Unraveling Neuro-Proteogenomic Landscape and Therapeutic Implications for Human Behaviors and Psychiatric Disorders

Linda Repetto
University of Edinburgh
Jiantao Chen
State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University
Zhijian Yang
Sun Yat-sen University https://orcid.org/0000-0003-4803-8633
Ranran Zhai
Sun Yat-sen University https://orcid.org/0000-0002-5834-9120
Paul Timmers
University of Edinburgh https://orcid.org/0000-0002-5197-1267
Ting Li
Sun Yat-sen University https://orcid.org/0000-0002-9614-1566
Emma Twait
Utrecht University
Sebastian May-Wilson
Centre for Global Health Research, Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, EH8 9AG, UK https://orcid.org/0000-0003-2668-5717
Marisa Muckian
The University of Edinburgh
Bram Prins
University of Cambridge
Grace Png
Helmholtz Zentrum
Charles Kooperberg
Division of Public Health Sciences, Fred Hutchinson Cancer Center
Åsa Johansson
Uppsala University
Robert Hillary
The University of Edinburgh
Eleanor Wheeler
MRC Epidemiology Unit, University of Cambridge https://orcid.org/0000-0002-8616-6444

Lu Pan
Karolinska Institutet
Yazhou He
Sichuan University
Sofia Klasson
University of Gothenburg
Shahzad Ahmad
ErasmusMC
James Peters
Imperial College London
Arthur Gilly
Helmholtz Zentrum München
Maria Karaleftheri
Echinos Medical Centre
Emmanouil Tsafantakis
Anogia Medical Centre
Jeffrey Haessler
Fred Hutchinson Cancer Research Center
Ulf Gyllensten
Uppsala University
Sarah Harris
The University of Edinburgh https://orcid.org/0000-0002-4941-5106
Nicholas Wareham
University of Cambridge https://orcid.org/0000-0003-1422-2993
Andreas Göteson
University of Gothenburg
Cecilia Lagging
University of Gothenburg
M. Arfan Ikram
Erasmus University Medical Center https://orcid.org/0000-0003-0372-8585
Cornelia van Duijn
University of Oxford https://orcid.org/0000-0002-2374-9204
Christina Jern
University of Gothenburg
Mikael Landén
Gothenburg University https://orcid.org/0000-0002-4496-6451
Claudia Langenberg
Queen Mary University of London https://orcid.org/0000-0002-5017-7344
Ian Deary

University of Edinburgh https://orcid.org/0000-0002-1733-263X
Riccardo Marioni
University of Edinburgh https://orcid.org/0000-0003-4430-4260
Stefan Enroth
Uppasala University
Alexander Reiner
Fred Hutchinson Cancer Research Center https://orcid.org/0000-0002-1427-4470
George Dedoussis
Harokopio University of Athens
Eleftheria Zeggini
Helmholtz Zentrum München
Adam Butterworth
University of Cambridge
Anders Mälarstig
Karolinska Institutet
James Wilson
University of Edinburgh https://orcid.org/0000-0001-5751-9178
Pau Navarro
The University of Edinburgh
Xia Shen (xia.shen@ed.ac.uk)
University of Edinburgh https://orcid.org/0000-0003-4390-1979

Article

Keywords:

Posted Date: March 31st, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2720355/v1

Additional Declarations: Yes there is potential Competing Interest. P.R.H.J.T is a salaried employee of BioAge Labs, Inc. The remaining authors declare no competing financial interests. R.E.M has received a speaker fee from Illumina, is an advisor to the Epigenetic Clock Development Foundation, and a scientific consultant for Optima Partners. E.W. is now an employee of AstraZeneca.

Unraveling Neuro-Proteogenomic Landscape and Therapeutic Implications for Human Behaviors and Psychiatric Disorders

Linda Repetto[1,2,3,†], Jiantao Chen[2,3,4,†], Zhijian Yang[2,3,4,†], Ranran Zhai[2,3,4,†], Paul R. H. J. Timmers[1,5], Ting Li[2,3,4], Emma L. Twait[6], Sebastian May-Wilson[1], Marisa D. Muckian[1], Bram P. Prins[7], Grace Png[8,9], Charles Kooperberg[10], Åsa Johansson[11], Robert F. Hillary[12], Eleanor Wheeler[13], Lu Pan[14], Yazhou He[1], Sofia Klasson[15], Shahzad Ahmad [16], James E. Peters[17], Arthur Gilly[8], Maria Karaleftheri[18], Emmanouil Tsafantakis[19], Jeffrey Haessler[10], Ulf Gyllensten[11], Sarah E. Harris[20], Nicholas J. Wareham[13], Andreas Göteson[15], Cecilia Lagging[15,21], Mohammad Arfan Ikram [16], Cornelia M. van Duijn[16], Christina Jern[15], Mikael Landén[14,15], Claudia Langenberg[13,22], Ian J. Deary[20], Riccardo E. Marioni[12], Stefan Enroth[11], Alexander P. Reiner[23], George Dedoussis[24], Eleftheria Zeggini[8,25], Adam S. Butterworth[7,26,27,28,29], Anders Mälarstig[14,30], James F. Wilson[1,5,†], Pau Navarro[1,5,†], Xia Shen[1,2,3,4,14,†,*]
on behalf of the SCALLOP Consortium

[1]Centre for Global Health Research, Usher Institute, University of Edinburgh, Edinburgh, UK
[2]State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai, China
[3]Center for Intelligent Medicine Research, Greater Bay Area Institute of Precision Medicine (Guangzhou), Fudan University, Guangzhou, China
[4]Biostatistics Group, School of Life Sciences, Sun Yat-sen University, Guangzhou, China
[5]MRC Human Genetics Unit, MRC Institute of Genetics Molecular Medicine, University of Edinburgh, Edinburgh, UK
[6]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht and Utrecht University, Utrecht, Netherlands
[7]BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
[8]Institute of Translational Genomics, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany
[9]Technical University of Munich (TUM), School of Medicine, 81675 Munich, Germany
[10]Division of Public Health Sciences, Fred Hutchinson Cancer Center, Seattle USA
[11]Dept. Immunology, Genetics and Pathology, Science for life laboratory, Uppsala University, Sweden
[12]Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, The University of Edinburgh, Edinburgh, EH4 2XU, United Kingdom
[13]MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge School of Clinical Medicine, Cambridge, UK
[14]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
[15]Institute of Biomedicine, Department of Laboratory Medicine, the Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
[16]Department of Epidemiology, ErasmusMC, Rotterdam, The Netherlands
[17]Department of Immunology and Inflammation, Faculty of Medicine, Imperial College London, London, United Kingdom
[18]Echinos Medical Centre, Echinos, Greece
[19]Anogia Medical Centre, Anogia, Greece

[20]Lothian Birth Cohorts, University of Edinburgh, Edinburgh, EH8 9JZ, United Kingdom
[21]Region Västra Götaland, Sahlgrenska University Hospital, Department of Clinical Genetics and Genomics, Gothenburg, Sweden
[22]Computational Medicine, Berlin Institute of Health (BIH) at Charité – Universitätsmedizin Berlin, Germany
[23]Division of Public Health Sciences, Fred Hutchinson Cancer Center and Department of Epidemiology, University of Washington, Seattle USA
[24]Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University of Athens, Athens, Greece
[25]Technical University of Munich (TUM) and Klinikum Rechts der Isar, TUM School of Medicine, Munich, Germany
[26]British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, UK
[27]Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, UK
[28]National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge, Cambridge, UK
[29]National Institute for Health Research Cambridge Biomedical Research Centre, University of Cambridge and Cambridge University Hospitals, Cambridge, UK
[30]Emerging Science and Innovation, Pfizer Worldwide Research, Development and Medical, Cambridge, UK

*Correspondence should be addressed to: xia.shen@ed.ac.uk
†These authors contributed equally to this work.

Understanding the genetic basis of neuro-related proteins is essential for dissecting the molecular basis of human behavioral traits and the disease etiology of neuropsychiatric disorders. Here, the SCALLOP Consortium conducted a genome-wide association meta-analysis of over 12,500 individuals for 184 neuro-related proteins in human plasma. The analysis identified 117 cis-regulatory protein quantitative trait loci (cis-pQTL) and 166 trans-pQTL. The mapped pQTL capture on average 50% of each protein's heritability. Mendelian randomization analyses revealed multiple proteins showing potential causal effects on neuro-related traits such as sleeping, smoking, feelings, alcohol intake, mental health, and psychiatric disorders. Integrating with established drug information, we validated 13 out of 13 matched combinations of protein targets and diseases or side effects with available drugs, while suggesting hundreds of re-purposing and new therapeutic targets. This consortium effort provides a large-scale proteogenomic resource for biomedical research on human behaviors and other neuro-related phenotypes.

Certain patterns of human behaviors such as cigarette-smoking, alcohol consumption, and high fat may elevate the risk of developing a range of complex diseases[1,2]. While neuropsychiatric disorders are among the leading causes of life-long disability globally, affecting around 800 million people[3,4]. As of 2023, mental health remains a global crisis and priority brought to the forefront of public health discussions anew, after the impact of COVID-19 on people's lives, where stressors such as isolation, significant changes in habits, and global enhanced mortality and fear of contracting the disease have had severe consequences on mental well-being[5–7]. These conditions represent a significant challenge for medical research due to the high complexity of their neurobiological mechanisms and heterogeneity of symptoms which often overlap with other neurological, psychiatric, and non-psychiatric disorders[8–10].

In the past decade, genome-wide association studies (GWAS) have been successful in identifying numerous genetic variants that can partially account for variation in complex traits and diseases [11,12]. However, the effect of a genetic variant such as a single nucleotide polymorphism (SNP) on a complex disease is usually very small and often does not provide information on the phenotype's molecular architecture. Measuring proteins may overcome this obstacle as proteins are the product of translated DNA and functional elements that bridge the genetic codes and disease outcomes. Circulating proteins in blood plasma originate from various organ tissues and cell types in the human body and have fundamental roles in different biological processes[13–15]. Thus, such proteins are often used in clinical practice as disease biomarkers. Circulating neurology-related

proteins have the potential to provide insight into the pathophysiology of neurological and mental disorders and the genetic architecture of their molecular pathways, setting the basis for the improvement of diagnostic instruments and targeted therapy[16].

Protein levels are more linked to variation in cognitive function than genetic variants alone. Current studies on neurology-related proteins either focussed on neurodegenerative disorders or cognitive function specifically or had a limited sample size[17–22]. In a recent study, neurology-related proteins were associated with general fluid cognitive abilities in late life, and a portion of these was observed to be mediated by brain volume, measured as a structural brain variable [20].

The field of proteomics has been rapidly expanding in recent years and produced results that have played a fundamental role in the decoding process of molecular mechanisms involved in several traits and diseases, from cardiovascular disease to general health [19,23–26]. The genomic studies of the human proteome have benefited from various high-throughput measurement techniques, such as mass spectrometry[14,27], aptamer-based assays[28], and antibody-based assays[15]. Among these, the antibody-based Proximity Extension Assay [29] has high measurement precision, especially for many functional but low-abundant proteins.

This study aims to identify genetic variants associated with 184 neurology-related blood circulating proteins via a large-scale genome-wide association meta-analysis (GWAMA) and investigate the proteins' genetic and potential causal relationships with potential disease-causing behaviors, common psychiatric disorders, as well as related comorbidities. We systematically investigate the proteins' therapeutic implications based on established drug information. We provide an atlas for the genetic architecture of these proteins as a resource for biomedical research on human behaviors and psychiatric disorders.

Results

GWAMA identified 283 loci associated with 184 neuro-related proteins

In the discovery phase, we conducted a GWAMA using data from up to 12,176 individuals (mean age = 61.9, percentage females = 44.6%) for 92 proteins in the Olink©Neurology panel, and up to 5013 individuals (mean age = 49.6, percentage females = 56.1%, see Supplementary Tables 11-23 for details) for 92 proteins in the Olink©Neuro-Exploratory panel, from a total of twelve participating cohorts (Supplementary Tables 12-23). Overall, we identified 266 top variants distributed across a total of 117 cis-pQTL and 166 trans-pQTL with the significance threshold of $P < 5 \times 10^{-8}$ for the cis-loci and $P < 1.76 \times 10^{-10}$ for the trans-loci (Supplementary Table 1, Supplementary Fig. 7-8).

Out of the 137 proteins with detected pQTL, 68 proteins had significantly associated variants both in cis- and trans-regulatory loci.

As expected, the identified trans-pQTL, in general, were more weakly associated than the cis-pQTL, nevertheless, we found that 24 proteins shared a total of 14 trans-pQTL. For example, well-known pleiotropic loci such as the *HLA* region and the *ABO* locus showed trans-regulatory effects across a number of plasma proteins (**Fig. 1a**). For instance, 19 proteins showed significant trans-pQTL at the *ABO* locus, nevertheless, the associations were not completely due to the same causal variants (Supplementary Fig. 3). Most of the mapped pQTL were also found to be expression QTL (eQTL) significantly associated with the expressions of the corresponding/nearest genes, however, compared to trans-pQTL, cis-pQTL were much more likely to colocalize with eQTL, in terms of the underlying genetic regulation (Supplementary Fig. 1-2). The lead variants of the cis-pQTL were also more centered around the transcription start sites (TSS) of the corresponding coding genes, compared to those of the trans-pQTL around the TSS of the nearest coding genes (**Fig. 1b**). The cis-pQTL also had stronger effects, less correlated with the minor allele frequencies (MAFs), compared to the trans-pQTL (**Fig. 1c-d**).

The fact that the trans-pQTL were not colocalized with eQTL could be partly due to the weaker signals of the trans-pQTL than those of the cis-pQTL. However, we hypothesized that the trans-pQTL may not necessarily reflect the biological regulatory mechanisms of the corresponding proteins, but rather driven by underlying features of the blood samples, due to their influence on the immuno-reaction of the Olink assay. For example, the pleiotropic trans-pQTL across the proteins highlight major blood coagulation and clotting factors such as KLKB1 (Plasma kallikrein), KNG1 (Kininogen-1), and F12 (Coagulation factor XII), as well as glycosylation locus *ST3GAL4*. We thus also looked into the functional pathways and gene sets that involve the closest genes to our trans-pQTL, using the gene set enrichment analyses (Supplementary Fig. 6). With a false discovery rate $< 5\%$, 997 significant pathways were found to be enriched for the genes of our trans loci, of which 443 (44.4%) were driven or partly driven by the HLA genes. Most top enriched pathways were clustered into inflammatory and immune responses, coagulation processes, cell-to-cell signaling and adhesion, and protein glycosylation (Supplementary Table 8). Particularly, the trans-pQTL were found to be enriched in 1) established GWAS traits such as blood protein levels, platelet count, and platelet crit; 2) GO pathways such as biological adhesion, wound healing, coagulation, and glycosylation; 3) Hallmark gene sets including coagulation; 4) Reactome pathways including hemostasis and clotting formation; 5) microRNA targets and Wiki pathways for blood clotting cascade.

We assessed the overall heritabilities across the 184 analyzed plasma proteins. Methods based on summary association statistics have been developed to infer heritability and genetic correla-

tion parameters for complex traits with GWAS results; however, consistent estimates can only be obtained for genetic correlations [30–32]. Thus, we used a standard polygenic mixed model on the individual-level data collected in the ORCADES cohort to assess the narrow-sense heritability for each protein [33]. Across the analyzed proteins, we found that the higher the protein's heritability, the more pQTL detected for the protein (**Fig. 1e**), the stronger the cis-pQTL effects are (**Fig. 1g**), and the higher amount of phenotypic variance captured by the detected pQTL (**Fig. 1f**). On average, the mapped pQTL together explain 49% of the proteins' heritability. This indicates that proteins as molecular phenotypes have strong major regulatory loci. Nevertheless, their genetic effects can still be widespread across the genome, having a polygenic genetic architecture.

Using data from the ORCADES cohort, we found TDGF1 (Teratocarcinoma-Derived Growth Factor 1) to have the highest heritability ($h^2 = 0.85$), followed by MDGA1 (MAM Domain-Containing Glycosylphosphatidylinositol Anchor Protein 1, $h^2 = 0.75$), CLM1 (CD300 Molecule Like Family Member F, $h^2 = 0.72$), and LAIR2 (Leukocyte Associated Immunoglobulin Like Receptor 2, $h^2 = 0.70$). In contrast, CTF1 (Cardiotrophin 1), EPHA10 (Ephrin Type-A Receptor 10), GSTP1 (Glutathione S-Transferase Pi 1), HSP90B1 (Heat Shock Protein 90 Beta Family Member 1), IFI30 (Gamma-Interferon-Inducible Lysosomal Thiol Reductase), NDRG1 (N-Myc Downstream Regulated 1) and SFRP1 (Secreted Frizzled Related Protein 1) all had an estimated h^2 value close to 0, while having at least one pQTL.

We used the PhenoScanner pQTL database [34,35] to determine whether the pQTL sentinel variants or variants in linkage disequilibrium (LD) with them ($r^2 > 0.8$) that we identified had been previously found to be significantly associated with the corresponding proteins (Supplementary Table 2). 113 of our discovered loci were already discovered in previous studies. We also checked whether the hits from the meta-analysis were significant in the individual cohorts and observed that 73 of the sentinel variants were found to be statistically significant only in the meta-analysis. We also extracted the established associations between our mapped cis-pQTL and complex traits from the PhenoScanner database (Supplementary Table 3). At a 5% false discovery rate, 39 cis-pQTL showed significant association with both complex traits and other proteins (mostly based on an aptamer-based assay). We found that the level of pleiotropy at the protein level, i.e., being trans-pQTL for other proteins, is associated with the level of pleiotropy on the complex traits (Supplementary Fig. 4).

We performed linkage disequilibrium (LD) pruning ($r^2 < 0.001$) to identify secondary independent associations at the cis-pQTL. We identified a total of 769 additional variants across all the 117 proteins with cis-pQTL mapped (Supplementary Table 4).

This meta-analysis within our SCALLOP collaborative framework is a follow-up of a previous

study on the proteins from the Olink Neurology and Neuro-exploratory panels, where data were collected from the two Greek cohorts that we included in this study[36]. Our results replicated over 90% of the established loci, including the previous main discoveries of the cis-pQTL for CD33, GPNMB, and MSR1. Furthermore, we cross-referenced the significant loci discovered in the meta-analysis with the currently available pQTL data from the UK Biobank Pharma Proteomics Project (UKB-PPP)[37]. 114 proteins in our meta-analysis were also included in the UKB-PPP analysis. For these proteins, 91 out of the 102 cis-pQTL and 89 out of the 125 trans-pQTL were also reported in the UKB-PPP results (Supplementary Table 1).

Mendelian randomization analysis identifies plausible causal protein markers for neuro-related phenotypes

In order to make statements on potential causality from the proteins to complex traits and diseases, we focused on the genetic associations at the cis-pQTL, which provide strong and most likely valid genetic instruments in Mendelian randomization (MR) analysis. We first considered the 152 neuro-related traits whose GWAS summary statistics are available through LD-Hub[38] as the outcome data. We performed an inverse-variance weighted (IVW) two-sample MR analysis using the 886 LD-pruned genetic instruments across the 117 cis-pQTL on the 152 phenotypes. With a false discovery rate 5% threshold, we obtained 24 significant potential causal associations for 13 proteins on 22 traits, where three proteins are currently druggable targets (**Fig. 2**, Supplementary Table 5).

In order to control for false positive inference due to LD, we adopted the HEIDI (heterogeneity in dependent instruments)[39] test statistic to examine the colocalization between each pQTL and its association with the corresponding downstream outcome phenotypes. Nine out of the 24 plausible causal associations had colocalization support by HEIDI ($p > 0.05$) (**Fig. 2-3**, Supplementary Table 5). Among these, the single protein CDH6 showed a potential causal effect on neurological and behavioral traits including mood swings, miserableness, leg pain, smoking, and neuroticism, where the effect on smoking had a different direction compared to on the others. CTSC and LGALS8 were both plausible causal markers for alcohol intake but with opposite effects directions. CDH17 showed an positive effect on intelligence. DPEP1 showed a negative effect on napping, while as a druggable target it also showed a potential risk-increasing effect on schizophrenia.

Mendelian randomization analysis provides evidence for the proteins' causal effects on other complex diseases

Expanding our cis-pQTL-based MR analysis to a broader range of complex traits, we used the UK Biobank GWAS summary-level data for 4,085 phenotypes by the Neale's lab (see Data Availability) as the outcome data. We performed the same analysis procedure as above, and with a false discovery rate 5% threshold, the analysis yielded in 472 significant potential causal associations for 82 proteins on 221 traits. Among these discoveries, 59 were for 47 diseases with 33 plausible causal protein markers.

Again, we utilized the HEIDI test statistic to examine the colocalization between each pQTL and the disease genetic associations. 29 out of the 59 plausible causal associations with disease outcomes showed colocalization supported by HEIDI ($p > 0.05$) (**Fig. 4**, Supplementary Table 6), including 8 druggable protein targets and 14 new targets.

Except for the effect of TPPP3 (tubulin polymerization–promoting protein family member 3) on hypothyroidism/myxoedema, reverse generalized summary-statistics-based MR (GSMR)[40] did not show evidence for reverse causality of the other significant MR discoveries on the complex diseases. In general, the MR estimated odds ratios (FDR < 0.05) were found to be ranging from 0.49 to 2.48, consistent with previous studies evaluating the causal effects of blood circulating proteins on other complex traits [15,41].

Systematic analysis of established, re-purposing, and new drug targets

Based on the MR causal inference, we systematically investigated the protein markers in the DrugBank database (see Data Availability). There were 13 protein-trait combinations from the significant MR discoveries that matched established drugs. We found that for all the 13 established drug targets (**Fig. 5a-b**), the MR-inferred causal effects directions matched the corresponding targeting drugs' pharmacological effects (including side effects) (**Fig. 5c**). For instance, hyaluronic acid is a liver disease biomarker, the protein NCAN binds with hyaluronic acid thus reduces liver cirrhosis. Gemtuzumab ozogamicin is a monoclonal anti–CD33 antibody, reducing white blood cell count. Benralizumab is an antibody for IL5RA, treating eosinophilic asthma by affecting its causal effect on eosinophil counts. Overdosed acetaminophen increases the mean corpuscular volume and mean corpuscular haemoglobin, due to the insufficient enzyme activity of Glutathione S-transferase P (GSTP1).

Clenbuterol was used as a bronchodilator in the treatment of asthma patients. But it can cause long and short-term side effects, including hypertension. Our MR analysis showed that the increased

level of beta-nerve growth factor (beta-NGF), which could be caused by Clenbuterol, could lead to a higher risk of hypertension (**Fig. 5d**).

The MR analysis reveals that protein CTSS (cathepsin S) can increase platelets in the blood and reduce mean platelet volume. Fostamatinib can inhibit the protein CTSS, known as an approved medication for chronic immune thrombocytopenia (ITP) by inhibiting the spleen tyrosine kinase (SYK). It indicates that fostamatinib treats ITP via both protein SYK and CTSS (**Fig. 5e**).

Cilastatin is a dehydropeptidase 1 (DPEP1) inhibitor used to prevent degradation of imipenem, both were used together to treat infections. We found that inhibiting DPEP1 can increase the risk of high blood pressure, while decrease the risk of schizophrenia (**Fig. 5f**). This indicates clinical re-purposing potential of Cilastatin, and other DPEP1 inhibitors, as treatments for schizophrenia, though further investigations are needed.

Overall, besides the validated targets, we also identified 273 suggestive drug re-purposing target-disease pairs for 18 proteins (**Fig. 5a-b**, Supplementary Table 9). There already exist established drugs for these protein targets, making these drugs potentially useful upon further clinical trials. At last, 144 new target-disease combinations were suggested, based on our causal inference (Supplementary Table 10).

Discussion

We identified novel pQTL for 137 of 184 neuro-related proteins, provided insights into their molecular mechanisms and effects on complex diseases and traits, and highlighted useful therapeutic targets with established drugs. On average, we identified half of the genetic architecture underlying the concentration of these proteins. We provide a well powered genetic landscape for these proteins with large-scale summary-level data for future research.

Although the proteins were found to have small effects individually in the MR analysis, our results indicated that for most of the identified proteins, having low levels in plasma leads to a higher chance of having poorer health conditions (Supplementary Fig. 5). These conditions include both deterioration of mental health and related non-neurological comorbidities. Such results on the neuro-related proteins are consistent with the notion that psychiatric and neurological disorders are multi-factorial and not limited to the central nervous system, but rather are products of interactions among multiple systems within the organism[42–45]. The intertwining of neuropsychiatric, inflammatory, and cardiovascular disorders has long presented a challenge in clinical research due to the difficulties in discerning the relationships among them[46,47]. Our results suggest that these disorders may share molecular mechanisms and pathways and provide the basis for developing

new diagnostic tools and treatment strategies. We also reported a large number of drug re-purposing targets, suggesting the potential use of established drugs in new clinical trials for treatment of different symptoms and disorders.

Regarding the MR methodology, we found that the MR analysis with a single genetic instrument at the cis-pQTL tended to generate a stronger estimated causal effect (**Fig. 4**). This is partly due to power, as compared to multi-instrument MR, single-instrument MR tends to produce causal effects estimates with larger standard errors, so that only the results with large causal effects estimates could reach statistical significance. Thus it indicates: 1) Single genetic instrument analysis may be more prone to winner's curse, i.e., more likely to detect an overestimated effect on the outcome trait; 2) using multiple independent instruments within a locus may not only improve power but also control false discoveries due to overestimated effects in the outcome GWAS.

As expected, the mapped trans-pQTL did not show good colocalization with nearby genes, and they were enriched in blood clotting and coagulation pathways. For instance, a blood clotting factor *KLKB1* appeared to be a trans-regulatory hub for multiple proteins. We thus infer that some of the trans-pQTL discovered are not directly involved in the genetic mechanisms of the corresponding proteins, but rather they regulate blood characteristics that affect the performance of the antibody-based assays. This is an important discovery for biotechnological development in proteomics, suggesting that the features of the plasma samples could be non-negligible factors in circulating protein quantification.

This study significantly advances our understanding of the genetics of neuro-related proteins and provides new targets for drug discovery. The pQTL discovery and causal inference with disease outcomes can inform clinical studies to identify actionable drug targets and enable integration into multi-omics analyses. The UK Biobank Pharma Proteomics Project and more cohorts could provide additional insights through larger meta-analyses and replication analyses, potentially revealing secondary signals in the pQTL. The inclusion of cohorts with diverse ancestries could further elucidate pQTL alleles that are not sufficiently polymorphic in European populations, identifying distinct molecular mechanisms underlying complex diseases.

Methods

Proteins This study focussed on proteins from the Olink Neurology and Olink Neuro-exploratory panels. Circulating protein levels were quantified using Proximity Extension Assay technology, consisting of pairs of oligonucleotides-labelled antibodies to bind target proteins and hybridize to have their sequence extended and amplified through polymerase chain reaction (PCR). The level of

amplified DNA is then quantified by microfluidic qPCR[29].

Proteins were selected by a panel of experts to include protein biomarkers that are known to be associated with neurological disorders and conditions through existing literature. The functions of these proteins comprise axonal development, metabolism, immune response, and cell-to-cell communication. The proteins have been included in their respective panel on the basis of their observed involvement in neurological conditions and disorders, as well as the general performance of the assay.

Cohorts and data collection We obtained summary statistics from the GWAS analyses performed on the Olink Neurology proteins from 10 cohorts and the Olink Neuro-exploratory proteins from 6 cohorts. Cohorts comprised population-based and case-control studies. The summary statistics information for each cohort can be found in Supplementary Tables 11-25. The total sample size for the Neurology panel meta-analysis was 12,176, whereas the Neuro-exploratory panel meta-analysis included up to 5,013 individuals. The participating cohorts used whole-genome sequencing data or imputed data using the 1000 Genomes Project (Phase1 and Phase3) or the Haplotype Reference Consortium (HRC) as reference panels. An average of 14.5 million SNPs were tested per protein, and the lowest per-SNP filter imputation quality ranged from 0.4 to 0.3 depending on the cohort. Each cohort carried out quality control according to their study design, as reported in Supplementary Table 11.

Data below the Olink limit of detection (LOD) is calculated based on the negative controls included in each PCR run. Data below the LOD was available only for some cohorts participating in the meta-analysis. As the proteins were quantified at different times across cohorts, not all studies have data on all proteins in the two Olink panels.

Genome-wide association analysis of the proteins The Normalized Protein expression values (NPX), Olink's unit of protein abundance level on a log2 scale [29], were rank-based inverse normal transformed before running the per-protein GWAS analyses. Genotypic data were the allelic dosages resulting from imputation using the Haplotype reference consortium (HRC) or the 1000 genomes data as reference panel. Monomorphic SNPs were excluded. The genotype-phenotype association analysis was performed using regression models adjusting for sex, age, plate number, plate column, plate row, sample time in storage, season of sample collection, population structure (when appropriate), and other study-specific covariates.

Meta-analysis The summary association statistics from each participating cohort were uploaded through a secured FTP channel to the University of Edinburgh's ECDF Eddie Mark 3 cluster. The

meta-analysis was run per protein in METAL (version 2018-08-28)[48] using the inverse variance weighted method. We defined cis-pQTL to be 500kb upstream or downstream of the gene coding for the respective protein and set the trans-pQTL window to be 1Mb around the top variants that were found outside the defined cis- window. A 1% MAF filter was applied to the meta-analysis summary statistics for subsequent analyses. The variants that existed in only one participating cohort were also removed before subsequent analyses. The significance threshold was set to be 5×10^{-8} for the top variants of cis-regulatory variants and $5 \times 10^{-8}/184 = 2.73 \times 10^{-10}$ for the variants in trans-regions.

Heritability analysis We used a standard polygenic mixed model implemented in GenABEL[33] on the individual-level data collected in the ORCADES cohort to assess the narrow-sense heritability for each protein. The heritability captured by each pQTL is calculated as $2f(1-f)\hat{\beta}^2$, where f and $\hat{\beta}$ are the coding allele frequency and estimated genetic effect, respectively, assuming Hardy-Weinberg equilibrium.

Established genetic associations We used PhenoScanner v2[34,35] to cross-reference the lead (most significant) genetic variants in the cis-pQTL from our meta-analysis with other phenotypes. PhenoScanner is an extensive database of over 65 billion associations from publicly available GWAS studies. We used the lead variants of our cis-loci as input without the additional option of using proxy markers. When checking the novelty of our mapped cis-pQTL, we consider established pQTL associations with $P < 5 \times 10^{-6}$ as known. When extracting the established complex traits associations, we set the p-value threshold to 1 to include all possible associations. Thereafter, results with false discovery rate less than 0.05 are considered. We excluded the studies with non-European ancestry.

Cross-referencing with other Olink-based pQTL studies We cross-referenced the discovered pQTL with results from the two Greek cohorts that we included in this study[36] and those reported by the UK Biobank Pharma Proteomics Project (UKB-PPP)[37]. For each cis-pQTL, we checked whether a cis-pQTL was also reported for the same protein in either one of the two pQTL studies. For each trans-pQTL, we checked whether a trans-pQTL was reported within a $\pm$500Kb window of the lead variant of our discovered trans-pQTL.

Gene set enrichment and functional annotation of GWAS trans loci We performed our gene set enrichment analyses using the GENE2FUNC in FUMA v1.3.7 [49,50], which returns functional annotation to ENSEMBL v92 gene models for the submitted list in a biological context. We identified the genes closest to the top SNPs in our trans loci using the locuszoom v0.12[51,52] database and then submitted the list of genes to the FUMA website. We selected all types of genes to use as background for

this analysis, including over 57,000 genetic elements. We set the maximum FDR adjusted p-value for gene set association to 1.

Mendelian randomization analysis We performed a two-sample Mendelian randomization (MR) analysis using the inverse-variance weighted (IVW) method to evaluate causal effects between the proteins with genome-wide significant cis-pQTL and the traits from the UK Biobank GWAS results by the Neale's lab. Multiple sentinel variants of our cis-pQTL after LD pruning ($r^2 < 0.001$) were used jointly as instrumental variables. We report the significant discoveries at a level of 5% false discovery rate, for which we also performed a reverse generalized summary-statistics-based MR (GSMR) from the complex trait exposures to protein outcomes.

Colocalization analysis For the MR-positive discoveries, the pQTL-complex-trait colocalization analysis was performed using the SMR/HEIDI tool in the GCTA software[39]. We considered a pair of QTL associations to be colocalized if the HEIDI test p-value was greater than 0.05.

For eQTL-pQTL colocalization analysis, we adopted the v7 release of both the GTEx eQTL and eQTLGen summary-level data. We used the Bayesian colocalization analysis tool `coloc`, with the posterior probabilities testing the H4 colocalization hypothesis, which tests for one shared variant between the pair of corresponding eQTL and pQTL[53]. For each cis-pQTL, we tested colocalization with the cis-eQTL of the corresponding coding gene in each tissue. For each trans-pQTL, we tested colocalization with the cis-eQTL of the nearest coding gene.

Drug target investigation For the protein markers from IVW MR results with false discovery rate less than 5%, we systematically investigated available drugs targetting these markers using the DrugBank database. We considered a drug target validated if an MR discovery between the protein marker and the trait/disease suggested the same effect direction as the drug's effect on the protein target. The protein targets that have available drugs but not directly related to the MR discovered outcomes were regarded as re-purposing targets. The remaining MR discoveries were reported as new targets.

Code availability

METAL: https://genome.sph.umich.edu/wiki/METAL_Documentation; PLINK: https://www.cog-genomics.org/plink/; GCTA-GSMR: https://yanglab.westlake.edu.cn/software/gcta/#GSMR; PhenoScanner: http://www.phenoscanner.medschl.cam.ac.uk; SMR & HEIDI: https://yanglab.westlake.edu.cn/software/smr/#SMR&HEIDIanalysis; FUMA: https://fuma.ctglab.nl.

Data availability

The full genome-wide summary association statistics for the 184 proteins will be made publicly available **upon publication of the paper**; GTEx data: https://gtexportal.org/home/datasets; 1000 Genomes phase 3 genotype data: https://www.cog-genomics.org/plink/2.0/resources#phase3_1kg; Neale's lab UK Biobank round2 GWAS summary-level data: http://www.nealelab.is/uk-biobank; DrugBank: https://www.drugbank.com.

Acknowledgements

X.S. was in receipt of Swedish Research Council (Vetenskapsrådet) grants (No. 2017-02543 & No. 2022-01309), a National Natural Science Foundation of China (NSFC) grant (No. 12171495), a Natural Science Foundation of Guangdong Province grant (No. 2114050001435), and a National Key Research and Development Program grant (No. 2022YFF1202105). P.R.H.J.T. and J.F.W. acknowledge support from the Medical Research Council Human Genetics Unit program grant "Quantitative Traits in Health and Disease" (U. MC_UU_00007/10). The work from C.K. and A.P.R. is supported in part by NIH grant R01-HL136574.

We thank the members of the cited consortia of genome-wide association studies for making their data available. Cohort-specific acknowledgements are given in the **Supplementary Information**.

Author contributions

X.S., P.N., and J.F.W. initiated and coordinated the study. L.R., J.C., Z.Y., and R.Z. performed data analysis. L.R., P.R.H.J.T., E.L.T., P.N., and X.S. contributed to the analysis pipeline. S.M.W., M.D.M., B.P.P., A.J., R.F.H., E.W., S.K., S.A., L.P., S.B., Y.H., G.P., C.K., J.E.P., U.G., S.E.H., N.J.W., C.L., M.A.T., A.G., A.G., M.K., E.T., J.H., A.P.R., G.D., E.Z., M.L., C.M.V.D., C.J., C.L., I.J.D., R.E.M., S.E., A.S.B., and A.M. contributed to the cohort-level analysis. L.R., J.C., Z.Y., R.Z., P.N., and X.S. wrote the paper. All authors approved the submitted version of the paper.

Competing interests statement

P.R.H.J.T is a salaried employee of BioAge Labs, Inc. The remaining authors declare no competing financial interests. R.E.M has received a speaker fee from Illumina, is an advisor to the Epigenetic Clock Development Foundation, and a scientific consultant for Optima Partners. E.W. is now an employee of AstraZeneca.

References

[1] Danaei, G. *et al.* The preventable causes of death in the united states: Comparative risk assessment of dietary, lifestyle, and metabolic risk factors. *PLOS Medicine* **6**, 1–23 (2009). URL https://doi.org/10.1371/journal.pmed.1000058.

[2] Wang, X. *et al.* Fruit and vegetable consumption and mortality from all causes, cardiovascular disease, and cancer: systematic review and dose-response meta-analysis of prospective cohort studies. *BMJ* **349** (2014). URL https://www.bmj.com/content/349/bmj.g4490. https://www.bmj.com/content/349/bmj.g4490.full.pdf.

[3] WHO. Who | mental disorders (2019). URL https://www.who.int/news-room/fact-sheets/detail/mental-disorders.

[4] Ritchie, H. & Roser, M. Mental health (2020). URL https://ourworldindata.org/mental-health.

[5] Hossain, M. M. *et al.* Epidemiology of mental health problems in covid-19: a review. *F1000Research* **9** (2020). URL /pmc/articles/PMC7549174//pmc/articles/PMC7549174/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7549174/.

[6] Greenberg, N. Mental health of health-care workers in the covid-19 era. *Nature Reviews Nephrology 2020 16:8* **16**, 425–426 (2020). URL https://www.nature.com/articles/s41581-020-0314-5.

[7] Jones, E. A., Mitra, A. K. & Bhuiyan, A. R. Impact of covid-19 on mental health in adolescents: A systematic review. *International Journal of Environmental Research and Public Health 2021, Vol. 18, Page 2470* **18**, 2470 (2021). URL https://www.mdpi.com/1660-4601/18/5/2470/htmhttps://www.mdpi.com/1660-4601/18/5/2470.

[8] Bearden, C. E., Reus, V. I. & Freimer, N. B. Why genetic investigation of psychiatric disorders is so difficult. *Current Opinion in Genetics Development* **14**, 280–286 (2004).

[9] Sullivan, P. F. & Geschwind, D. H. Defining the genetic, genomic, cellular, and diagnostic architectures of psychiatric disorders. *Cell* **177**, 162–183 (2019). URL https://doi.org/10.1016/j.cell.2019.01.015.

[10] Taylor, M. J. *et al.* Association of genetic risk factors for psychiatric disorders and traits of these disorders in a swedish population twin sample. *JAMA Psychiatry* **76**, 280–289 (2019). URL https://jamanetwork.com/journals/jamapsychiatry/fullarticle/2718628.

[11] Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of gwas discovery. *American journal of human genetics* **90**, 7–24 (2012). URL https://pubmed.ncbi.nlm.nih.gov/22243964/.

[12] Visscher, P. M. *et al.* 10 years of gwas discovery: Biology, function, and translation. *The American Journal of Human Genetics* **101**, 5–22 (2017). URL http://dx.doi.org/10.1016/j.ajhg.2017.06.005.

[13] Chames, P., Regenmortel, M. V., Weiss, E. & Baty, D. Themed section: Vector design and drug delivery review therapeutic antibodies: successes, limitations and hopes for the future. *British Journal of Pharmacology* **157**, 220–233 (2009). URL http://www3.interscience.wiley.com/journal/121548564/issueyear?year=2009.

[14] Solomon, T. *et al.* Identification of common and rare genetic variation associated with plasma protein levels using whole exome sequencing and mass spectrometry. *Circulation. Genomic and precision medicine* **11**, e002170 (2018). URL /pmc/articles/PMC6301071//pmc/articles/PMC6301071/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6301071/.

[15] Folkersen, L. *et al.* Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nature Metabolism* **2**, 1135–1148 (2020). URL http://dx.doi.org/10.1038/s42255-020-00287-2.

[16] Westwood, S. *et al.* Plasma protein biomarkers for the prediction of csf amyloid and tau and [18f]-flutemetamol pet scan result. *Frontiers in Aging Neuroscience* **10**, 409 (2018).

[17] Dencker, M., Björgell, O. & Hlebowicz, J. Effect of food intake on 92 neurological biomarkers in plasma. *Brain and Behavior* **7**, e00747 (2017). URL http://doi.wiley.com/10.1002/brb3.747.

[18] Jabbari, E. *et al.* Proximity extension assay testing reveals novel diagnostic biomarkers of atypical parkinsonian syndromes. *Journal of Neurology, Neurosurgery and Psychiatry* **90**, 768–773 (2019).

[19] Hillary, R. F. *et al.* Genome and epigenome wide studies of neurological protein biomarkers in the lothian birth cohort 1936. *Nature Communications* **10**, 3160 (2019). URL http://www.nature.com/articles/s41467-019-11177-x.

[20] Harris, S. E. *et al.* Neurology-related protein biomarkers are associated with cognitive ability and brain volume in older age. *Nature Communications* (2020). URL https://doi.org/10.1038/s41467-019-14161-7.

[21] Rodrigues-Amorim, D. *et al.* Plasma β-iii tubulin, neurofilament light chain and glial fibrillary acidic protein are associated with neurodegeneration and progression in schizophrenia. *Scientific Reports 2020 10:1* **10**, 1–10 (2020). URL https://www.nature.com/articles/s41598-020-71060-4.

[22] Sandberg, J. V. *et al.* Proteins associated with future suicide attempts in bipolar disorder: A large-scale biomarker discovery study. *Molecular Psychiatry* **27**, 3857–3863 (2022). URL https://doi.org/10.1038/s41380-022-01648-x.

[23] Folkersen, L. *et al.* Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. *PLOS Genetics* **13**, e1006706 (2017). URL https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006706.

[24] Williams, S. A. *et al.* Plasma protein patterns as comprehensive indicators of health. *Nature Medicine 2019 25:12* **25**, 1851–1857 (2019). URL https://www.nature.com/articles/s41591-019-0665-2.

[25] Lehallier, B. *et al.* Undulating changes in human plasma proteome profiles across the lifespan. *Nature Medicine 2019 25:12* **25**, 1843–1850 (2019). URL https://www.nature.com/articles/s41591-019-0673-2.

[26] Wingo, A. P. *et al.* Integrating human brain proteomes with genome-wide association data implicates new proteins in alzheimer's disease pathogenesis. *Nature Genetics 2021 53:2* **53**, 143–146 (2021). URL https://www.nature.com/articles/s41588-020-00773-z.

[27] Jensen, S. B. *et al.* Discovery of novel plasma biomarkers for future incident venous thromboembolism by untargeted synchronous precursor selection mass spectrometry proteomics. *Journal of Thrombosis and Haemostasis* **16**, 1763 (2018). URL /pmc/articles/PMC6123273//pmc/articles/PMC6123273/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6123273/.

[28] Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).

[29] Assarsson, E. *et al.* Homogenous 96-plex pea immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLOS ONE* **9**, e95192 (2014). URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0095192.

[30] Bulik-Sullivan, B. *et al.* Ld score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics 2015 47:3* **47**, 291–295 (2015). URL https://www.nature.com/articles/ng.3211.

[31] Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nature Genetics 2015 47:11* **47**, 1236–1241 (2015). URL https://www.nature.com/articles/ng.3406.

[32] Ning, Z., Pawitan, Y. & Shen, X. High-definition likelihood inference of genetic correlations across human complex traits. *Nature Genetics* **52**, 859–864 (2020). URL https://pubmed.ncbi.nlm.nih.gov/32601477/.

[33] Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. Genabel: an r library for genome-wide association analysis. *Bioinformatics* **23**, 1294–1296 (2007). URL https://academic.oup.com/bioinformatics/article/23/10/1294/198080.

[34] Staley, J. R. *et al.* Phenoscanner: a database of human genotype-phenotype associations. *Bioinformatics (Oxford, England)* **32**, 3207–3209 (2016). URL https://pubmed.ncbi.nlm.nih.gov/27318201/.

[35] Kamat, M. A. *et al.* Phenoscanner v2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics (Oxford, England)* **35**, 4851–4853 (2019). URL https://pubmed.ncbi.nlm.nih.gov/31233103/.

[36] Png, G. *et al.* Mapping the serum proteome to neurological diseases using whole genome sequencing. *Nature Communications 2021 12:1* **12**, 1–12 (2021). URL https://www.nature.com/articles/s41467-021-27387-1.

[37] Sun, B. B. *et al.* Genetic regulation of the human plasma proteome in 54,306 uk biobank participants. *bioRxiv* **20**, 2022.06.17.496443 (2022). URL https://www.biorxiv.org/content/10.1101/2022.06.17.496443v1https://www.biorxiv.org/content/10.1101/2022.06.17.496443v1.abstract.

[38] Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2016).

[39] Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature genetics* **48**, 481–487 (2016). URL http://eutils.ncbi.nlm.

nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=27019110\TU\textbackslash&
retmode=ref\TU\textbackslash&cmd=prlinks.

[40] Zhu, Z. *et al.* Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nature Communications* **9**, 224 (2018).

[41] Bretherick, A. D. *et al.* Proteome-by-phenome mendelian randomisation detects 38 proteins with causal roles in human diseases and traits. *bioRxiv* 631747 (2019). URL https://www.biorxiv.org/content/10.1101/631747v1.

[42] Knardahl, S. Cardiovascular psychophysiology. *Annals of Medicine* **32**, 329–335 (2000). URL https://www.tandfonline.com/action/journalInformation?journalCode=iann20.

[43] Ioannidis, K., Askelund, A. D., Kievit, R. A. & Harmelen, A. L. V. The complex neurobiology of resilient functioning after childhood maltreatment. *BMC Medicine* **18**, 1–16 (2020). URL https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-020-1490-7.

[44] McLaughlin, K. A., Colich, N. L., Rodman, A. M. & Weissman, D. G. Mechanisms linking childhood trauma exposure and psychopathology: a transdiagnostic model of risk and resilience. *BMC Medicine 2020 18:1* **18**, 1–11 (2020). URL https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-020-01561-6.

[45] Fried, E. I. & Robinaugh, D. J. Systems all the way down: Embracing complexity in mental health research. *BMC Medicine* **18**, 1–4 (2020). URL https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-020-01668-w.

[46] Fleshner, M., Frank, M. & Maier, S. F. Danger signals and inflammasomes: Stress-evoked sterile inflammation in mood disorders. *Neuropsychopharmacology 2017 42:1* **42**, 36–45 (2016). URL https://www.nature.com/articles/npp2016125.

[47] Bauer, M. E. & Teixeira, A. L. Inflammation in psychiatric disorders: What comes first? *Annals of the New York Academy of Sciences* **1437**, 57–67 (2019).

[48] Willer, C. J., Li, Y. & Abecasis, G. R. Metal: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190 (2010). URL /pmc/articles/PMC2922887//pmc/articles/PMC2922887/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC2922887/.

[49] Watanabe, K., Taskesen, E., Bochoven, A. V. & Posthuma, D. Functional mapping and annotation of genetic associations with fuma. *Nature Communications* **8**, 1826 (2017). URL http://www.nature.com/articles/s41467-017-01261-5.

[50] Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Fuma: Functional mapping and annotation of genetic associations. *European Neuropsychopharmacology* **29**, S789–S790 (2019).

[51] Pruim, R. J. *et al.* Locuszoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336 (2010). URL /pmc/articles/PMC2935401//pmc/articles/PMC2935401/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC2935401/.

[52] Boughton, A. P. *et al.* Locuszoom.js: interactive and embeddable visualization of genetic association study results. *Bioinformatics* **37**, 3017–3018 (2021). URL https://academic.oup.com/bioinformatics/article/37/18/3017/6178278.

[53] Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genetics* **10**, e1004383 (2014). 1305.4022.

Legends to Main Figures

Figure 1: Overview of the mapped protein quantitative trait loci (pQTL). **a.** Pleiotropic trans-pQTL counts and overlap of the mapped pQTL with existing eQTL. The upper barplot shows the number of proteins share trans-pQTL (gene annotations based on gene closest to the trans-pQTL). The scatterplot shows the genomic location of significant cis-pQTL in red ($P < 5 \times 10^{-8}$), significant trans-pQTL in blue ($P < 5 \times 10^{-8}/184$), and the shading within the dots indicates significance of the corresponding/nearest cis-eQTL for the respective protein. **b.** Scatterplot of the pQTL lead variants association signals v.s. their distance to the transcription start site (TSS) of the corresponding/nearest coding genes. **c.** Scatterplot of the absolute estimated genetic effects of the pQTL lead variants v.s. their minor allele frequencies (MAFs). **d.** The scatterplot in **c** shown in logarithm scale. **e.** Number of mapped pQTL per protein v.s. the linear mixed model estimated heritability in the ORCADES cohort. **f.** The variance explained by the mapped pQTL summed up for each protein v.s. the estimated heritability. **g.** For the proteins with significant cis-pQTL mapped, the lead variant signal strength v.s. the estimated heritability of each protein.

Figure 2: Causality between the proteins and neuro-related phenotypes inferred by Mendelian randomization (MR) analyses. The forest plot shows the significant MR results (false discovery rate < 0.05) based on LD-pruned ($r^2 < 0.001$) instrumental variants within each cis-pQTL. Inverse-variance weighted (IVW) estimates are provided as the solid round dots, and the whiskers indicate 95% confidence intervals. The numbers of instrumental variants in the cis-pQTL are given to the right of the whiskers. As a colocalization measure, the HEIDI (heterogeneity in dependent instruments) test evidence ($p > 0.05$) are given as the diamonds, where the largest diamonds correspond to a p-value of 1. The upper part of the plot shows the results where the proteins are known druggable targets, while the lower part shows the results for new protein targets.

Figure 3: Regional association patterns of the pQTL and the colocalized neuro-related complex traits. The displayed protein-trait pairs correspond to the Mendelian randomization discoveries in Figure 2 with the HEIDI p-value > 0.05. Each subfigure shows the pQTL region of 1Mb centered at the lead variant. The vertical dashed line in each subfigure marks the transcription start site of the corresponding protein's coding gene.

Figure 4: Causality between the proteins and UK Biobank disease phenotypes inferred by Mendelian randomization (MR) analyses. The forest plot shows the significant MR results (false discovery rate < 0.05) based on LD-pruned ($r^2 < 0.001$) instrumental variants within each cis-pQTL. Inverse-variance weighted (IVW) estimates are provided as the solid round dots, and the whiskers indicate 95% confidence intervals. The numbers of instrumental variants in the cis-pQTL are given to the right of the whiskers. As a colocalization measure, the HEIDI (heterogeneity in dependent instruments) test evidence ($p > 0.05$) are given as the diamonds, where the largest diamonds correspond to a p-value of 1. The upper part of the plot shows the results where the proteins are known druggable targets, while the lower part shows the results for new protein targets.

Figure 5: Drug targets revealed by Mendelian randomization (MR) analyses. The MR results with 5% false discovery rate are considered. **a.** The number of MR inferred pairs of proteins and traits split into four categories: new (drug) targets, druggable targets that have drugs with unclear clinical function, re-purposing targets that have established drugs but for different diseases, and validated known targets where the established drugs have pharmacological effects that match the MR results. **b.** Numbers of re-purposing and validated drug targets per protein analysed. **c.** The validated known drug targets, the description of the drugs, and the corresponding consistent MR estimated effects. **d.** Potential mechanism of the adverse effect of Clenbuterol that targets NGF. **e.** Potential mechanism of Fostamatinib treating Chronic immune thrombocytopenia through CTSS. **f.** Potential pharmacology of DPEP1's re-purposing drug on schizophrenia.

Figures

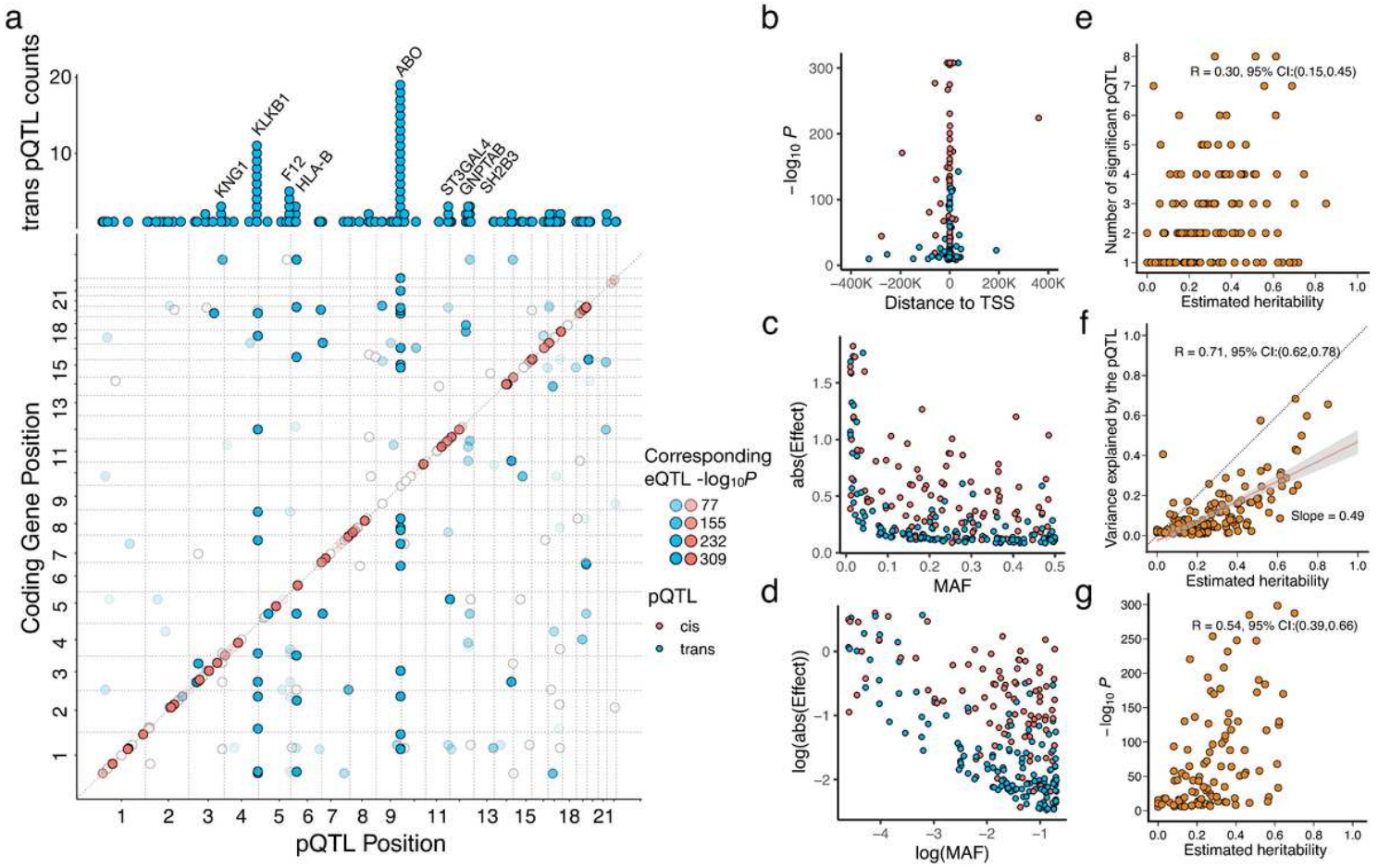

Figure 1

Overview of the mapped protein quantitative trait loci (pQTL). a. Pleiotropic trans pQTL counts and overlap of the mapped pQTL with existing eQTL. The upper barplot shows the number of proteins share trans-pQTL (gene annotations based on gene closest to the trans-pQTL). The scatterplot shows the genomiclocation of significant cis-pQTLin red (P < 5×10−8), significant trans-pQTLin blue (P< 5×10−8 /184), and the shading within the dots indicates significance of the corresponding/nearest cis-eQTL for the respective protein. b. Scatterplot of the pQTL lead vari ants association signals v.s. their distance to the transcription start site (TSS) of the corresponding/nearest coding genes. c. Scatterplot of the absolute estimated genetic effects of the pQTL lead variants v.s. theirminor allele frequencies (MAFs). d. The scatterplotinc showninlogarithm scale. e. Number of mapped pQTL per protein v.s. the linear mixed model estimated heritability in the ORCADES cohort. f. The variance explained by the mapped pQTL summed up for each protein v.s. the estimated heritability. g. For the proteins with significant cis-pQTL mapped, the lead variant signal strength v.s. the estimated heritability of each protein.

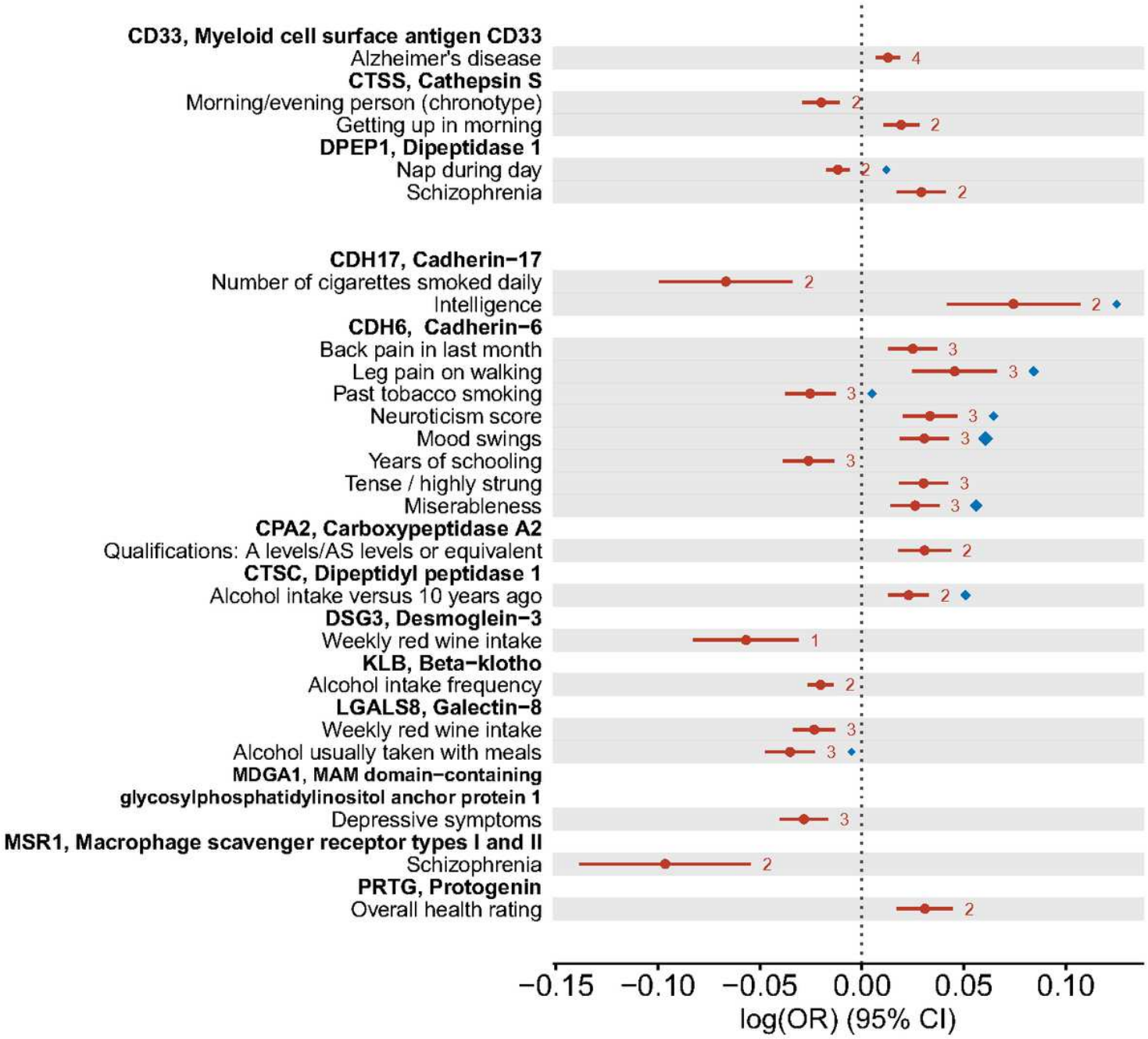

Figure 2

Causalitybetween theproteinsandneuro-relatedphenotypes inferredbyMendelian randomization (MR) analyses. The forest plot shows the significant MR results (false discovery rate< 0.05) based on LD-pruned (r 2 < 0.001) instrumental variants within each cis-pQTL. Inverse variance weighted (IVW) estimates are provided as the solid round dots, and the whiskers indicate 95% confidence intervals. The numbers of instrumental variants in the cis-pQTL are given to the right of the whiskers. As a colocalization measure, the HEIDI (heterogeneity in dependent instruments) test evidence (p > 0.05) are given as the diamonds, where the largest diamonds correspond to a p-value of 1. The upper part of the plot shows the results where the proteins are known druggable targets, while the lower part shows the results for new protein targets.

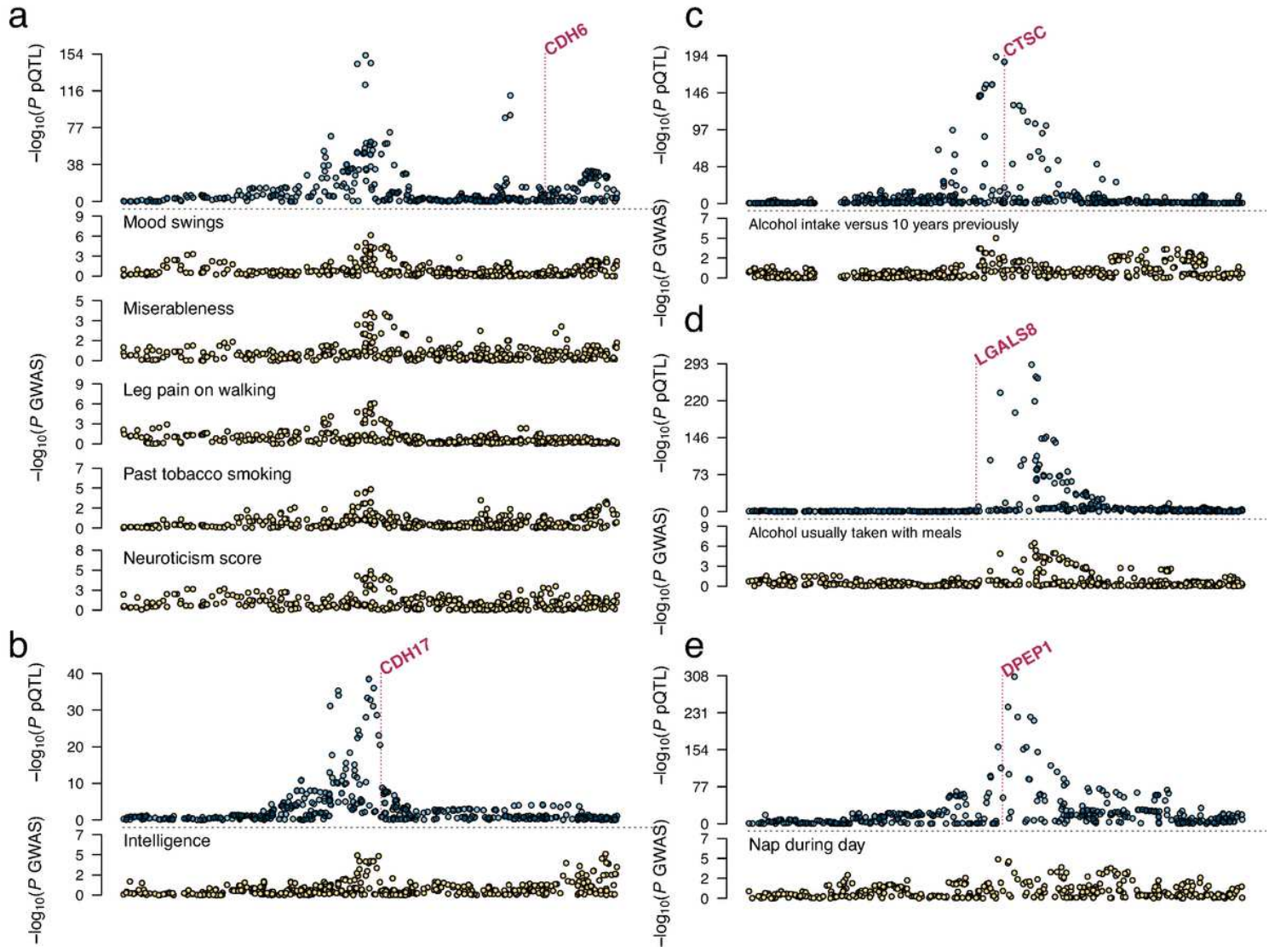

Figure 3

Regional association patterns of the pQTL and the colocalized neuro-related complex traits. The displayed protein-trait pairs correspond to the Mendelian randomization discoveries in Figure 2 with the HEIDI p-value > 0.05. Each subfigure shows the pQTL region of 1Mb centered at thelead variant. The vertical dashedlinein each subfiguremarks the transcription start site of the corresponding protein's coding gene.

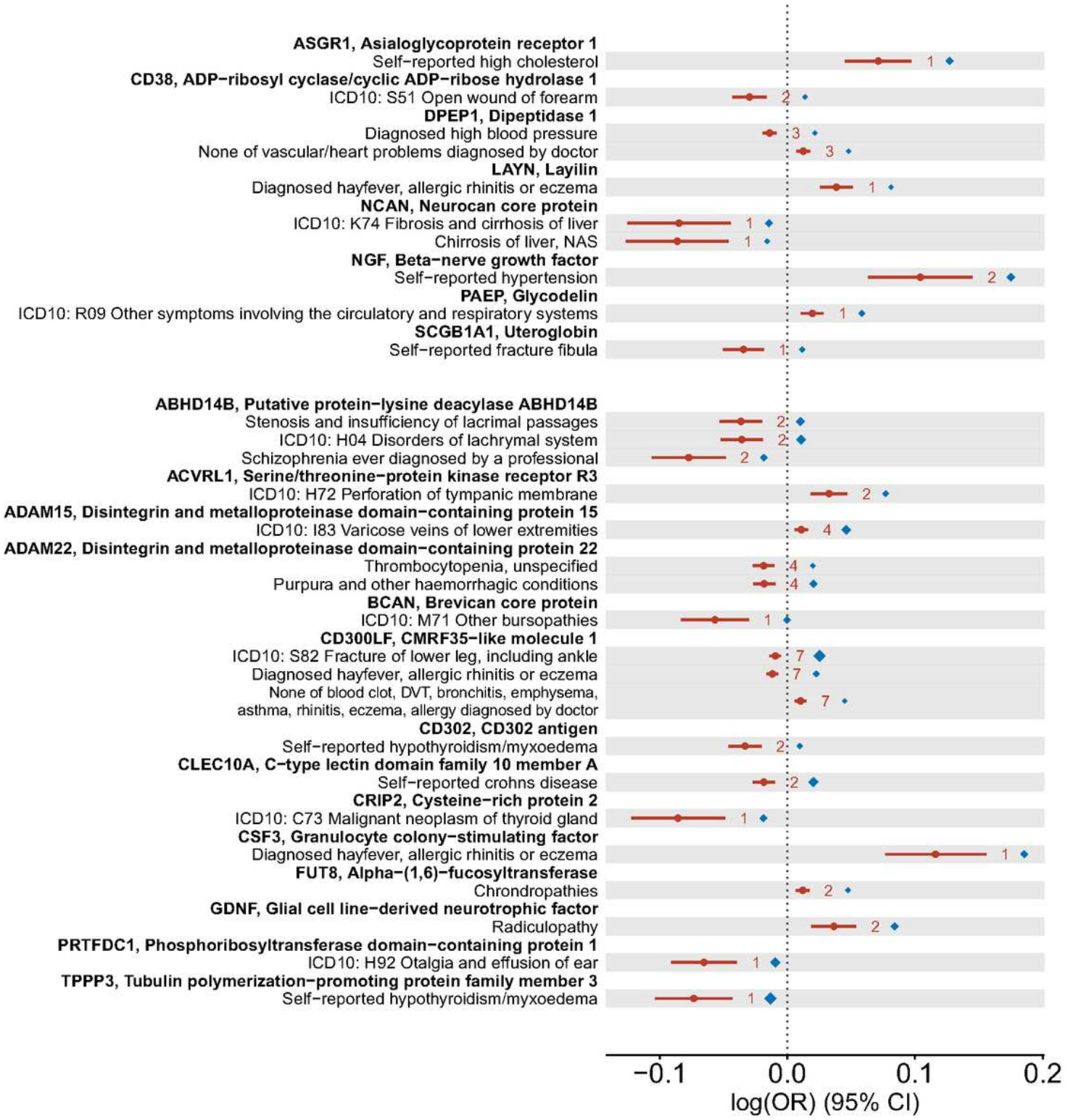

Figure 4

Causalitybetween theproteinsandUKBiobankdiseasephenotypes inferredbyMendelian randomization (MR) analyses. The forest plot shows the significant MR results (false discovery rate< 0.05) based on LD-pruned (r 2 < 0.001) instrumental variants within each cis-pQTL. Inverse variance weighted (IVW) estimates are provided as the solid round dots, and the whiskers indicate 95% confidence intervals. The numbers of instrumental variants in the cis-pQTL are given to the right of the whiskers. As a

colocalization measure, the HEIDI (heterogeneity in dependent instruments) test evidence ($p > 0.05$) are given as the diamonds, where the largest diamonds correspond to a p-value of 1. The upper part of the plot shows the results where the proteins are known druggable targets, while the lower part shows the results for new protein targets.

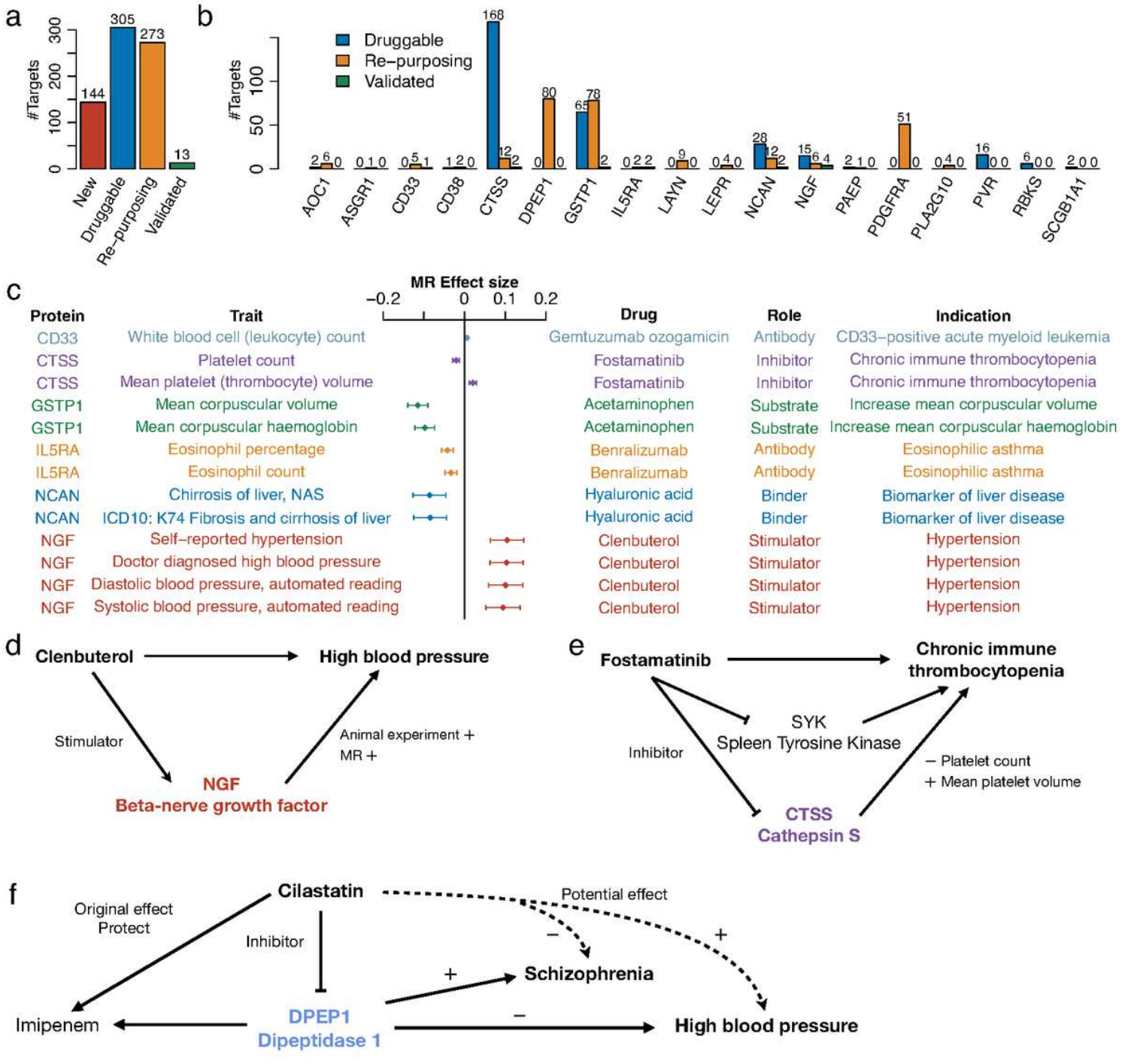

Figure 5

Drug targets revealed by Mendelian randomization (MR) analyses. The MR results with 5% false discovery rate are considered. a. The number of MR inferred pairs of proteins and traits split into four

categories: new (drug) targets, druggable targets that have drugs with unclear clinical function, re-purposing targets that have established drugs but for different diseases, and validated known targets where the established drugs have pharmacological effects that match the MR results. b. Numbers of re-purposing and validated drug targets per protein analysed. c. The validated known drug targets, the description of the drugs, and the corresponding consistent MR estimated effects. d. Potential mechanism of the adverse effect of Clenbuterol that targets NGF. e. Potentialmechanism of Fostamatinib treating Chronic immune thrombocytopenia through CTSS. f. Potential pharmacology of DPEP1's re-purposing drug on schizophrenia.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SCALLOPNEUXsuppinfocompressed.pdf