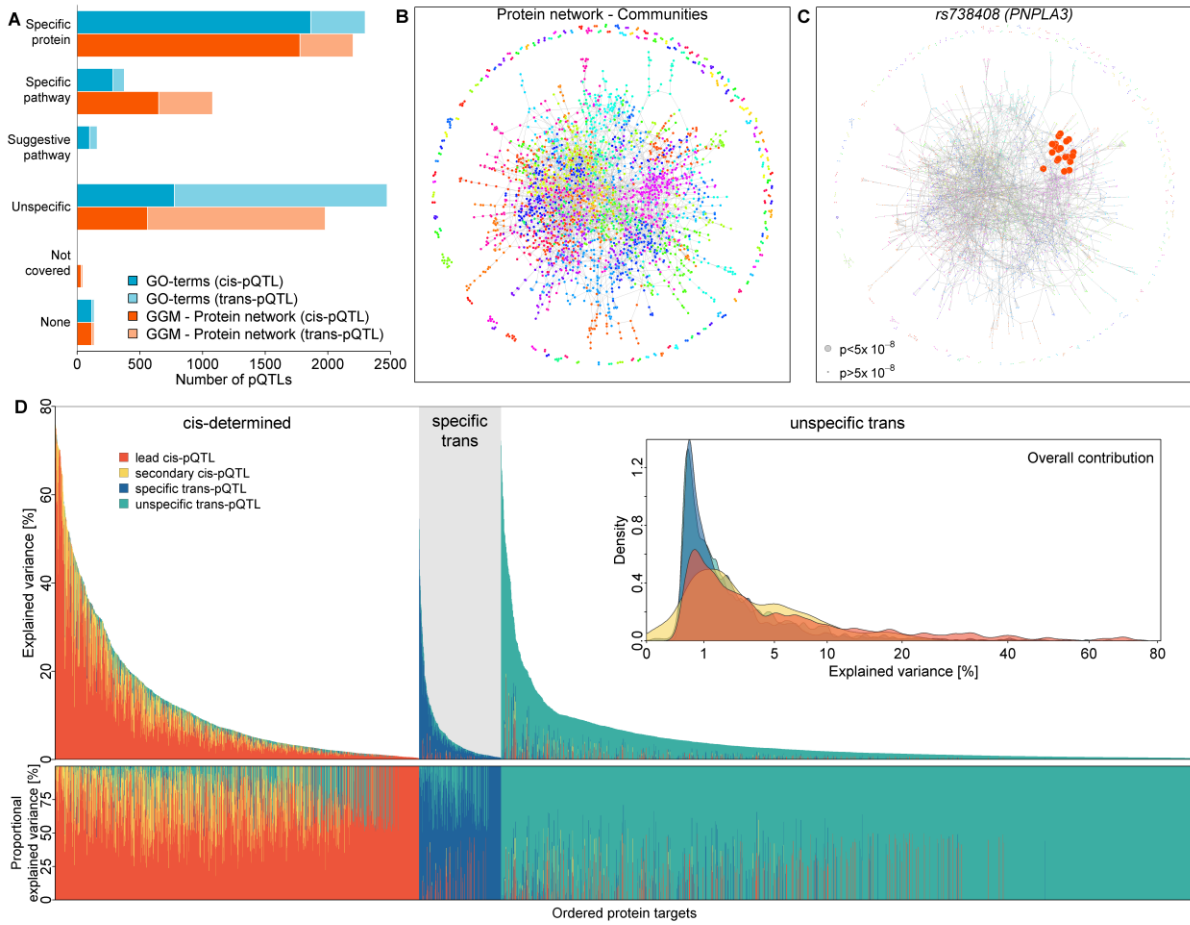
1085

1086 **Figure 1**
1087



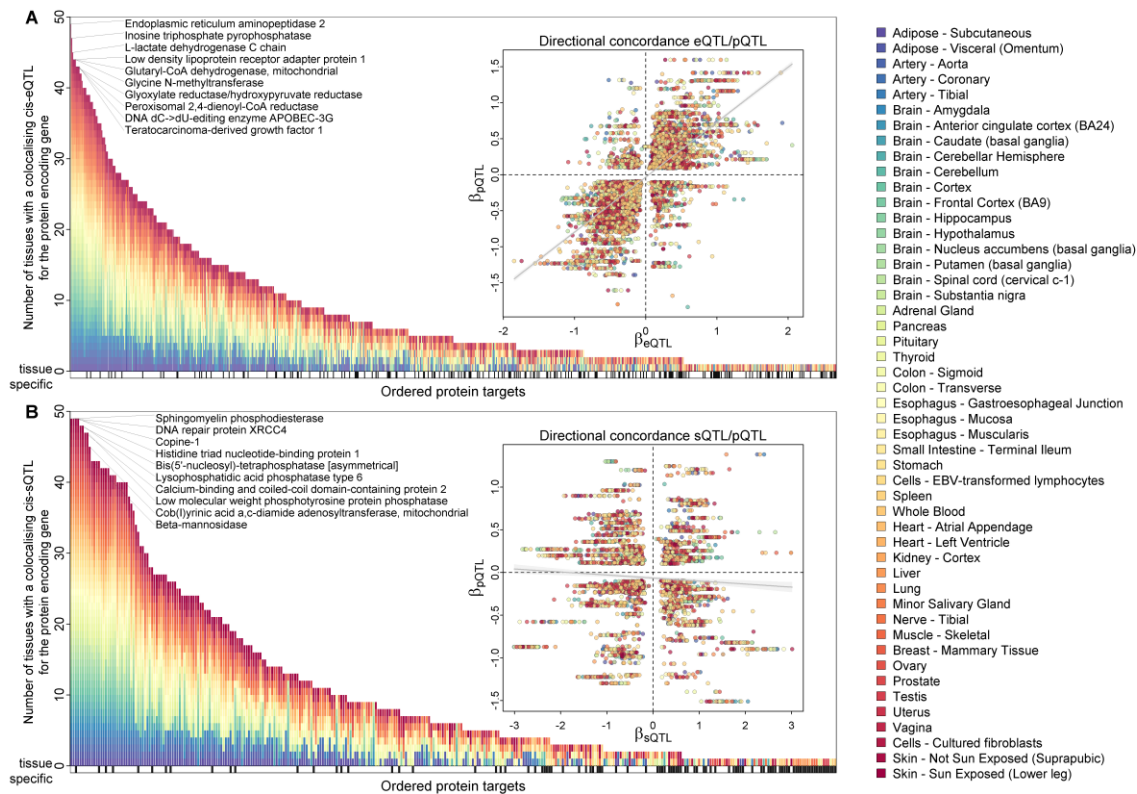
1088
1089

1090 **Figure 2**
1091

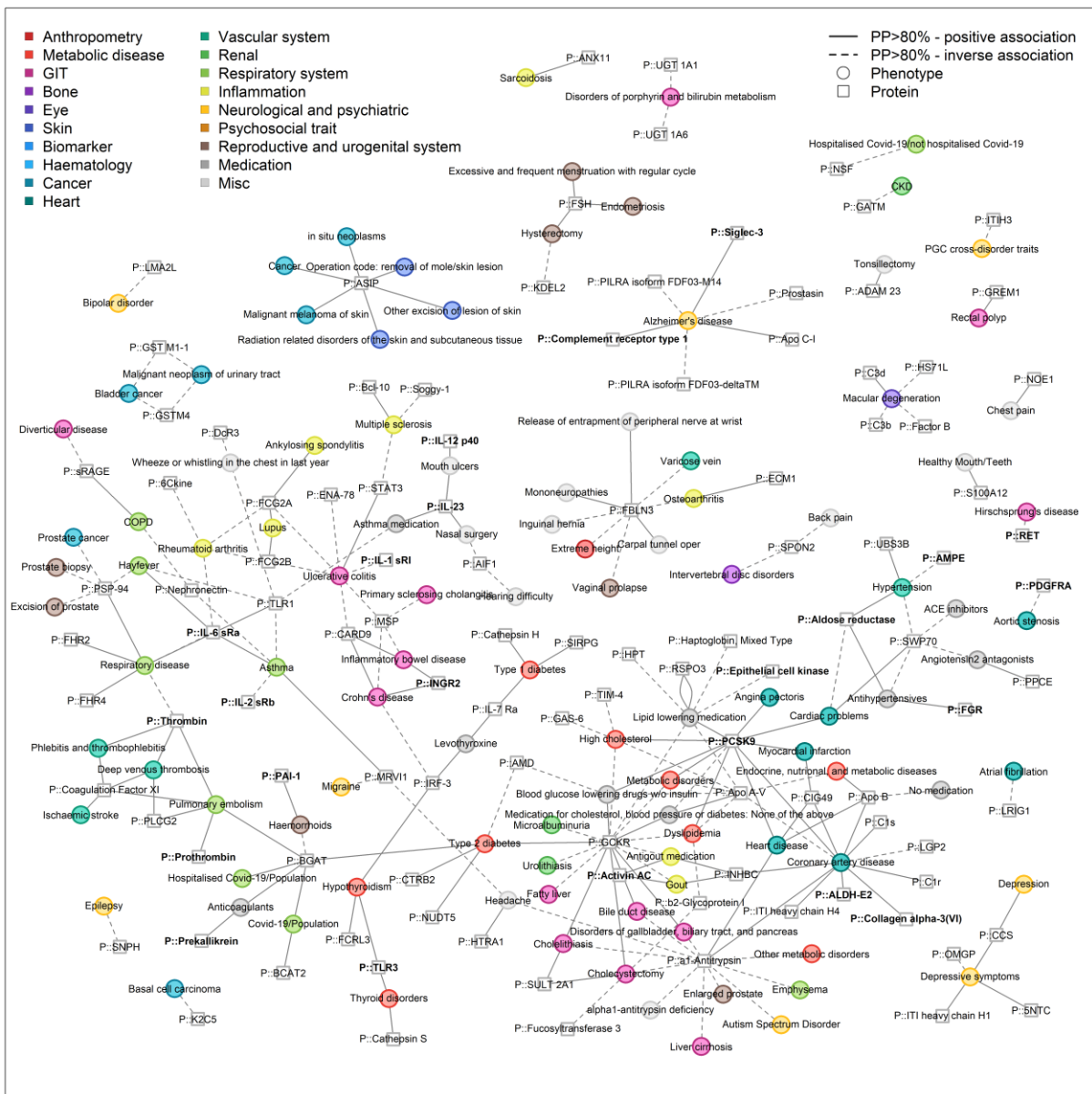


1092
1093

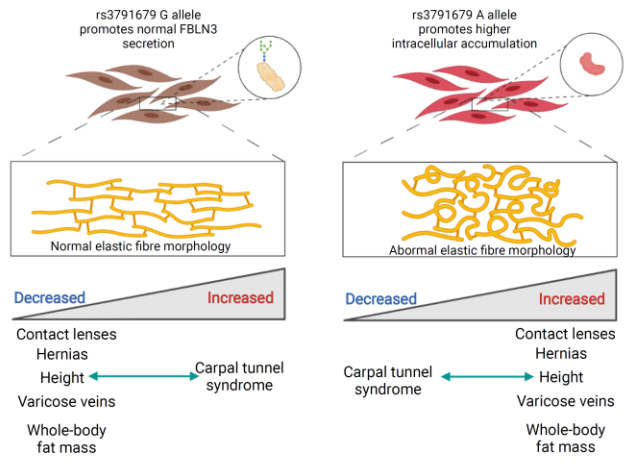
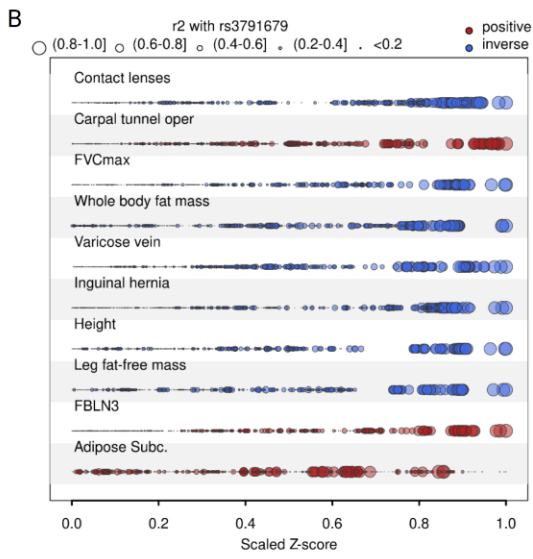
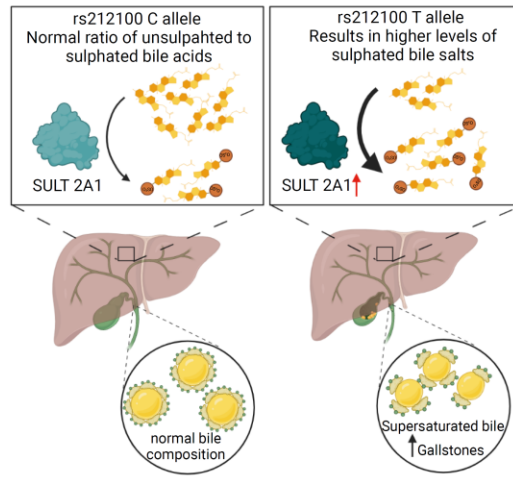
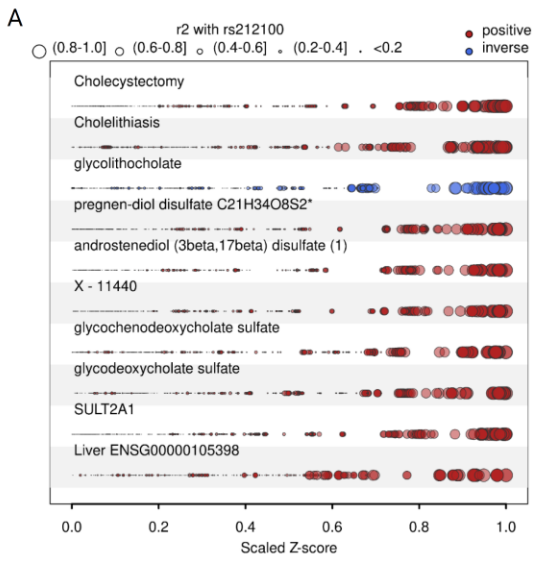
1094 **Figure 3**
 1095



1096
 1097



1106 **Figure 6**
 1107



1108
 1109

1110

1111

1112

1113

Supplementary Materials for

1114

Mapping the proteo-genomic convergence of human diseases

1116

1117 Maik Pietzner^{1,2*}, Eleanor Wheeler^{1*}, Julia Carrasco-Zanini¹, Adrian Cortes³, Mine Koprulu¹,

1118 Maria A. Wörheide⁴, Erin Oerton¹, James Cook¹, Isobel D. Stewart¹, Nicola D. Kerrison¹,

1119 Jian'an Luan¹, Johannes Raffler^{4,5}, Matthias Arnold^{4,6}, Wiebke Artl⁷, Stephen O'Rahilly⁸, Gabi

1120 Kastenmüller^{4,9}, Eric R. Gamazon^{10,11}, Aroon D. Hingorani^{12,13,14}, Robert A. Scott³, Nicholas

1121 J. Wareham^{1,13}, Claudia Langenberg^{1,2,13#}

1122

1123

Correspondence to: claudia.langenberg@mrc-epid.cam.ac.uk

1124

1125

1126 **This PDF file includes:**

1127

1128 Materials and Methods

1129 Figs. S1 to S10

1130 Tables S1 to S9 (Legends only)

1131

1132

1133

1134

1135

1136 **Materials and Methods**

1137

1138 *Study participants*

1139 The Fenland study is a population-based cohort of 12,435 participants of predominantly white-
1140 European ancestry born between 1950 and 1975 who underwent detailed phenotyping at the
1141 baseline visit from 2005-2015. Participants were recruited from general practice surgeries in
1142 the Cambridgeshire region in the UK. Exclusion criteria were: physician diagnosis of diabetes
1143 mellitus, inability to walk unaided, terminal illness, clinically diagnosed psychotic disorder,
1144 pregnancy, or lactation. The study was approved by the Cambridge Local Research Ethics
1145 Committee (NRES Committee – East of England Cambridge Central, ref. 04/Q0108/19), and
1146 all participants provided written informed consent. Use of human biological samples was in
1147 accord with the terms of the informed consents under an IRB/EC approved protocol.

1148 *Genotyping and imputation*

1149 Fenland participants were genotyped using three genotyping arrays: the Affymetrix UK
1150 Biobank Axiom array (OMICs, n=8994), Illumina Infinium Core Exome 24v1 (Core-Exome,
1151 n=1060) and Affymetrix SNP5.0 (GWAS, n=1402). Samples were excluded for the following
1152 reasons: 1) failed channel contrast (DishQC <0.82); 2) low call rate (<95%); 3) mismatch
1153 between reported and genetic sex; 4) heterozygosity outlier; 5) unusually high number of
1154 singleton genotypes; or 6) impossible identity-by-descent values. Single nucleotide
1155 polymorphisms (SNPs) were removed if: 1) call rate < 95%; 2) clusters failed Affymetrix
1156 SNPolisher standard tests and thresholds; 3) MAF was affected by plate; 4) SNP was a
1157 duplicate based on chromosome, position, and alleles (selecting the best probeset according
1158 to Affymetrix SNPolisher); 5) Hardy-Weinberg equilibrium $p < 10^{-6}$; 6) did not match the
1159 reference; or 7) MAF=0.

1160 Autosomes for the OMICS and GWAS subsets were imputed to the HRC (r1) panel using
1161 IMPUTE448, and the Core-Exome subset and the X-chromosome (for all subsets) were
1162 imputed to HRC.r1.1 using the Sanger imputation server (71). All three array subsets were
1163 also imputed to the UK10K+1000Gphase3 (72) panel using the Sanger imputation server to
1164 obtain additional variants that do not exist in the HRC reference panel. Variants with MAF <
1165 0.001, imputation quality (info) < 0.4, or Hardy Weinberg Equilibrium $p < 10^{-7}$ in any of the
1166 genotyping subsets were excluded from further analyses.

1167 *Proteomic measurements*

1168 Proteomic profiling of fasted EDTA plasma samples from 12,084 Fenland Study participants
1169 collected at baseline was performed by SomaLogic Inc. using an aptamer-based technology
1170 (SOMAscan proteomic assay). Relative protein abundances of 4,775 human protein targets

1171 were evaluated by 4,979 aptamers (SomaLogic V4), and a detailed description can be found
1172 elsewhere (67). Briefly, the SOMAscan assay uses a library of short single-stranded DNA
1173 molecules, which are chemically modified to specifically bind to protein targets, and the relative
1174 amount of aptamers binding to protein targets is determined using DNA microarrays. To
1175 account for variation in hybridization within runs, hybridization control probes are used to
1176 generate a hybridization scale factor for each sample. To control for total signal differences
1177 between samples due to variation in overall protein concentration or technical factors such as
1178 reagent concentration, pipetting, or assay timing, a ratio between each aptamer's measured
1179 value and a reference value was computed, and the median of these ratios was computed for
1180 each of the three dilution sets (20%, 0.5%, and 0.005%) and applied to each dilution set.
1181 Samples were removed if they were deemed by SomaLogic to have failed or did not meet our
1182 acceptance criteria of 0.25-4 for all scaling factors. In addition to passing SomaLogic QC, only
1183 human protein targets were taken forward for subsequent analysis (4,979 out of the 5,284
1184 aptamers). Aptamers' target annotation and mapping to UniProt accession numbers as well
1185 as Entrez gene identifiers were provided by SomaLogic, and we used those to obtain genomic
1186 positions of protein-encoding genes.

1187 *GWAS and meta-analysis*

1188 After excluding ancestry outliers and related individuals, 10,708 Fenland participants had both
1189 phenotypes and genetic data for the GWAS (OMICS=8,350, Core-Exome=1,026,
1190 GWAS=1,332). Within each genotyping subset, aptamer abundances were transformed to
1191 follow a normal distribution using the rank-based inverse normal transformation. Transformed
1192 aptamer abundances were then adjusted for age, sex, sample collection site, and 10 genetic
1193 principal components and the residuals used as input for the genetic association analyses.
1194 Genome-wide association was performed under an additive model using BGENIE (v1.3) (73).
1195 Results for the three genotyping arrays were combined in a fixed-effects meta-analysis in
1196 METAL (74). Following the meta-analysis up to 10.2 million genetic variants also present in
1197 the largest subset of the Fenland data (Fenland-OMICS) with overall MAF \geq 1% were taken
1198 forward for further analysis.

1199 *Conditional analysis*

1200 To identify conditionally independent signals in a genomic region associated with an aptamer,
1201 we performed conditional analysis as implemented in GCTA (75) using the `--s/ct` option, with
1202 a collinear cut-off of 0.1 and a p-value threshold of 1.004×10^{-11} . As a quality control step, we
1203 fitted a final model including all identified variants for a given genomic region using individual-
1204 level data in the largest available data set ('Fenland-OMICs') and discarded all variants no
1205 longer meeting genome-wide significance.

1206 We performed a forward stepwise selection procedure to identify secondary signals at each
1207 locus on the X-chromosome using SNPTEST v.2.5.2 to compute conditional association
1208 statistics based on individual-level data in the largest subset. Briefly, we defined conditionally
1209 independent signals as those emerging after conditioning on all previously selected signals in
1210 the locus until no signal was significant genome-wide.

1211 *Variant annotation*

1212 For each identified pQTL we first obtained all SNPs in at least moderate LD ($r^2 > 0.1$) using
1213 PLINK (version 2.0) and queried comprehensive annotations using the Variant Effect Predictor
1214 software (76) (version 98.3) with the --pick option. For each cis-pQTL we checked whether
1215 either the variant itself or a proxy in the encoding gene ($r^2 > 0.6$) was predicted to induce a
1216 change in the amino acid sequence of the associated protein, so-called protein-altering
1217 variants (PAVs). We further obtained domain information for each protein target via UniProt
1218 and tested whether the predicted amino acid exchange falls into any protein domain.

1219 *Locus definition*

1220 For each aptamer, we used a genome-wide significance threshold of 1.004×10^{-11} and defined
1221 non-overlapping regions by merging overlapping or adjoining 1 Mb intervals around all
1222 genome-wide significant variants (500 kb on either side), treating the extended MHC region
1223 (chr6:25.5–34.0 Mb) as one region. For each region we defined a regional sentinel variant as
1224 the most significant variant in the region. We defined genomic regions shared across aptamers
1225 based on a combination of multi-trait colocalization and LD-clumping ($r^2 > 0.8$). To this end, we
1226 first collapsed overlapping regions across all aptamers and divided those into approximately
1227 LD-independent regions (77). For each of the 776 regions we performed multi-trait
1228 colocalization as implemented in the R package *hyprcoloc* (46) to obtain clusters of aptamers
1229 sharing a common causal variant defined by a regional probability $> 80\%$ and a posterior
1230 probability of at least 50%. However, some clusters included protein targets for which the lead
1231 cis-pQTL was not in LD with the variant prioritized by HyPrColoc, most likely due to violation
1232 of the one causal variant assumption. We then subdivided those clusters manually and
1233 clumped signals based on a strong LD ($r^2 > 0.8$). This procedure resulted in 2,548 genomic
1234 regions with at least one cis-pQTL. Genetic variants from conditional analysis were assigned
1235 to the same locus as the sentinel signal. We classified pQTLs as cis-acting instruments if the
1236 variant was less than 500 kb away from the gene body of the protein-encoding gene.

1237 *Replication using the Olink PEA technology*

1238 To test for replication of pQTLs identified using the SomaScan platform with a complementary
1239 technique, we used in-house GWAS results for 1,069 proteins measured using Proximity
1240 Extension Assays provided by Olink in a subset of 485 individuals of the Fenland study. A

1241 detailed description of the GWAS can be found elsewhere (10) but largely followed the same
1242 protocol as described for SomaScan in the present paper. We first established which pQTLs
1243 were expected to replicate in the smaller sample size measured with both technologies by
1244 rerunning SNP – protein associations in the smaller subset. We identified 797 pQTLs that a)
1245 mapped to proteins covered on the Olink platform (N=543) and b) were significant at p-
1246 value<0.01. For each of the 797 SNP -protein pairs we repeated the same analysis but now
1247 instead of using SomaScan protein targets using the corresponding Olink protein as outcome.
1248 We considered associations to be replicated if they were directionally concordant and at least
1249 nominally significant at p-value<0.05 in this analysis.

1250 *Data-driven protein target network*

1251 We constructed a data-driven protein network using Gaussian graphical modeling similar to
1252 previous work (2). First, for each pair of highly correlated aptamers targeting the same protein
1253 (Pearson correlation>0.5), we dropped one at random to avoid artificial null results in the
1254 network, leaving 4,929 aptamers. We next computed residual plasma abundances by
1255 accounting for the effects of age, sex, test site, and the first three proteomic principal
1256 components using linear regression models. We finally used the R implementation
1257 `ggm.estimate.pcor` from the package *GeneNet* to estimate full-order partial correlations among
1258 residual aptamer abundances and only kept edges meeting a stringent Bonferroni cut-off
1259 ($p < 2.05 \times 10^{-9}$). The final network consisted of 2,936 aptamers and 4,669 edges. We performed
1260 community detection using the Girvan-Newman algorithm as implemented in the R package
1261 *igraph* and obtained 191 distinct protein communities in the network.

1262 *Classification of pQTLs*

1263 We classified the specificity of pQTLs based on two complementary approaches. We first
1264 derived the set of associated aptamers at a nominal GWAS-threshold of significance ($p < 5 \times 10^{-8}$)
1265 for each of the 5,442 uniquely identified variants. Next, we tested whether all aptamers 1)
1266 belonged to the same protein target (including complexes), 2) could be assigned to a common
1267 GO term (as previously described (15)), or 3) belonged to the same protein community in the
1268 data-driven protein network. We classified variants fulfilling 1) as protein-specific, variants
1269 fulfilling 2) and 3) as pathway-specific, variants fulfilling either 2) or 3) as suggestive pathway-
1270 specific, and any other variant as non-specific.

1271 *Variance explained*

1272 We estimated the variance explained by pQTLs for plasma levels of each aptamer with at least
1273 one associated pQTL using different sets of genetic instruments. To this end, we successively
1274 included 1) the lead cis-pQTL, 2) secondary cis-pQTLs, 3) specific trans-pQTLs, and 4) all
1275 trans-pQTLs in a linear regression model using residual aptamer abundances as outlined in

1276 the GWAS section. We used the R^2 of the entire model as an estimate for variance explained.
1277 We did this analysis in the largest set of Fenland participants genotyped on a single array.

1278 *Candidate causal gene assignment*

1279 We integrated functional assignments with protein-protein interaction network data to assign
1280 putative causal genes for each variant - aptamer pair. Briefly, for each variant close to the
1281 protein-encoding gene (± 500 kb) we assigned this gene as candidate causal gene. For pQTLs
1282 in *trans* we used a scoring system by integrating 1) search for functional variants (VEP score
1283 1-12 and $R^2 > 0.6$), 2) LD with an eQTL ($R^2 > 0.8$), 3) distance of the gene products from the 20
1284 closest genes to the protein target associated with the pQTL in the STRING protein-protein
1285 interaction, and 4) the closest gene. We assigned a score of two for 1) - 3) and a score of one
1286 for the closest gene and retained the gene(s) with the highest score(s) as (a) possible
1287 candidate(s). In a second step, we aligned gene assignments across all aptamers based on
1288 the definition of genomic loci, that is, for each locus shared across multiple aptamers we
1289 repeated the scoring system taking into account all possible candidate genes using a score of
1290 three for cis assignments. This procedure allowed us to refine assignment at otherwise poorly
1291 defined trans loci and to obtain higher confidence scores at each locus.

1292 *Incorporation of gene expression data*

1293 We incorporated gene expression and splicing QTL data by cross-referencing all cis-pQTLs
1294 with cis-eQTL/sQTLs identified in the GTEx version 8 release across 49 distinct tissues using
1295 an LD threshold of $r^2 > 0.8$ to identify likely similar signals (24). If at least on cis-pQTL mapped
1296 to a corresponding cis-eQTL or cis-sQTL for the protein-encoding gene, we used statistical
1297 colocalization (69) to test for a shared genetic signal between protein abundance measured
1298 in plasma and expression of the respective gene across all available tissues. We considered
1299 a PP > 80% as evidence for a highly likely shared signal. We used a 500 kb window around the
1300 cis-pQTL for colocalization analysis. To identify cis-pQTL/cis-eQTL pairings which are likely
1301 to be tissue-specific, we obtained Z-scores for the candidate variant across all tissues and
1302 divided by the square root of the respective tissue sample size to normalize across tissues.
1303 We defined such pairings as tissue-specific if the normalized Z-score was more than ± 5 times
1304 the median absolute deviation (MAD) away from the median normalized Z-score, which
1305 represents a robust measure of outlier detection. For transcripts expressed in fewer than five
1306 tissues, we considered the number of colocalizing tissues as a threshold for tissue specificity.
1307 To account for possible measurement artefacts, we repeated this process for 277 protein
1308 targets with evidence that a secondary signal was in LD with a cis-eQTL/sQTL using summary
1309 statistics conditioned on the lead pQTL in the region for colocalization. All GTEx variant-gene
1310 cis-eQTL and cis-sQTL associations from each tissue were downloaded in January 2020 from

1311 <https://console.cloud.google.com/storage/browser/gtex-resources>. We further tested for
1312 colocalization with gene expression determined from whole blood using data from the
1313 eQTLGen consortium (26), which included data on more than 30,000 participants.

1314 *Annotation of GWAS catalog loci*

1315 We downloaded genome-wide significant summary statistics from the GWAS catalog (date
1316 25/01/2021) and tested whether any of the identified pQTLs or proxies ($r^2 > 0.8$) have been
1317 reported to be associated with any non-proteomic trait, that is omitting any results that related
1318 to multiplex proteomic assays. Out of 227,631 entries, 113,618 entries passed this and
1319 additional filtering steps (missing effect estimates, missing risk allele, and not passing
1320 genome-wide significance). We next assessed whether for 3,139 lead and secondary pQTLs
1321 in cis, linkage by LD ($r^2 > 0.8$) to findings reported in the GWAS catalog may help to prioritize
1322 potential causal genes for the phenotypic trait. We compared the reported or mapped gene
1323 (closest gene assigned by the GWAS catalog) to the protein-encoding gene at the locus. This
1324 left us with 4,133 entries, including 590 cis-regions and 556 mapped GWAS traits. We collated
1325 loci into single entries to account for the variety of entries at pleiotropic loci and further dropped
1326 32 cis-regions for which more than one protein target mapped to the signals in the region,
1327 resulting in 3,868 entries.

1328 *Phenome-wide scans at protein-encoding loci*

1329 We performed phenome-wide scans using statistical colocalization for 1,584 protein targets
1330 where we had evidence for at least one cis-pQTL. Briefly, we queried the Open GWAS
1331 database (78) as well as an in-house database of curated GWAS summary statistics hosted
1332 by GSK using a defined region (± 500 kb) around the protein-encoding gene body and tested
1333 whether any of the traits in the databases showed a high posterior probability (PP) of shared
1334 genetic signal with plasma concentrations of the encoded protein target using statistical
1335 colocalization (69). We chose a cut-off of $PP > 80\%$ to declare that a protein target and a
1336 phenotypic trait are highly likely to share a genetic signal at a locus. We used a conservative
1337 prior setting, prior probability of 1×10^{-6} that both traits have a common genetic signal, along
1338 with a check that the regional lead signals for the protein and the trait are in strong LD ($r^2 > 0.8$)
1339 to declare colocalization. We repeated this analysis using conditional statistics for the protein
1340 target accounting for a possible binding artefact introduced by the lead signal at the locus. We
1341 manually curated common trait names to reduce redundancy of phenotypes across both
1342 databases and kept the association with the largest PP for each mapped trait for a shared
1343 signal when all other definitions supported colocalization. Finally, we collapsed all pairs of
1344 protein targets and phenotypic traits with high evidence for colocalization into a protein-
1345 disease network by drawing an edge between a protein target and a phenotypic trait if there

1346 was a high PP (>80%) for a shared signal. We used the lead signal at the locus aligned to the
1347 protein-increasing allele to indicate effect directions and visualized the network using the
1348 *igraph* R package. We report Mendelian Randomization estimates for binary outcomes derived
1349 from the UK Biobank as odds ratios, by transforming β -effect estimates according to the
1350 following formula: $\log(\text{odds ratio}) = \beta / (\mu * (1 - \mu))$, where μ = case fraction, since analysis
1351 were performed using a linear regression frame work for computational efficacy.

1352 *COVID-19 summary statistics*

1353 We downloaded genome-wide summary statistics for four different outcome definitions of
1354 COVID-19 from the Human Genetics Initiative (<https://www.covid19hg.org/>). These included
1355 A2 (very severe respiratory confirmed COVID-19 vs. population), B1 (hospitalized COVID-19
1356 vs. not hospitalized COVID-19), B2 (hospitalized COVID-19 vs. population), C2 (COVID-19
1357 vs. population). To map the LD panel for colocalization analysis, we restricted those statistics
1358 to participants of European ancestry and excluded results contributed by 23&me.

1359 *Metabolite GWAS results for SULT2A1*

1360 We extracted summary statistics for 69 metabolites associated with the lead *cis*-pQTL
1361 ($p < 1 \times 10^{-6}$) for SULT2A1 from an in-house metabolome-wide GWAS based on the EPIC-
1362 Norfolk cohort, methods of which have been described previously(79).

1363 *Pathway enrichment analysis*

1364 We performed GO term enrichment as implemented in the R package *clusterProfiler* (80)
1365 separately for each mapped trait in the protein-phenotype network that had at least three
1366 colocalizing protein targets. We used all three GO term categories for this purpose and
1367 retained only pathways that met statistical significance after correction for multiple testing
1368 using the Benjamini-Hochberg procedure controlling the false-discovery rate at 5%.

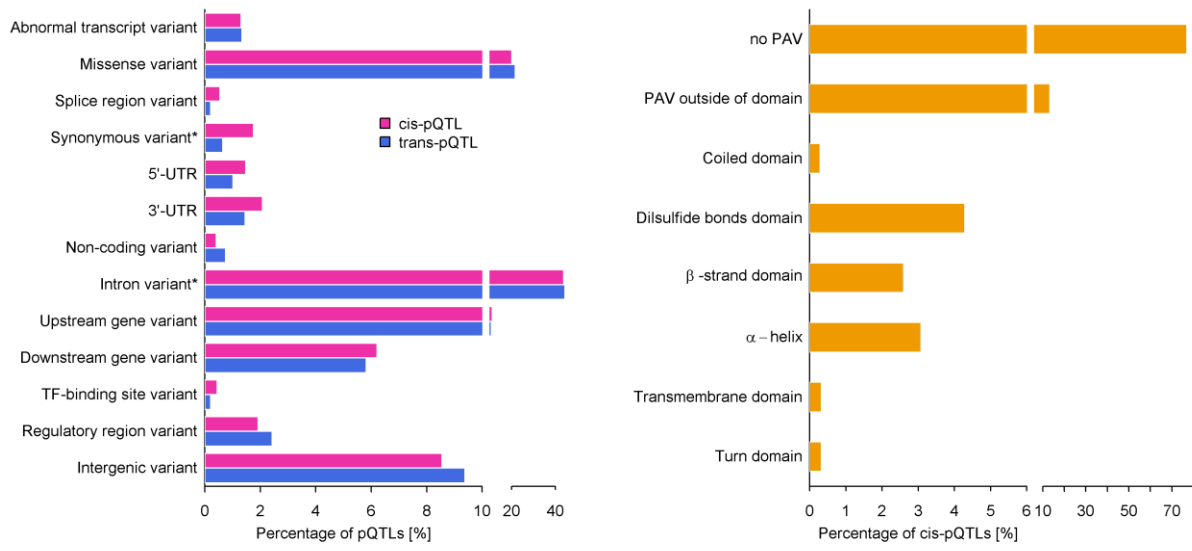
1369 *Testing for effect modification by age and sex*

1370 We included an interaction term between the *cis*-pQTL and age (continuous) or sex in a linear
1371 regression model with the same adjustments as in the main analysis to test for potential
1372 differences of the 417 *cis*-pQTLs included in the proteogenomic map by age or sex. For
1373 interactions significant below the Bonferroni-corrected significance level of $0.05/(2*417)$, we
1374 estimated group-specific estimates, dichotomizing age at the median age of 49 years, by
1375 running linear regression models within each stratum using the largest set of individuals with
1376 the same genotype platform (N=8350).

1377 *Mapping of druggable targets*

1378 To annotate druggable targets we merged the list of proteins targeted by the SomaScan V4
1379 platform with an updated list of druggable genes from Finan *et al.* (63) based on common gene
1380 entries. We deprioritized drugs with multiple reported side effects in ongoing or completed
1381 clinical trials, missing efficacy, discordant effect directions between the estimated effect of life-
1382 long higher protein concentrations mediated by genetic variants and action of the drug, a lack
1383 of clinical data, or having been withdrawn from major markets.

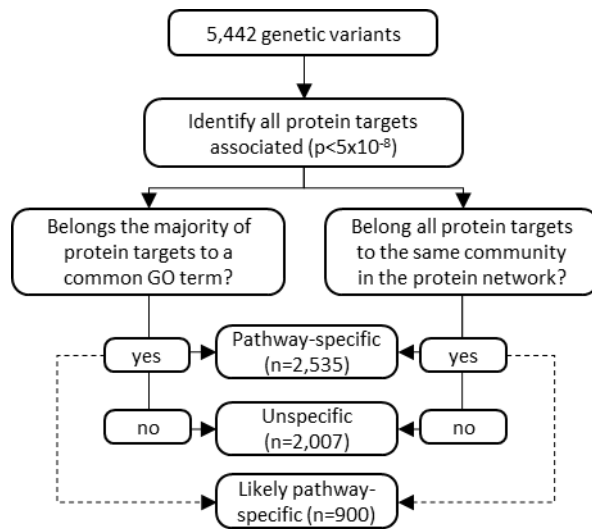
1384



1385

1386
 1387
 1388
 1389
 1390
 1391
 1392
 1393
 1394
 1395
 1396

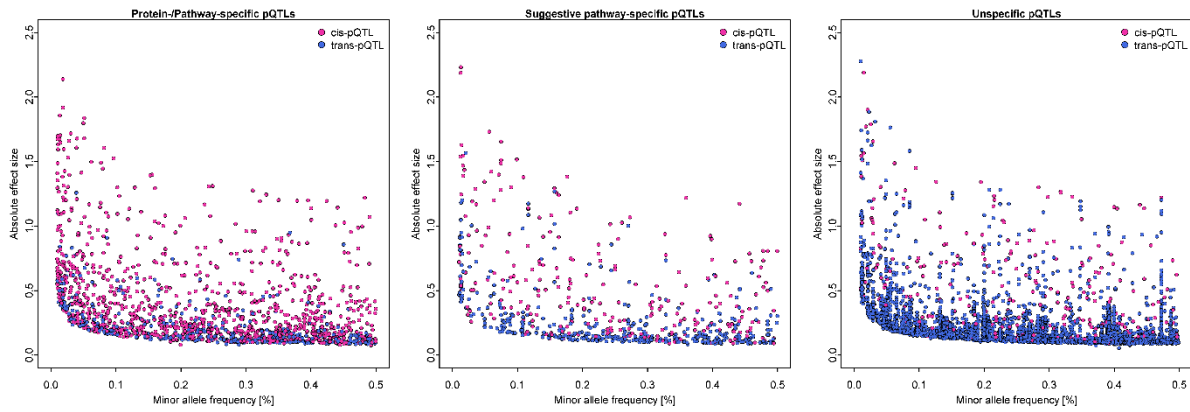
Fig. S1. Functional annotation of genetic variants associated with at least one protein target. **Left.** Bar chart of the distribution of annotations of 4,976 genetic variants associated with at least one protein target based on the Variant Effect Predictor tool. Colors indicate protein quantitative trait loci (pQTLs) either in close proximity to the protein-encoding gene (cis-pQTL, pink) or elsewhere in the genome (trans-pQTL, blue). Categories significantly enriched for cis-pQTLs are indicated with an asterisk (Fisher's-exact test, p-value < 0.003). The list of all 5,442 sentinel and secondary signals representing distinct variants in strong LD ($R^2 > 0.6$). **Right.** Fraction of cis-pQTLs (N=3,139) harboring a protein-altering variant (PAV) and further split based on the location of the PAV in any common protein domain.



1397

1398 **Fig. S2. Flowchart for pQTL classification.** For each of the 5,442 distinct genetic variants
 1399 we identified the set of significantly associated aptamers ($p < 5 \times 10^{-8}$) across the entire data. In
 1400 the next step we tested whether all of the targeted proteins fall within a GO term and/or belong
 1401 to the same community of protein targets in a data-driven protein network.

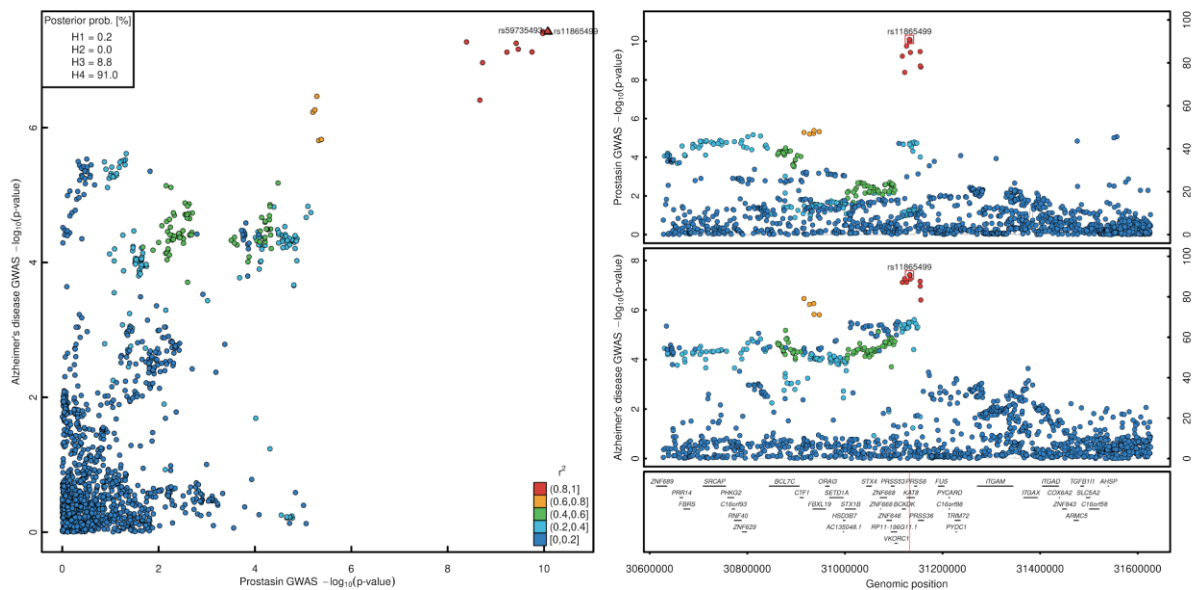
1402



1403

1404 **Fig. S3. Distribution of absolute effect sizes for identified protein quantitative trait loci**
 1405 **(pQTLs).** pQTLs are separated by specificity across the proteome and absolute effect sizes are
 1406 plotted against the minor allele frequency.

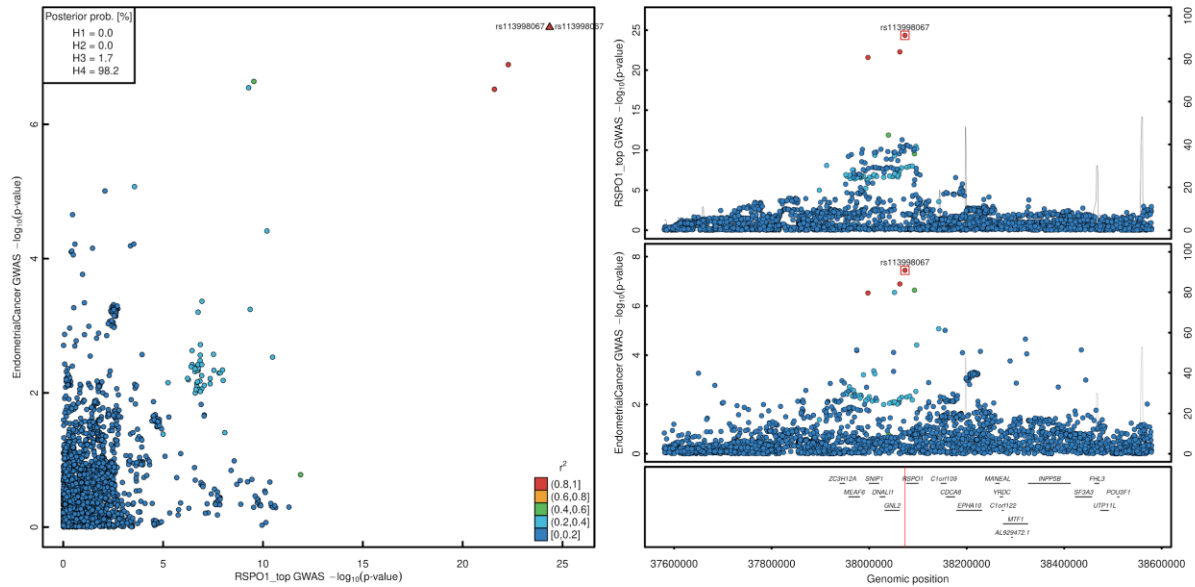
1407



1408

1409 **Fig. S4. Locusplot comparing association statistics for prostaticin and Alzheimer's disease.**
 1410 The left panel displays a comparison of $-\log_{10}$ -transformed p-values from GWAS summary
 1411 statistics for genetic variants in a 500 kb region around the lead signal on chromosome 16.
 1412 Coloring was done based on linkage disequilibrium with the lead variant for the protein. The
 1413 right panel is a stacked locuszoom plot with annotation of protein-encoding genes underneath.
 1414 Location of the lead variant is indicated by a red line.

1415



1416

1417 **Fig. S5. Locusplot comparing association statistics for RSPO1 and endometrial cancer.**

1418 The left panel displays a comparison of $-\log_{10}$ -transformed p-values from GWAS summary

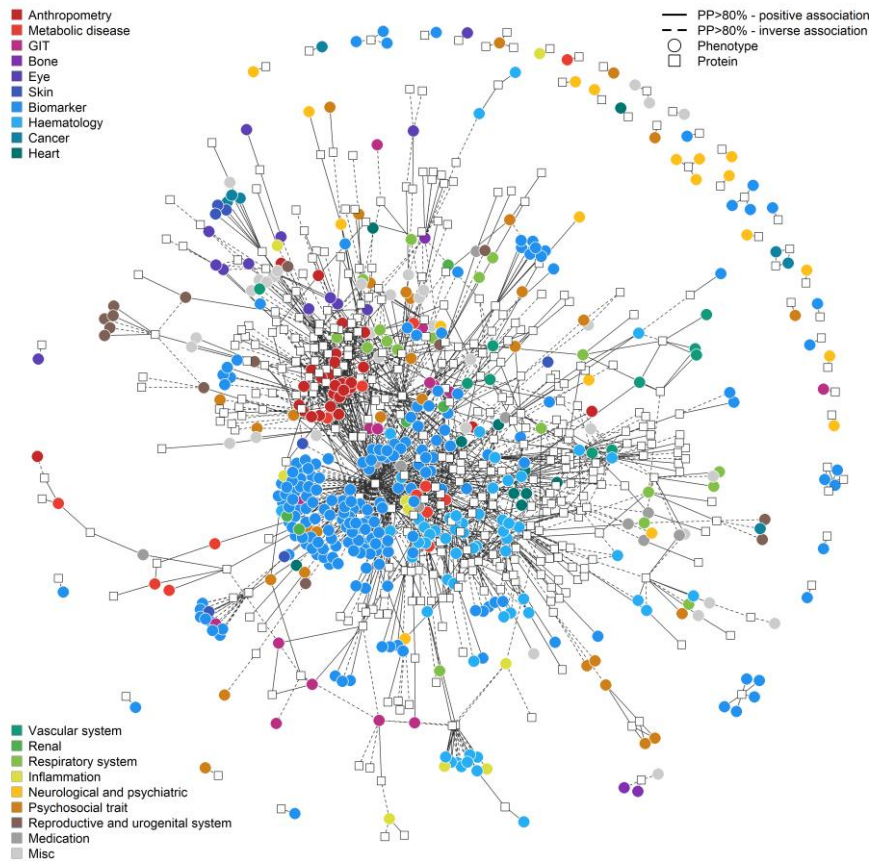
1419 statistics for genetic variants in a 500 kb region around the lead signal on chromosome 1.

1420 Coloring was done based on linkage disequilibrium with the lead variant for the protein. The

1421 right panel is a stacked locuszoom plot with annotation of protein-encoding genes underneath.

1422 Location of the lead variant is indicated by a red line.

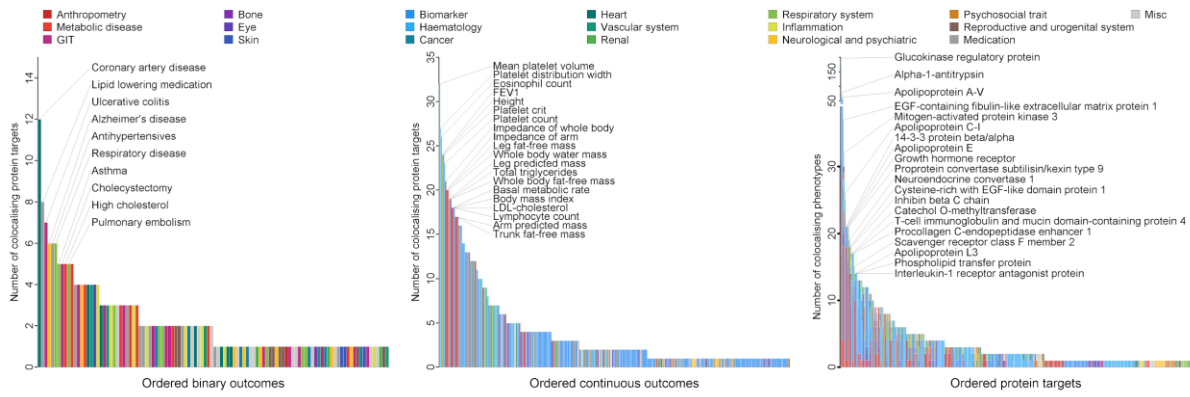
1423



1424

1425 **Fig. S6 Proteo-genomic map of human health.** The network is composed of 412 protein
 1426 targets (squares) and 506 phenotypes (circles) as nodes, which are connected (n=1,859 edges)
 1427 if there is evidence of a shared genetic signal (posterior probability >80%). Only protein targets
 1428 and phenotypes with at least one connection are included. Effect directions are indicated by the
 1429 line type aligned with the allele associated with higher amounts of the protein target (solid –
 1430 positive, dashed – inverse association with the phenotype). Colors indicate categories of
 1431 phenotypes. The inset represents the entire network, including continuous and binary
 1432 phenotypes, whereas the larger figure is restricted to binary phenotypes. An interactive version
 1433 of the figure can be found at www.omicscience.org/apps/pgwas.

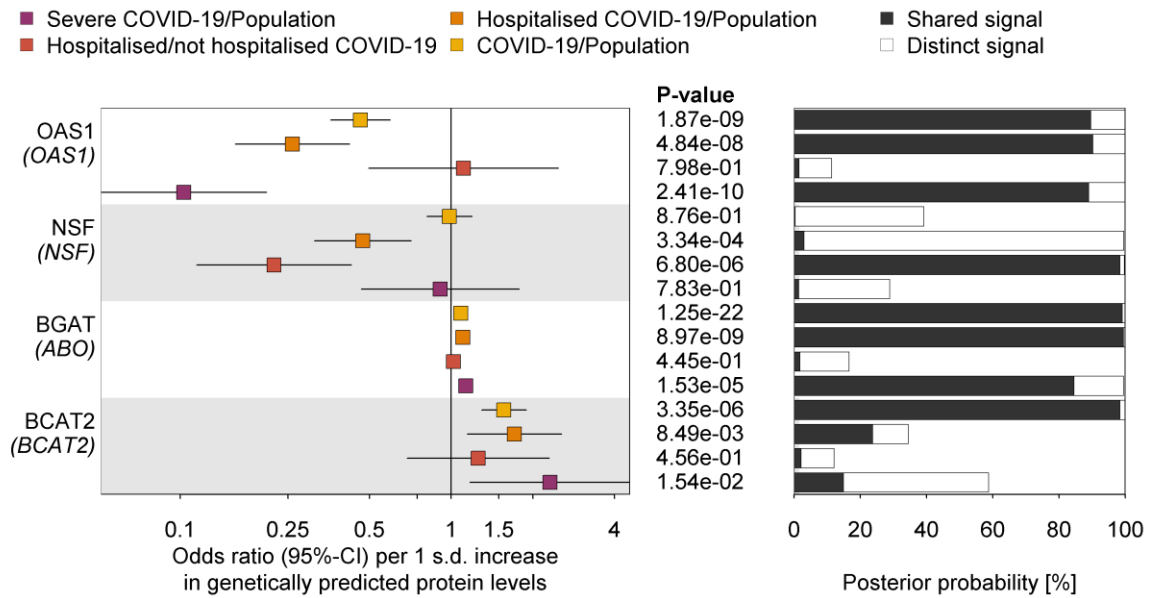
1434



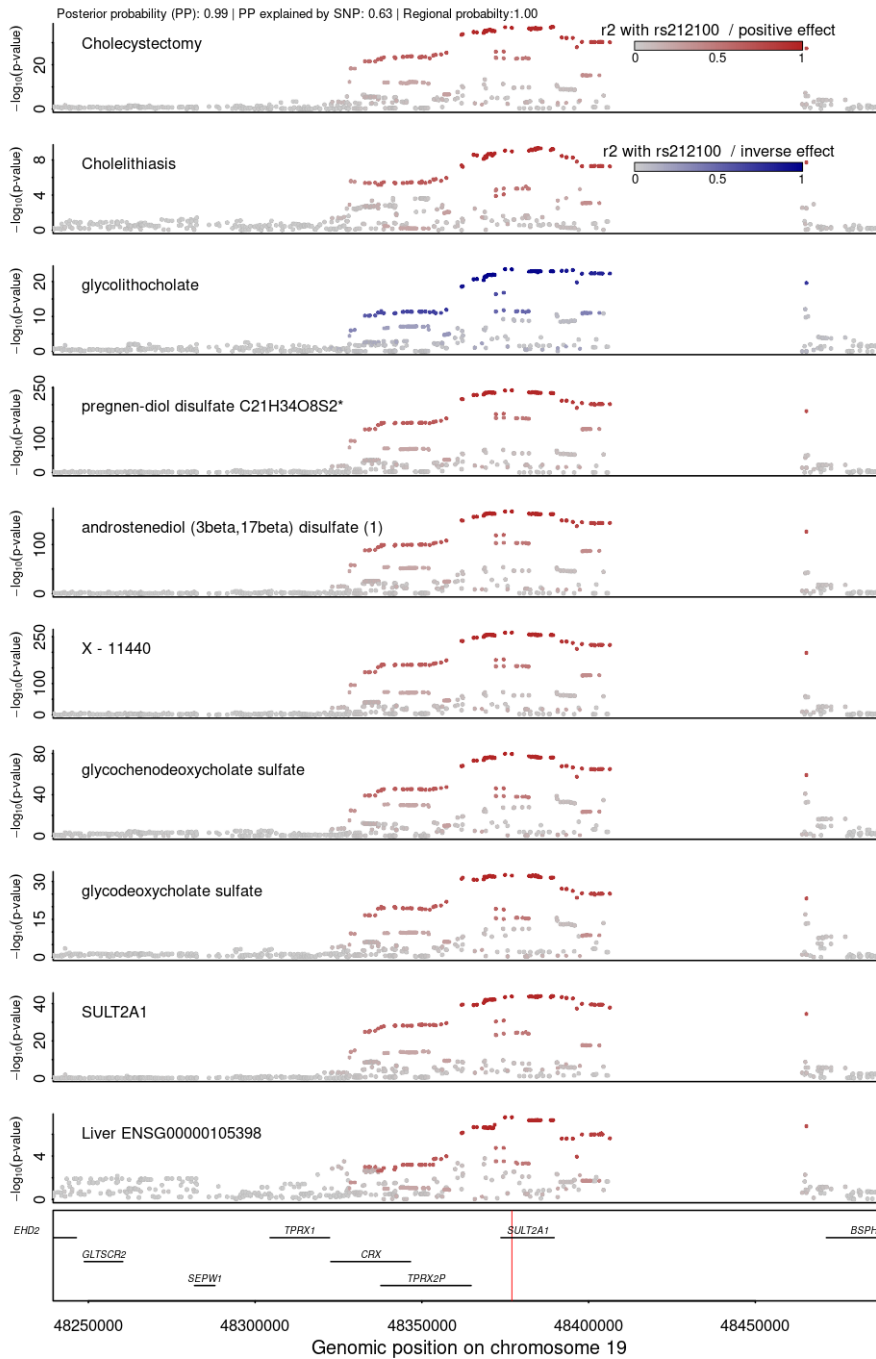
1435

1436 **Fig. S7 Summary of cis-region-based phenome-wide colocalization analysis.** The left panel
 1437 represents the number of colocalizing protein targets (posterior probability >80% for a shared
 1438 genetic signal) for each binary outcome, whereas the panel in the middle does the same for
 1439 continuous traits. Traits were ordered by the number of colocalizing protein targets, and top
 1440 traits are annotated. The right panel displays the number of colocalizing phenotypes for each
 1441 protein target, and stacked bar charts were used to indicate diversity of phenotypes based on
 1442 the categories indicated above the plots.

1443



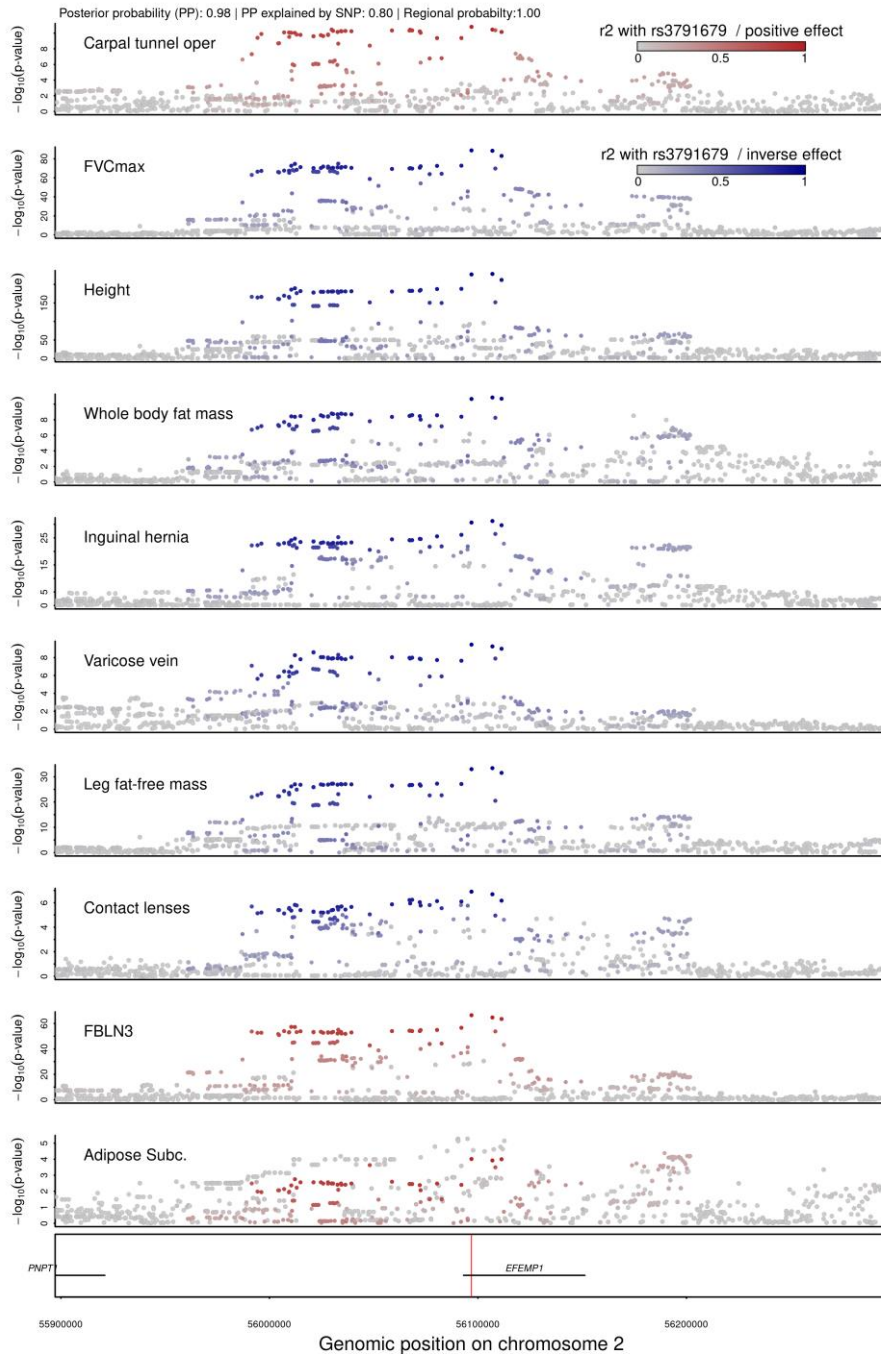
1444
 1445 **Fig. S8 Protein targets related to COVID-19.** Odds ratios and 95%-CIs for the genetically
 1446 predicted effect of protein levels on four different outcome definitions and control populations
 1447 for COVID-19 (left), including protein targets with strong evidence for statistical
 1448 colocalization for at least one definition (right). The column in the middle reports p-values.
 1449



1450

1451 **Fig. S9 Stacked regional association plots for phenotypes colocalizing with bile salt**
 1452 **sulfotransferase (SULT2A1).** Phenotypes are annotated and coloring is based on effect
 1453 directions aligned to the protein-increasing allele for rs212100 (red-positively, blue-inversely).
 1454 Darker colors indicate higher linkage disequilibrium with rs212100, and gray dots are below
 1455 0.1. Posterior and regional probabilities from multi-trait colocalization are given at the top of
 1456 the plot.

1457



1458

1459 **Fig. S10 Stacked regional association plots for phenotypes colocalizing with FBLN3.**

1460 Phenotypes are annotated and coloring is based on effect directions aligned to the protein-

1461 increasing allele for rs3791679 (red-positively, blue-inversely). Darker colors indicated higher

1462 linkage disequilibrium with rs3791679, and gray dots are below 0.1. Posterior and regional

1463 probabilities from multi-trait colocalization are given at the top of the plot.

1464

1465 **Table titles and legends**

1466
1467

1468 **Table S1. Demographics of the Fenland study population.**

1469 **Table S2. Summary of variant – protein target associations.** The table includes all lead and
1470 secondary signals across all 2,584 identified genomic regions, including summary statistics,
1471 functional annotations, variant classification, and gene assignments.

1472 **Table S3. Explained variance for 4,030 distinct aptamers targeting 3,892 proteins with**
1473 **at least one pQTL.** The table includes the amount of variance explained in protein
1474 abundances separately for each of the three classification criteria of pQTLs and contains
1475 further information on possible druggable targets.

1476 **Table S4. Integration of gene and splicing QTLs from the GTEx version 8 release.** For
1477 each protein target the strongest eQTL and/or sQTL are listed along with colocalization
1478 priority and possible tissue specificity is indicated.

1479 **Table S5. Summary of pQTL mapping to known GWAS loci.** For each mapping cis-pQTL
1480 – GWAS variant pair all curated traits are listed, and a column indicates whether the protein-
1481 encoding genes has been reported at this locus.

1482 **Table S6. Protein target – phenotype connections with strong evidence of colocalization**
1483 **at the protein encoding locus.** The table contains all protein – phenotype connections as
1484 shown in Figure 5 with further information on association statistics for the lead cis-pQTL.

1485 **Table S7. Results from pathway enrichment analysis.** For each curated phenotype with at
1486 least three associated proteins in the proteo-genomic map GO term enrichment was performed
1487 and results are presented collapsing pathways with the same gene set into one entry.

1488 **Table S8. Results from age- and sex-interaction analysis for cis-pQTLs.** The table lists for
1489 14 identified cis-pQTLs with a potential age- or sex-differential effect results from interaction
1490 testing and results within each stratum, that is, for each sex separately as well as for middle age
1491 and older individuals.

1492 **Table S9. Summary on identified druggable targets with potential for repurposing.**

1493
1494
1495