Uncovering bi-directional causal relationships between plasma proteins and psychiatric disorders: A proteome-wide study and directed network analysis

Carlos Kwan-long Chau[1], Alexandria Li-Ching Lau[1], Pak-Chung Sham[2,3], Hon-Cheong So[1,4-7*]

[1]School of Biomedical Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong

[2]Depeartment of Psychiatry, University of Hong Kong, Hong Kong

[3]Center for Genomic Sciences, University of Hong Kong, Hong Kong

[4]KIZ-CUHK Joint Laboratory of Bioresources and Molecular Research of Common Diseases, Kunming Institute of Zoology and The Chinese University of Hong Kong, China

[5]Department of Psychiatry, The Chinese University of Hong Kong, Shatin, Hong Kong

[6]CUHK Shenzhen Research Institute, Shenzhen, China

[7]Margaret K.L. Cheung Research Centre for Management of Parkinsonism, The Chinese University of Hong Kong, Shatin, Hong Kong

Correspondence to: Hon-Cheong So, Lo Kwee-Seong Integrated Biomedical Sciences Building, The Chinese University of Hong Kong, Shatin, Hong Kong. Tel: +852 3943 9255; E-mail: hcso@cuhk.edu.hk

Submission date: 6th July 2019

No. of pages: 26

Abstract

Psychiatric disorders as a whole represents a major public health burden yet their etiologies remain poorly understood, and treatment advances have been limited. In addition, there are no reliable biomarkers for diagnosis or progress monitoring of psychiatric disorders.

Here we performed a proteome-wide causal association study covering 3522 plasma proteins and 24 psychiatric traits or disorders. We conducted Mendelian randomization (MR) analysis leveraging GWAS summary data of plasma protein measurements and psychiatric disorders; in total about 95000 MR tests were conducted. The analysis was bi-directional: we studied how proteins may affect psychiatric disorder risks, but also looked into how psychiatric traits/disorders may be causal risk factors for changes in protein levels. We also performed a variety of additional analysis to prioritize protein-disease associations, including HEIDI test for distinguishing functional association from linkage, analysis restricted to *cis*- acting variants and replications in independent datasets from the UK Biobank. Based on the MR results, we constructed directed networks linking proteins and different psychiatric traits, hence shedding light on proteins mediating psychiatric comorbidities. For example, we observed that certain cytokine pathways may underlie the heightened risks of suicide and depression in schizophrenia subjects.

Identifying proteins causal to disease development have important implications on drug discovery or repurposing. Findings from this study may also guide the development of blood-based biomarkers for the prediction or diagnosis of psychiatric disorders, as well as assessment of disease progression or recovery.

Introduction

Psychiatric disorders as a whole represent a major public health burden and are leading causes of disability worldwide[1]. Nevertheless, there remain significant challenges in the diagnosis and treatment of psychiatric disorders. The etiologies of most psychiatric disorders are not fully understood yet, and we still have limited knowledge about the factors leading to disease progression or recovery. There have been limited advances on the treatment of psychiatric disorders, especially on the discovery of drugs with novel mechanisms of actions.

Another longstanding challenge is that unlike many other medical diseases, there are yet *no* reliable biomarkers for diagnosis, prognosis or monitoring progress of psychiatric disorders[2]. The biomarker should preferably easily accessible and measured, and blood/plasma proteins represent one type of biomarkers fulfilling such requirement. A number of studies have investigated into differences in protein levels in various psychiatric disorders, many of which are hypothesis-driven[2,3]. The rise of high-throughput proteomic platforms now enables interrogation of a large panel of

proteins at the same time. However, as pointed out by a recent review, there were several important limitations in these studies, for example inadequate sample sizes, methodological heterogeneity, and risk of confounding[4].

An important question underlying many protein marker (or proteomics) study in observational samples is that it is often difficult to assess causality and the direction of effect. Is the change in the level of a protein the cause or consequence of the disorder? Ideally one may follow-up a group of subjects at risk for the disorder and regularly monitor the protein level to ascertain the temporal relationship between disease onset and rise or fall in protein levels. However, this is often impractical due to low incidence of some disorders, the cost and difficulty in identifying at-risk subjects, and loss to follow-up. Another important concern is confounding bias. For example, patients are often medicated at the time of recruitment, and numerous other known or unknown confounders (e.g. age, gender, stressors, comorbid medical conditions) can result in spurious associations between a protein and the psychiatric disorder.

The recent decade has witnessed a rapid growth in genomic research in psychiatric disorders; in particular, high-throughput technologies such as genome-wide association studies (GWAS) have been highly successful in unraveling genetic variants associated with many neuropsychiatric traits. With the advent of GWAS, a new approach to causal inference, known as Mendelian randomization (MR), is gaining increased attention. An MR study can be considered a "naturalistic" randomized controlled trial (RCT)[5,6]. The basic principle is to use genetic variants as "instruments" to represent the actual risk factor, and analyze its relationship with the outcome. For example, a person who has inherited a low-density lipoprotein (LDL) lowering allele (or allelic score) at a relevant genetic locus is analogous to receiving a LDL-lowering drug, and vice versa. To estimate the causal effects of reduced LDL, we can then test whether people who have inherited the lipid-lowering genetic variants are less likely to develop cardiovascular diseases. The random allocation of alleles at conception is analogous to randomization in clinical trials; as such, MR studies are less vulnerable to confounding bias than observational studies[5]. And since a person's genetic profile is fixed at birth, this approach also avoids problems of reversed causality.

In this work we aim to investigate potential casual relationships between plasma protein levels and a range of psychiatric disorders/traits, based on the Mendelian randomization approach. As discussed above, a major advantage is that MR is less susceptible to confounder bias and reversed causality. MR also enables us to test direction of causal effects in both directions (protein-to-disease as well as disease-to-protein). In addition, MR can be performed using GWAS summary statistics, which are usually based on very large samples (usually >10000). On the contrary, the sample sizes of previous proteomics studies are typically small (*N* mostly <100; see ref[4]). The MR approach also provides a cost-effective way to explore a wide panel of proteins against a variety of psychiatric traits, which

will be prohibitively costly to carry out in a conventional longitudinal or observational study. Here we made use of recently published results of proteomic quantitative trait loci (pQTLs) and the latest large-scale GWAS data from the Psychiatric Genomics Consortium (PGC), UK Biobank and other sources to perform MR analysis.

While many of the MR studies were hypothesis-driven and focused on testing causal associations of known risk factors, MR can also be extended to screen for a larger number of risk factors to uncover novel causal associations. In the field of psychiatric genetics, hypothesis-driven candidate gene studies have now turned out to be largely irreproducible[7,8], suggesting that candidate-based approaches to 'omics' studies have important limitations. Here we adopted a systematic and unbiased 'hypothesis-generating' or 'hypothesis-free' approach and scans all available proteomic markers for casual relationships with psychiatric disorders in both directions.

By exploring proteins causal to the development of psychiatric disorders, we may have a better understanding into the pathogenesis of psychiatric disorders, gaining important mechanistic insight. Of note, many susceptibility SNPs found in GWAS reside in non-coding or inter-genic regions, and their functional impact remains to be elucidated[9]. Linking SNPs to protein QTLs will help to reveal the functional and biological impact of susceptibility variants. From a clinical point of view, drugs that target the causal proteins may be effective for disease prevention and treatment. If such drugs are already in use, they can be repositioned for the disorders, saving the high cost and time for new drug development. In addition, proteins with causal relationship with the disease may also serve as predictive biomarkers for at-risk patients before full development of the disorder.

On the other hand, the study of proteins whose levels are affected *as a consequence* of disease liability also has important implications. The discovery of such proteins may enhance our understanding of the molecular mechanisms underlying disease progression and complications. From a clinical perspective, such proteins may be candidate biomarkers for monitoring and assessing disease progress or recovery. An important definition of causality is that intervention on the cause will lead to changes in the outcome. If a disease is a *causal* risk factor for protein level change (let's say increase), then when the disease is cured, the protein level will change (decrease) accordingly. The reverse (i.e. rise in protein level) may be observed when the disease becomes more severe or during relapses. This may be particularly relevant in psychiatry, as patient progress is usually monitored by clinical observations and self-reported symptoms, with no objective biomarkers available. In addition to the above, sometimes a protein whose level changes as a result of a disease may yet be a risk factor for another condition. For example, depression is associated with increased risk of cardiometabolic disorders[10], and some have postulated inflammatory pathways and proteins as mediating factors[11,12]. Similarly, we may decipher mediating proteins between different psychiatric disorders/traits by studying bi-directional causal links between proteins and diseases. In

this study, we have presented a framework for directed network (a network in which edges have directions) analysis of psychiatric traits and proteins based on our bi-directional MR results, and external protein-protein interaction databases.

As a summary, in this work we first employed two-sample MR to identify proteins causally associated with psychiatric disorders/traits, with replication in the UK Biobank sample. Next we conducted MR analysis with psychiatric disorders/traits as exposure and protein levels as outcome. These analyses were conducted in a large-scale and systematic manner covering up to 3522 plasma proteins and 24 GWAS datasets of psychiatric disorders/traits. Based on these results, we constructed directed protein-disease networks which aimed to shed light on the pathways between various proteins and psychiatric disorders. The full analysis results are made publicly available, which we believe will serve as a useful resource for further clinical or experimental studies. An overview of our analytic framework is presented in Figure 1.

Methods

Blood Plasma Protein pQTL Datasets

Blood plasma protein pQTL data were taken from the recently published KORA and INTERVAL studies [13,14]. Full summary statistics of the two proteomics GWAS are downloaded from https://metabolomics.helmholtz-muenchen.de/pgwas/index.php?task=download and https://www.phpc.cam.ac.uk/ceu/proteins/. KORA is a German cohort with GWAS performed on levels of 1124 protein levels from a population-based cohort of 1000 subjects. Proteins were quantified by an aptamer-based affinity proteomics platform known as SOMAscan. While the original paper by Suhre et al.[13] also included a sample from the QMDiab study, that sample was recruited from Arab and South Asian countries, and was excluded from our analysis since most other GWAS datasets we analyzed were from Europeans. The INTERVAL study is a similar proteomics study from a UK cohort, investigating the association of 3622 plasma proteins in 3301 healthy subjects[14]. Again the SOMAscan platform was employed for protein quantification. For details of each dataset, please refer to the original papers[13,14].

We have included these two studies based on their relatively large sample sizes, the use of same platform for protein quantification, and the fact that both studies have full GWAS summary statistics available for download. Full proteomics GWAS data are required for our MR analysis to identify which proteins may be affected as a consequence of liability to psychiatric disorders and for HEIDI analysis.

Both studies classified pQTL-plasma-protein association as *cis* or *trans*, and in this work we would perform analysis on *cis* + *trans* pQTLs combined and *cis*-pQTLs only. *Cis*-pQTLs have

stronger prior biological evidence and are less susceptible (but not entirely immune) to horizontal pleiotropy in MR as they are close to the proteins studied. Here horizontal pleiotropy refers to instrument SNPs having effect on the outcome through pathways other than the exposure. However, both proteomic GWAS also identified a substantial number of *trans*-pQTLs passing genome-wide significance, and that modeling more SNPs will improve the power of MR analysis and enable correction for horizontal pleiotropy. For this reason, we conducted analysis with *cis & trans* pQTLs, as well as *cis*-pQTLs only.

Psychiatric GWAS datasets

We extracted psychiatric GWAS summary datasets mainly from the Psychiatric Genomic Consortium (PGC) and UK Biobank (UKBB). We included in total 24 GWAS datasets, which encompass a comprehensive range of psychiatric disorders (and traits closely linked to the disorders). They include attention deficit and hyperactivity disorder (ADHD)[15], anorexia nervosa (AN)[16], Alzheimer's disease (ALZ)[17], autistic spectrum disorders (ASD)[18], anxiety disorders[19] (including generalized anxiety disorder, phobic disorder, social phobia, agoraphobia, and specific phobias; results from both case-control study and derived factor score included), depressive symptoms (DS)[20], major depressive disorder (MDD)[21,22] (two sources, one sample[22] comprised female patients enriched for severe melancholic depression), longest period of depressed mood, number of episodes of depressed mood (GWAS from UK-Biobank data from Neale's lab[23]), neuroticism (NEU)[20], obsessive compulsive disorder (OCD)[24], post-traumatic stress disorder (PTSD)[25], bipolar disorder (BIP)[26], schizophrenia (SCZ)[27] and history of deliberate self-harm (DSH) or suicide[23]. We also include GWAS on antidepressant response (as both binary and quantitative outcome)[28], and a GWAS on SCZ versus bipolar disorder patients[29]. For ADHD and PTSD, we also performed separate analysis on male-only, female-only and combined samples, as the corresponding summary statistics has been made available and there is good evidence for gender differences in these disorders[30,31].

For some of the psychiatric disorders/traits, a large-scale ('primary') study is available from PGC or a related source, while a corresponding trait is also recorded in UKBB. These psychiatric disorders/traits include anxiety disorders, ASD, BIP, OCD, PTSD, AN, SCZ and ALZ (parent's ALZ disease status was used as a proxy for analysis). The details are given in Table S3. For these disorders/traits, we also included the UKBB association dataset as a replication set when we studied the causal effects of proteins on diseases.

MR analysis: protein as exposure and psychiatric disorder/trait as outcome

We first treated plasma protein levels as exposure and psychiatric disorders/traits as outcome. As discussed above, for some of the disorders/traits, a large-scale study dedicated to the study of the specific disorder is available in PGC (or a similar source such as IGAP[17]), while a similar corresponding trait was also recorded in the UKBB sample. In such cases, we would conduct a 'primary' analysis based on the PGC data, then extract the top results (FDR<0.1, based on combined

cis & trans pQTLs) for replication analyses using the UKBB data.

We extracted pQTL according to those provided in the original papers (Supplementary data 1 in Suhre et al.[13] and Supplementary Table 4 in Sun et al.[14]). Exposure instrument SNPs were harmonized with psychiatric GWAS summary data with the harmonise_data function in the R package "TwoSampleMR". LD-clumping was performed with the clump_data function in "TwoSampleMR" at an r-squared threshold of 0.1 (within 10000 kb). We set a more liberal r-squared threshold than the default (0.001) to allow possibly more instruments to be included. The threshold was set to a relatively low level such that only weak LD is allowed, in order to prevent instability in MR estimates. Residual SNP correlations were properly taken into account of in our inverse-variance weighted (IVW) and MR-Egger analyses (see Supp. Text)[32,33].

Statistical analysis methodology

The primary method of MR analysis is based on the number of SNPs (Table 1). If there is evidence of significant pleiotropy as revealed by a significant intercept term in Egger regression, we would employ the MR-Egger approach, which enables valid causal estimates to be given in the presence of imbalanced horizontal pleiotropy. Otherwise, the standard Wald ratio or IVW approach was used. The R package "MendelianRandomization" was used, which allowed correlations between instrument SNPs to be accounted for.

To control for multiple testing, we employed the false discovery rate (FDR) approach, which controls the expected proportion of false positives within the rejected hypotheses. FDR was computed by the Benjamini-Hochberg method[34] and a modified approach by Storey[35]. The latter approach (implemented in the R package "qvalue") also estimates the proportion of null which could lead to improvement in power (this is the default method used in this study). Note that the FDR approach is also valid under positive (regression) dependency of hypothesis tests[36]. The FDR-adjusted *p*-value, or *q*-value, refers to the lowest FDR incurred when calling a feature significant[35].

Other sensitivity tests

<u>Analysis restricted to *cis*-acting pQTLs</u>

We primarily present results from analysis of combined *cis*- and *trans*- pQTLs, as this would allow more instruments to be included, improving the power and reliability of results, and would enable methods such as MR-Egger to be performed to correct for imbalanced pleiotropy. We noted that a substantial proportion of pQTLs identified by Suhre and Sun et al. and were in fact *trans*- pQTLs (~27% and 70% respectively), hence discarding *trans*-acting variants will result in substantial waste of available data and loss of statistical power.

Nevertheless, c*is*-acting pQTL have strong prior probability of association and are less susceptible

to horizontal pleiotropy, although they can also affect the outcome via other pathways. Here we also performed additional analysis by sub-setting the original data to produce 2 extra exposure datasets (named *cis*-KORA, *cis*-INTERVAL), and MR was repeated with *cis*-pQTLs only.

HEIDI test

Sometimes the instrument SNP can be in LD with another SNP, which is causally associated with the outcome. This scenario is known as 'linkage' by Zhu et al.[37], which may be distinguished with functional associations (same SNP affecting both protein level and outcome) by the SMR-HEIDI test[37]. The test was designed especially for single-SNP MR analysis with the assumption of a single causal variant in the region. We employed the SMR program (https://cnsgenomics.com/software/smr/#SMR) written by the authors. Default settings were used, in which up to the top 20 pQTLs in the *cis* region were extracted for the HEIDI test. The test requires summary statistics of the proteomics GWAS, and is relatively computationally time-consuming to perform. We conducted HEIDI for single-SNP MR results with FDR<0.2 to limit the total number of tests to several hundreds. An HEIDI *p*-value >0.05 is marked as 'passing' the test as per the original authors, but due to multiple testing of a large number of hypotheses, we also presented FDR-adjusted *p*-values.

In practice, we note that occasionally HEIDI would fail to produce results due to presumably insufficient number of SNPs in the *cis* region of the target pQTL. Another important limitation is that currently HEIDI or other co-localization methods cannot produce reliable results with the number of causal variants >=2. For these reasons we included HEIDI as a sensitivity test, but we do not exclude associations failing this test from subsequent analyses. If multiple testing is accounted for by the FDR approach, a relatively small proportion of HEIDI results remained significant after FDR correction (see Results).

Replication Analysis

We sought replication for the results from our primary analysis (mainly PGC studies) in an independent sample from the UK Biobank. Replication was performed for three sets of MR results (each result represents a protein-disorder pair): (1) MR results based on KORA sample with FDR<0.1; (2) Results based on INTERVAL sample with FDR<0.1; (3) Results from joint KORA + INTERVAL meta-analysis with estimated FDR<0.1 (based on Simes *p*-value). In (3) only overlapping proteins in the two samples were included.

As for the methodology of joint analysis, we computed the maximum *p*-value (max_*p*) as a test of the hypothesis of non-null associations in *both* samples, as described by Nichols et al[38]. On the other hand, the Simes test[39] was conducted to combine the *p*-values to test the global null hypothesis (i.e. association in *any* sample). The max_p and Simes test procedures are both robust to (positive) dependencies of hypothesis tests[38,39]. We also computed a standard IVW meta-analysis with

"metagen" in R to obtain an aggregate estimate of causal effects from both PGC studies and UKBB. (Strictly speaking the statistical significance from a standard IVW meta-analysis may be inflated due to dependency of effect estimates, but this was included as an extra approach to prioritizing associations and for computing the overall effect sizes). We adopted a one-sided test for computing *p*-value in the replication sample.

We also attempted a more novel analysis known as replicability FDR (repFDR), which assesses the expected proportion of false replications[40]. This is a far more stringent test which tests the "union null hypothesis", i.e. H_0 is that the result is null in *any* one sample. H_1 (alternative hypothesis) represents non-null results or true associations in *every* testing sample, taking into account the number of hypothesis tested. Intuitively, when a large number of hypothesis is tested, to achieve low repFDR the hypothesis needs to be highly significant (having low FDR) in every sample. The authors proposed the "*r* value", which is defined as "the lowest FDR level at which we can say that the finding is among the replicated ones"[40]. The analysis was carried out with default settings using an online program provided by the authors (http://www.math.tau.ac.il/~ruheller/App.html).

MR in the other direction: Psychiatric disorders/traits as exposure and proteins as outcome

We then performed MR analysis in the other direction in which the 24 psychiatric disorders/traits were treated as exposure, while plasma protein levels were considered as outcome.

SNPs in psychiatric GWAS datasets were first selected based on the genome-wide significance threshold ($p < 5E-08$), then clumped at $r^2=0.1$ (within 10000 kb). The rest of the analysis follows what was described in the previous sections. The HEIDI test was performed for single-SNP analyses. In this case we have two independent outcome datasets (KORA and INTERVAL), so we also performed *replication* of the results from the first dataset on the second one. We conducted meta-analysis on the intersecting proteins and repFDR was computed as described above.

Conversion of coefficients from UKBB summary data

The UKBB GWAS results are based on linear regression of the outcome, regardless of whether it is binary or quantitative (details available at https://github.com/Nealelab/UK_Biobank_GWAS). For consistency and ease of meta-analysis with other GWAS data based on logistic regression, we converted the linear regression coefficients to those under a logistic regression model, based on the methodology by Lloyd-Jones et al.[41]. Standard error of the converted coefficients was derived by the delta method.

Directed network analysis of proteins and psychiatric traits/disorders

In ordinary MR analysis, we can just conclude whether one exposure is causally associated with one outcome. However, it is often of interest to investigate more complex relationships between the variables, in this cases proteins and diseases. For example, SCZ is highly comorbid with MDD and is likely a causal risk factor for MDD, but what is the biological mechanism behind this association? For example, it will be useful to find protein(s) which mediate the effect of SCZ on MDD. Previously a "two-step MR" procedure has been proposed to investigate a chain of causal relationships between three variables (e.g. the effect of an exposure on outcome via DNA methylation changes)[42]. The principle is to perform two MR analyses, the first from the exposure risk factor to methylation and the second from methylation to outcome. The same can be applied to model disease-protein relationships. Here we extended this analysis to an "*N*-step" MR, which allows longer pathways of causal relationships to be modeled. We also constructed *directed* protein-disease networks in which the nodes are either a protein or psychiatric disorder/trait.

We restricted our attention to intersecting proteins in the KORA and INTERVAL datasets. To be included in the network, the exposure-outcome pair should have a joint analysis *p*-value (from Simes test or maximum *p*) of less than 0.05 and the direction of effects should agree in both samples. To avoid too many isolated edges, proteins are included in the network only if they have (at least) one incoming and one outgoing edge. The edge weights reflect the strength of association between the disease and protein, derived from the regression coefficients from MR. The above criteria were chosen to balance interpretability and comprehensiveness of the network. The *p*-value threshold was set at a relatively liberal level to allow more proteins and traits to be modeled. However, the analytic framework presented here can also be applied with more stringent or liberal inclusion criteria.

We also constructed more complex networks by overlaying the above graph with protein-protein-interaction (PPI) sourced from the SIGnaling Network Open Resource – SIGNOR[43] (https://signor.uniroma2.it/). Only protein pairs whose members were both already present in our network and has a reported confidence level were include, which amounted to 121 pairs. The confidence levels from SIGNOR were then used as edge weightings.

Results

MR analysis: Proteins as exposure and psychiatric disorders as outcome

Main analysis based on the KORA and INTERVAL studies

The pQTLs of KORA and INTERVAL studies are listed in Tables S1a and S1b. In total 539 pQTLs were extracted from the KORA sample, of which 391 are *cis* variants. In total 1980 pQTLs were extracted from the INTERVAL data, of which 589 are *cis* variants. The KORA and INTERVAL datasets covered 284 and 1478 proteins respectively.

We have conducted totally 6038 and 25034 MR tests based on the KORA and INTERVAL samples respectively. Each test examined the effect of a protein on the psychiatric disorder/trait under study. We also repeated the analysis restricted to *cis* variants only, with 4285 and 8201 MR tests performed on KORA and INTERVAL respectively.

The main MR analysis results are presented in Table 2 and Supplementary Tables S1c to S1d. For the analysis based on KORA pQTLs, 13 protein-disease pairs were significant at an FDR level of 0.1 and 20 at FDR<0.2. For the analysis using pQTLs from INTERVAL sample, we found 391 associations at FDR<0.1 and 456 at FDR<0.2.

HEIDI test

We also performed HEIDI test as a sensitivity test to detect associations that may be due to linkage (Table S1c, S1d). We note that occasionally the program may fail to produce an output, presumably due to inadequate number of SNPs in the *cis* region that pass the preset p-value threshold (1.57E-3, equivalent to chi-square statistic of 10), or some variants may be filtered away due to too low or too high LD. The original authors used the conventional threshold $p<0.05$ as the threshold for passing or failing the test; however, we also note there is a high chance of false positives in view of the large number of tests performed. We computed FDR with the Benjamini-Hochberg method; among the protein-disease associations tested by HEIDI, a relatively modest proportion have low FDR-adjusted *p*-values <0.05 (overall 4.1 %), <0.1 (5.2%) or <0.2 (6.2%). It should be noted that failing the HEIDI does not always mean the result is invalid. For example, there can be other true causal variant(s) affecting the protein and outcome at the same time, which is/are also in LD with the target pQTL. However, the test is still useful in prioritizing proteins for further studies.

Joint analysis

We also performed a joint analysis on the intersecting proteins between the two pQTL datasets (Table 2 and Table S1e) to estimate of the aggregate causal effects. The results of IVW meta-analysis, maximum *p* (a test for non-null association in every sample) and Simes test (a test for non-null association in at least one sample) are given in Table S1e. Top 30 results sorted by Simes *p*-value are presented in Table 2.

We also performed ***cis*-only** analysis and the results are summarized in Table S2. The *cis*-pQTLs are summarized in Table S2a and S2b for reference. Results of KORA- or INTERVAL-only analysis are given in Table S2c and S2d, while joint analysis is presented in Table S2e. In total 12 and 15 associations achieved FDR<0.1 and <0.2 in the KORA sample respectively; 21 and 32 associations achieved FDR<0.1 and <0.2 in the INTERVAL sample respectively.

Cis-only analysis

There is significant overlap in the results between the *cis*- only and combined *cis & trans* pQTL

analyses. For example, 68 and 45 protein-disease associations showed Simes $p < 0.01$ in the *cis & trans* and *cis*-only analysis respectively, among which 44 of them overlapped (Table S1e). Note that we also provide a summary of the *cis*-only analysis results side-by-side with the main results (*cis + trans*) for easy reference.

Highlights of top protein-to-disease associations

We found a number of interesting protein-disease associations among the top results, and we shall highlight a few here. Here our discussion is restricted to the top ~50 results in the joint analysis, as listed in Table S1e. For instance, cathepsin B is found to be associated with increased ASD risk. Interestingly, a recent study reported that inhibition of cathespin B led to a reverse of increased leukocyte-endothelial adhesion (which is implicated in neuro-inflammation) in the cerebral vessels in an autistic mouse model[44]. We also found that interleulin-6 receptor subunit alpha is positively associated with SCZ. Inflammation is believed to play an important role in the pathogenesis of SCZ[45], and clinical studies have also found rise in inflammatory markers in SCZ[45]. Soluble IL-6 receptor (sIL-6R) may be involved in 'trans-signaling', in which the complex of IL-6 and sIL-6R may bind to gp130 and exert pro-inflammatory effects[46]. Another recent MR study also reported increased sIL-6R being associated with SCZ[47], but here the proteomics dataset used is different and the association is more specific (for the alpha subunit). The exact role of IL-6 pathways in SCZ remains to be elucidated in future studies. Also in line with the inflammation hypothesis in SCZ and other psychiatric disorders (e.g. ASD[48]), we observed that the protein endothelial cell-selective adhesion molecule (ESAM) was positively associated with SCZ and ASD risks. Neurovascular endothelial dysfunction as well as hyper-permeability of the blood brain barrier (BBB) may be linked to neuroinflammation, which has been implicated in a number of neuropsychiatric disorders[49,50]. Another interesting finding is that increased beta-endorphin levels was found to be causally associated with elevated risk of melancholic depression (the MDD-converge sample mainly recruited subjects with melancholic depression). Melancholic depression has been reported to be associated with hypothalamic-adrenal (HPA) axis hyper-activation, while beta-endorphin may be involved in the stimulation of HPA axis[51,52]. In clinical studies, both increase and decrease in beta-endorphin levels have been reported; yet some studies specifically focused on melancholic patients indeed reported elevation of beta-endorphin levels (please refer to Table 1 of Hegadoren et al.[53]). These findings call for the need of more homogeneous subgrouping of MDD patients in biomarker studies. Yet another finding that may be of clinical relevance was that lower ferritin was found to be casually linked to increased ASD risks. Reduced ferritin and iron deficiency were reported to be common among children with ASD[54,55]. Interestingly, a clinical study reported that iron supplementation improves sleep disturbance in ASD patients[56], although this is a case-only study without a placebo control arm.

Taken together, our findings may help shed light on the pathogenesis of psychiatric disorders, and in some occasions if drugs are available to target the tentative causal proteins, such drugs may be

repurposed to treat the corresponding disorders.

Replication analysis

Replication analysis was performed for MR results having FDR<0.1 in the KORA sample, INTERVAL sample or in joint analysis (Table S3). For the KORA dataset, altogether 6 associations were replicated with one-sided p-value <0.05 in the UKBB dataset (spanning three unique protein-disease pairs) (Table S3b). Three had replication FDR<0.05. Note that sometimes more than one UKBB dataset may correspond to the original phenotype and we included all of them in the replication set. The use of parental history of Alzheimer disease as a proxy of the disease phenotype has been successfully adopted in previous works[57] and this strategy was also employed in our replication analysis. In the INTERVAL sample, 97 associations achieved replication *p* (one-sided test) <0.05 in UKBB, and totally 77 had replication FDR<0.05 (Table S3c). We observed that many of the top results were from Alz-IGAP or SCZ; this is likely due to relatively large sample size and power of these two datasets. However, some of the top results for Alzheimer disease showed heterogeneity of instrument effects. This can be due to various reasons[58], but horizontal pleiotropy could be one, therefore such results might be viewed with caution. For comprehensiveness, we also extracted a set of intersecting proteins between the two pQTL datasets with FDR<0.1 in joint analysis (by Simes test), and carried out meta-analysis combining PGC results with UKBB, using either KORA (Table S3d) or INTERVAL (Table S3e) pQTL data.

We briefly highlight a few replicated protein-to-disease associations (we only focus on those listed in Table S3d/e due to the large number of associations). For example, protease C1 inhibitor, a protein involved in immune functioning, was among the top replicated associations for SCZ. Inflammation and immune pathways have been strongly implicated in the etiology of SCZ[45]. Protease C1 inhibitor an inhibitor of the classical pathway of complement and the contact system[59], which was reported to reduce the release of various cytokines in a sepsis model of primates. We observed that reduced levels of C1 inhibitor was causally linked to SCZ. This suggests that increased inflammatory responses, at least partially mediated by the complement system, may be a causal risk factor in the pathogenesis of SCZ. Further experimental work are required to reveal the mechanisms involved. Myeloid cell surface antigen (CD33) is another replicated causal association with Alzheimer disease (Table S3d/e), which has been reported to be associated with increased neuritic amyloid pathology and microglial activation[60], and has been suggested as a promising therapeutic target[61].

Relationship to drug repositioning or discovery

If a protein is causal to a psychiatric disorder, then a drug which targets the protein as an antagonist or inhibitor may be able to treat the disease if increased protein levels are casually associated with the disease. On the contrary, the psychiatric disorder may be a side-effect of that drug if reduced protein levels are associated with disease. The reverse is true for the drugs acting as agonists. To facilitate the discovery of candidates for drug repositioning and the exploration of

medication side-effects, we also searched ChEMBL database for existing drugs for top proteins in joint analysis (Tables S1i).

MR analysis: Psychiatric disorders/traits as exposure and proteins as outcome

MR analysis results

We also performed MR in the other direction, with psychiatric disorders/traits as exposure and proteins as outcome. The results are presented in Table S4. In total 16 psychiatric disorders have at least one genetic variant having p<5e-8 (Table S4a). However, some of these genome-wide significant SNPs cannot map to the SNP panels of KORA and INTERVAL. Finally, 11 and 15 psychiatric traits remained to be analyzed with KORA and INTERVAL, considering plasma protein levels as outcome.

Totally 12142 and 39009 MR tests were performed based on the KORA and INTERVAL datasets respectively. Note that since KORA and INTERVAL are independent outcome samples, we may consider the analyses as 'replications' of one another. Based on the KORA sample alone, 15 disease-protein associations achieved FDR<0.2 (Table S4b). Based on the INTERVAL sample, 495 FDR<0.1 and 1005 had FDR<0.2 (Table S4c). Finally we conducted meta-analysis on the pairs of associations analyzed in both pQTL samples (Table S4d). In total 111 and 236 meta-analysis results achieved FDR <0.1 and <0.2 respectively. In occasional cases single-SNP MR has to be performed; HEIDI results are presented alongside the main analysis in Table S4b and S4c.

Highlights of top disease-to-protein associations

We observed a number of interesting disease-to-protein associations supported by literature. For example, we found SCZ is causally linked to increase in complement C4 levels (Table 3 and S4d). The complement system, especially C4, has been shown to be implicated in SCZ[62,63]. Interestingly, a recent clinical study reported that C4 was significantly increased in chronic SCZ patient and was associated with positive and negative symptom severity[64], highlighting the potential of C4 as a disease activity biomarker. Another interesting finding was that SCZ was found to be a causal factor for rise in thyroid-stimulating hormone (TSH) (Table S4d). In a clinical study involving 40,483 patients with hypothyroidism and similar number of controls, Sharif et al. observed a higher prevalence of schizophrenia (2.01% vs. 1.25%, respectively, $p < 0.0001$)[65]. There is likely a complex link between neurotransmission and thyroid dysregulation[66] which may explain the association. Our findings suggest that the disease of SCZ itself may contribute to increased TSH (and hypothyroidism), amongst other possible risk factors such as medications. Moreover, we observed SCZ is a causal risk factor for decreased Interleukin (IL)-17 receptor C levels. The soluble secreted IL-17 receptors present in the blood may serve as extracellular antagonists to cytokine signaling[67] (i.e. reduced soluble receptors lead to rise in cytokine levels). Interestingly, IL-17 was observed to be positively correlated with total PANSS scores in schizophrenia. After risperidone treatment, the

proportion of Th17 cells (these cells produce IL-17) decreased and the change in proportion was correlated with the rate of change of PANSS score[68]. This is consistent with our present finding, and suggests proteins in the IL-17 signaling pathway may be useful biomarkers to evaluate disease progress or treatment response.

Directed network analysis of protein-disease associations

We constructed directed networks based on our bi-directional MR results, and supplemented by external PPI databases. The resulting networks are presented in Figures 2 and 3. For easier visualization and interpretability, we present separate networks with positive or negative associations only (Figures 2 and 3) or both (Figure S1). Networks with or without added edges from external PPI database are also presented separately. Network analyses based on significant edges from the Simes method are presented in Figures 2, 3 and S1 while those based on the max *p* procedure are presented in Figure S2.

We highlight a few examples in which our network analysis helps to reveal potential mediating proteins between psychiatric disorder/traits, hence shedding light on the mechanism underlying disease comorbidities. For example, we observed that kyneuraniase (KYNU) may mediate the relationship between SCZ and MDD (i.e. SCZ -> KYNU -> MDD). KYNU is an enzyme that catalyzes the conversion of 3-hydroxykynurenine (OHK), a metabolite of kynurenine (KYN), to 3-hydroxyanthranilic acid (3-HAA). 3-HAA can be further degraded to quinolinic acid (QUIN) which is considered neurotoxic because of its agonistic action on the NMDA receptor. In an alternative pathway, KYN is converted to kynurenic acid (KYNA). Interestingly, in a recent study on patients with affective psychosis, the KYNA/QUIN ratio was found to be reduced[69]. In major depression, elevated QUIN and reduction of KYNA/QUIN ratio has been observed, [70]which was found to be reversed with electroconvulsive therapy[71]. Our findings, coupled with the previous literature, suggest that the kynurenine pathway may play a role in the pathogenesis of comorbid depression and psychosis. Another interesting finding is that SCZ may be causally associated with increased in levels of CCL23 which in turns increases the risk of deliberate self-harm/suicide (SCZ -> CCL23 -> self-harm/suicide). While CCL23 has not been specifically studied in psychiatric disorders, chemokines in general play an important role in immune functioning and inflammatory processes, and altered levels of certain chemokines have been previously reported for SCZ and other psychiatric disorders[72]. For instance, a recent study reported an increase in “chemokine index” (derived from two chemokines) with duration of schizophrenia and severity of negative symptoms[73]. On the other hand, elevated cytokines have been observed to be associated with suicidality in a meta-analysis[74]. Along the line of inflammation in psychiatry disorders, we also observed IL-6 receptor subunit alpha as a potential mediating factor between SCZ and MDD as well as PTSD. As discussed above, soluble IL-6 receptor plays a part in ‘trans-signaling’ and can promote

inflammation. The IL-6 signaling pathway is implicated in many psychiatric disorders, and raised sIL-6R levels have been observed in depression[75] and PTSD[76,77]. Our present findings suggest that increased IL-6 signaling may represent one of the mechanisms underlying the heightened risks of MDD[78] and PTSD[79] within SCZ patients.

Discussion

Here we have conducted a causal proteome-wide association study of over 3000 plasma proteins with 24 GWAS datasets of psychiatric traits/disorders. We leveraged GWAS summary statistics of plasma protein measurements and psychiatric disorders, and perform bi-directional Mendelian randomization analysis to uncover casual relationships between the two entities. We have performed ~95,000 MR tests in total, and also performed several additional analyses such as HEIDI test, analysis in *cis*- variants only as well as replication in independent datasets (for both protein-to-disease and disease-to-proteins analyses). We believe this work will shed light on the pathophysiology underlying the development and progression of psychiatric disorders. It may also facilitate the development of novel biomarkers for prediction, diagnosis and monitoring of disease progression or recovery. By targeting proteins that may be causal to the disease, our findings may also help drug development or repurposing for psychiatric disorders. Importantly, all our analysis results are also made available online and are easily searchable, which we believe will represent a useful resource for basic and clinical researchers interested in further studies of plasma proteins in psychiatric disorders.

Here we conducted *bi*-directional analyses to investigate not only how proteins affect psychiatric disorder risks, but also how psychiatric traits/disorders may be causal risk factors for changes in protein levels. This is a unique feature of the current study, as previous works on 'omics'-based MR usually focused on, for example, how expression or methylation changes contribute to diseases[37]. Causal relationships in the reverse direction are however far less explored. By the definition of causality, if an intervention is applied to the cause (disease), then the outcome (protein) should change. Therefore such biomarkers are theoretically justified to evaluate the disease progression, treatment response or resolution from disease symptoms. In this work, we have found a number of putative disease-to-protein associations worthy of further investigations, as already highlighted above.

As a technical note, when the MR exposure is a binary variable (disease or not), it is preferable to conceptualize it as a (possibly latent) continuous risk factor[80]. For example, if the studied exposure is hypertension, then blood pressure would be the underlying continuous risk factor or trait[80]. For most psychiatric disorders, we may also conceptualize a continuous factor underlying the disorder. For instance, depressive symptoms can be the latent factor underlying MDD. Similarly, one can conceptualize a continuous scale of (or tendency to) anxiety, psychosis, OCD etc. as the latent risk

factor or trait. The principles of MR causal inference still apply.

In addition to the potential clinical applications described above, we also wish to highlight two other aspects of translational potential. Besides diagnosis-based phenotypes, we also included in our analysis a GWAS on antidepressant response (as both binary and continuous outcome), and a GWAS on SCZ *versus* bipolar disorder (BD) patients. The former analysis aims to reveal proteins causally associated with response to antidepressants, and may help discover candidate protein markers to predict treatment response. MR analysis of the latter dataset (SCZ vs BD) could help to explore differences in proteomic profiles between the two highly related disorders. The results may ultimately help to prioritize protein biomarkers that may contribute to differential diagnosis (DDx) between BD and SCZ, given that response to medications such as lithium and valproate can differ between the disorders. Obviously similar analysis may also be conducted for other kinds of psychiatric disorders, hence the presented analytic framework also points to a new direction for finding biomarkers in predicting drug response and DDx in general.

Another contribution of this study is a directed network analysis of disease protein associations of psychiatric traits/disorders. A network approach to analyzing genetic and protein associations enables us to gain more comprehensive insight into the pathways and complex pathophysiology underlying complex diseases. Several methods have been developed for network analysis of GWAS data[81,82]. However, previous analytic methods have mainly focused on *un*directed networks; to our knowledge, this is the first study to present a *directed* network analysis framework across *multiple* complex disease traits, based on GWAS data and the principle of MR. Such networks are useful in unraveling the mechanisms underlying disease comorbidities and complications.

There are several limitations that are worth noting. The MR approach is less vulnerable to reverse causality and confounding compared to observational studies, however in some scenarios it is still not easy to establish causality. In the study of causal effects of proteins to psychiatric disorders, the number of instrument SNPs is usually quite small, and it is quite common to rely on only one instrument (i.e. single-SNP MR). There are no methods to distinguish between causality and pleiotropy in single-SNP MR, although the HEIDI test can help to identify 'linkage' from the same SNP being associated with protein levels and disease at the same time[37]. When the number of instruments is larger, more tools are available to address the assumptions of MR, especially pleiotropy[58]. However, the number of instruments is still typically small when proteins are treated as exposure; the power of methods like Egger regression to detect pleiotropy is relatively weak under such scenarios[33]. Methods based on outlier removal (e.g. MR-PRESSO[81]) are also not applicable in the presence of few instruments. On the whole, we may still prioritize proteins that more likely causal to diseases, but fully establishing causality remains difficult. Nevertheless, at least the current study could show which proteins may be *associated with* various psychiatric disorders, which helps to limit the search space for suitable protein biomarkers or target proteins for drug development. For MR analysis with psychiatric disorders as exposure, the above problem of inadequate instruments is

usually of a lesser concern, especially for disorders like schizophrenia and Alzheimer's disease with large GWAS sample sizes. Nevertheless, for some other disorders such as anorexia, the number of available instruments is small so the same problem can occur.

The MR approach itself also has limitations. As MR is based on genetic instruments to represent the exposure, the results reflect a chronic exposure of (genetically predicted) higher/lower protein levels to the psychiatric disorder. The effects of short-term exposure, such as taking an inhibitor targeting the protein, may not be the same as effects from long-term exposure as predicted by MR.

With regards to the study samples, many psychiatric disorders are highly heterogeneous with regards to their presentations, course of disease, prognosis etc[84]. A protein may be associated with a subgroup of patients with certain clinical characteristics only (e.g. female patients with early onset) but not with the other groups. A future direction is to conduct the analysis in stratified or more deeply-phenotyped samples (e.g. male or female-only; early or late-onset only; with a particular clinical presentation etc.). Also, for some of the samples (especially those from UKBB), the disease diagnosis or symptoms are self-reported instead of being assessed by a clinical professional. The latter is probably not practical in a biobank due to high cost of performing detailed assessments. Nevertheless, further replications in more carefully assessed and phenotyped samples are warranted.

Also related to sample characteristics, we have tried to employ the latest and largest GWAS samples. Nevertheless, for some of the psychiatric disorders (e.g. OCD or PTSD), the sample size may still be relatively small. Also we have performed replication analysis in the UKBB, however the UKBB in general does not contain a lot of patients with psychiatric disorders (except for depression and anxiety). As a result, the power of replication study is relatively limited and absence of significant replications does not exclude genuine associations.

One potential translational aspect is the development of blood-based biomarkers for psychiatric disorders. While this work we mainly study the protein candidates one by one, in practice and future studies, we may combine different biomarkers and even other modalities (e.g. neuroimaging) to improve the performance of the test. This will be left as a direction for future investigations.

In summary, we have conducted a proteome-wide association study to decipher the causal relationship between plasma proteins and psychiatric disorders. We believe this work sheds light on the pathophysiology of psychiatric disorders, and will help facilitate the development of protein biomarkers as well as new or repurposed drugs for the treatment of psychiatric disorders.

Acknowledgements

We would like to thank Prof. Stephen Tsui and the Hong Kong Bioinformatics Center for computing support. This study was partially supported by the Lo Kwee Seong Biomedical Research Fund, the Health and Medical

Research Fund and a Chinese University of Hong Kong Direct Grant.

Author contributions

Conceived and designed the study: HCS. Supervised the study: HCS. Main data analysis: CKLC. Data interpretation: HCS, CKLC, PCS. Data analysis methodology: HCS, with input from PCS. Literature review, assistance in data analysis: ALCL. Protein Network analysis: ALCL. Drafted the manuscript: HCS.

Conflicts of interest

The author declares no conflict of interest.

Supplementary Text

MR analysis methods

The IVW framework is very widely used in MR studies. Here we used an IVW approach that is able to account for SNP correlations as described in Burgess et al.[32]. Briefly, assume $\hat{\beta}_{YG}$ to be the vector of estimated regression coefficients when the outcome is regressed on genetic instruments and σ_{YG} to be the corresponding standard errors (SE), and $\hat{\beta}_{XG}$ to be the estimated coefficients when the risk factor is regressed on the genetic instruments with SE σ_{XG}. We also assume the correlation between two genetic variants G_1 and G_2 to be ρ_{G1G2}, and $\Sigma_{G_1G_2} = \rho_{G_1G_2}\sigma_{YG_1}\sigma_{YG_2}$.

The estimate from a weighted generalized linear regression can be formulated by

$$\hat{\beta} = \left(\hat{\beta}'_{XG}\Sigma\hat{\beta}_{XG}\right)^{-1}\hat{\beta}'_{XG}\Sigma^{-1}\hat{\beta}_{YG}$$

with SE

$$SE(\hat{\beta}) = \sqrt{\left(\hat{\beta}'_{XG}\Sigma\hat{\beta}_{XG}\right)^{-1}}$$

A similar approach may be used for MR-Egger, which allows an intercept term in the weighted regression. Please also refer to ref. [33,83] for details. The presence of imbalanced horizontal pleiotropy could be assessed by the significance of the intercept term (testing whether intercept=0). We employed the R package "MendelianRandomization" for the above analyses.

As remarked by Burgess et al.[32], inclusion of a larger number of variants in partial LD may improve the power of MR, as the variance explained by the instruments might go up. Including "redundant" SNPs in addition to the causal variant(s) will not improve power; on the other hand it will not invalidate the results.

However, including too many variants with high correlations may also result in instability of MR estimates[84]. In this study, we performed LD-clumping of genetic instruments at an r^2 threshold of 0.1, balancing the benefits and risks of using a too high or too low threshold.

Figure Legends

Figure 1 An overview of the analytic framework. Briefly, we conducted MR analysis with psychiatric disorders/traits as exposure and protein levels as outcome. The analysis was conducted in a large-scale and systematic manner covering up to 3522 plasma proteins and 24 GWAS datasets of psychiatric disorders/traits. We also conducted several additional analyses, including a cis-only analysis (as cis-variants are as instruments less likely to be affected by pleiotropy), HEIDI test for selected associations to distinguish 'linkage' from functional association, as well as replication in UK Biobank of top protein-to-disease associations. Based on these results, we constructed directed protein-disease networks which aimed to shed light on the pathways between various proteins and psychiatric disorders.

Figure 2 Directed network based on MR results (protein-protein interactions not considered). Here we only considered proteins present in both KORA and INTERVAL samples and with Simes $p < 0.05$ in the joint analysis of protein-to-disease or disease-to-protein associations. For easier visualization and interpretability, Figure 2A is restricted to positive causal associations while Figure 2B is restricted to negative associations only. Figure 2C includes both positive and negative edges.

Figure 3 Directed network based on MR results, incorporating protein-protein interaction network from the SIGNOR database. Here we only considered proteins present in both KORA and INTERVAL samples and with Simes $p < 0.05$ in the joint analysis of protein-to-disease or disease-to-protein associations. Figure 3A is restricted to positive causal associations while Figure 3B is restricted to negative associations only. Please also refer to Figure S1 which includes both types of associations.

Figure S1 Directed network based on MR results, incorporating protein-protein interaction network from the SIGNOR database. Here we only considered proteins present in both KORA and INTERVAL samples and with Simes $p < 0.05$ in the joint analysis of protein-to-disease or disease-to-protein associations. This figure includes both positive and negative associations between proteins and diseases.

Figure S2 Directed network based on MR results, incorporating protein-protein interaction network from the SIGNOR database. Here we considered proteins present in both KORA and

INTERVAL samples and with *maximum p* < 0.05 in the joint analysis of protein-to-disease or disease-to-protein associations. This figure includes both positive and negative associations between proteins and diseases.

References:

1. Vigo, D., Thornicroft, G. & Atun, R. Estimating the true global burden of mental illness. *Lancet Psychiatry* **3**, 171-8 (2016).
2. Venkatasubramanian, G. & Keshavan, M.S. Biomarkers in Psychiatry - A Critique. *Ann Neurosci* **23**, 3-5 (2016).
3. Mora, C., Zonca, V., Riva, M.A. & Cattaneo, A. Blood biomarkers and treatment response in major depression. *Expert Rev Mol Diagn* **18**, 513-529 (2018).
4. Comes, A.L. *et al.* Proteomics for blood biomarker exploration of severe mental illness: pitfalls of the past and potential for the future. *Transl Psychiatry* **8**, 160 (2018).
5. Bennett, D.A. & Holmes, M.V. Mendelian randomisation in cardiovascular research: an introduction for clinicians. *Heart* (2017).
6. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* **23**, R89-98 (2014).
7. Johnson, E.C. *et al.* No Evidence That Schizophrenia Candidate Genes Are More Associated With Schizophrenia Than Noncandidate Genes. *Biol Psychiatry* **82**, 702-708 (2017).
8. Border, R. *et al.* No Support for Historical Candidate Gene or Candidate Gene-by-Interaction Hypotheses for Major Depression Across Multiple Large Samples. *Am J Psychiatry* **176**, 376-387 (2019).
9. Zhang, F. & Lupski, J.R. Non-coding genetic variants in human disease. *Human Molecular Genetics* **24**, R102-R110 (2015).
10. Scott, K.M. Depression, anxiety and incident cardiometabolic diseases. *Curr Opin Psychiatry* **27**, 289-93 (2014).
11. Wong, B.C. *et al.* Differential associations of depression-related phenotypes with cardiometabolic risks: Polygenic analyses and exploring shared genetic variants and pathways. *Depress Anxiety* **36**, 330-344 (2019).
12. Halaris, A. Inflammation-Associated Co-morbidity Between Depression and Cardiovascular Disease. *Curr Top Behav Neurosci* **31**, 45-70 (2017).
13. Suhre, K. *et al.* Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun* **8**, 14357 (2017).
14. Sun, B.B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73-79 (2018).
15. Martin, J. *et al.* A Genetic Investigation of Sex Bias in the Prevalence of Attention-Deficit/Hyperactivity Disorder. *Biol Psychiatry* **83**, 1044-1053 (2018).
16. Duncan, L. *et al.* Significant Locus and Metabolic Genetic Correlations Revealed in

Genome-Wide Association Study of Anorexia Nervosa. *Am J Psychiatry* **174**, 850-858 (2017).
17. Lambert, J.C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* **45**, 1452-8 (2013).
18. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet* **51**, 431-444 (2019).
19. Otowa, T. *et al.* Meta-analysis of genome-wide association studies of anxiety disorders. *Molecular Psychiatry* **21**, 1391-1399 (2016).
20. Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet* **48**, 624-33 (2016).
21. Wray, N.R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* **50**, 668-681 (2018).
22. consortium, C. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* **523**, 588-91 (2015).
23. UK Biobank results from Neale's Lab (http://www.nealelab.is/uk-biobank/). .
24. International Obsessive Compulsive Disorder Foundation Genetics, C. & Studies, O.C.D.C.G.A. Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Mol Psychiatry* **23**, 1181-1188 (2018).
25. Duncan, L.E. *et al.* Largest GWAS of PTSD (N=20 070) yields genetic overlap with schizophrenia and sex differences in heritability. *Mol Psychiatry* **23**, 666-673 (2018).
26. Stahl, E.A. *et al.* Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature Genetics* **51**, 793-+ (2019).
27. Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-7 (2014).
28. Investigators, G., Investigators, M. & Investigators, S.D. Common genetic variation and antidepressant efficacy in major depressive disorder: a meta-analysis of three genome-wide pharmacogenetic studies. *Am J Psychiatry* **170**, 207-17 (2013).
29. Ruderfer, D.M. *et al.* Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Mol Psychiatry* **19**, 1017-1024 (2014).
30. Ramtekkar, U.P., Reiersen, A.M., Todorov, A.A. & Todd, R.D. Sex and age differences in attention-deficit/hyperactivity disorder symptoms and diagnoses: implications for DSM-V and ICD-11. *J Am Acad Child Adolesc Psychiatry* **49**, 217-28 e1-3 (2010).
31. Hourani, L., Williams, J., Bray, R. & Kandel, D. Gender differences in the expression of PTSD symptoms among active duty military personnel. *J Anxiety Disord* **29**, 101-8 (2015).
32. Burgess, S., Dudbridge, F. & Thompson, S.G. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Stat Med* **35**, 1880-906 (2016).
33. Bowden, J., Smith, G.D. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of*

Epidemiology **44**, 512-525 (2015).

34. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**, 289-300 (1995).
35. Storey, J.D. The positive false discovery rate: A Bayesian interpretation and the q-value. *Annals of Statistics* **31**, 2013-2035 (2003).
36. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* **29**, 1165-1188 (2001).
37. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* **48**, 481-7 (2016).
38. Nichols, T., Brett, M., Andersson, J., Wager, T. & Poline, J.B. Valid conjunction inference with the minimum statistic. *Neuroimage* **25**, 653-660 (2005).
39. Simes, R.J. An Improved Bonferroni Procedure for Multiple Tests of Significance. *Biometrika* **73**, 751-754 (1986).
40. Heller, R., Bogomolov, M. & Benjamini, Y. Deciding whether follow-up studies have replicated findings in a preliminary large-scale omics study. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 16262-16267 (2014).
41. Lloyd-Jones, L.R., Robinson, M.R., Yang, J. & Visscher, P.M. Transformation of Summary Statistics from Linear Mixed Model Association on All-or-None Traits to Odds Ratio. *Genetics* **208**, 1397-1408 (2018).
42. Relton, C.L. & Smith, G.D. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *International Journal of Epidemiology* **41**, 161-176 (2012).
43. Perfetto, L. *et al.* SIGNOR: a database of causal relationships between biological entities. *Nucleic Acids Research* **44**, D548-D554 (2016).
44. Wang, H. *et al.* Cathepsin B inhibition ameliorates leukocyte-endothelial adhesion in the BTBR mouse model of autism. *Cns Neuroscience & Therapeutics* **25**, 476-485 (2019).
45. Khandaker, G.M. *et al.* Inflammation and immunity in schizophrenia: implications for pathophysiology and treatment. *Lancet Psychiatry* **2**, 258-270 (2015).
46. Rose-John, S. IL-6 Trans-Signaling via the Soluble IL-6 Receptor: Importance for the Pro-Inflammatory Activities of IL-6. *International Journal of Biological Sciences* **8**, 1237-1247 (2012).
47. Hartwig, F.P., Borges, M.C., Horta, B.L., Bowden, J. & Smith, G.D. Inflammatory Biomarkers and Risk of Schizophrenia A 2-Sample Mendelian Randomization Study. *Jama Psychiatry* **74**, 1226-1233 (2017).
48. Kern, J.K., Geier, D.A., Sykes, L.K. & Geier, M.R. Relevance of Neuroinflammation and Encephalitis in Autism. *Frontiers in Cellular Neuroscience* **9**(2016).
49. Kealy, J., Greene, C. & Campbell, M. Blood-brain barrier regulation in psychiatric disorders.

Neurosci Lett, 133664 (2018).

50. Najjar, S. *et al.* Neurovascular Unit Dysfunction and Blood-Brain Barrier Hyperpermeability Contribute to Schizophrenia Neurobiology: A Theoretical integration of Clinical and experimental evidence. *Frontiers in Psychiatry* **8**(2017).
51. Yamauchi, N., Shibasaki, T., Wakabayashi, I. & Demura, H. Brain beta-endorphin and other opioids are involved in restraint stress-induced stimulation of the hypothalamic-pituitary-adrenal axis, the sympathetic nervous system, and the adrenal medulla in the rat. *Brain Research* **777**, 140-146 (1997).
52. Iyengar, S., Kim, H.S. & Wood, P.L. Mu-Opioid, Delta-Opioid, Kappa-Opioid and Epsilon-Opioid Receptor Modulation of the Hypothalamic Pituitary Adrenocortical (Hpa) Axis - Subchronic Tolerance Studies of Endogenous Opioid-Peptides. *Brain Research* **435**, 220-226 (1987).
53. Hegadoren, K.M., O'Donnell, T., Lanius, R., Coupland, N.J. & Lacaze-Masmonteil, N. The role of beta-endorphin in the pathophysiology of major depression. *Neuropeptides* **43**, 341-353 (2009).
54. Latif, A., Heinz, P. & Cook, R. Iron deficiency in autism and Asperger syndrome. *Autism* **6**, 103-114 (2002).
55. Herguner, S., Kelesoglu, F.M., Tanidir, C. & Copur, M. Ferritin and iron levels in children with autistic disorder. *European Journal of Pediatrics* **171**, 143-146 (2012).
56. Dosman, C.F. *et al.* Children with autism: Effect of iron supplementation on sleep and ferritin. *Pediatric Neurology* **36**, 152-158 (2007).
57. Marioni, R.E. *et al.* GWAS on family history of Alzheimer's disease. *Transl Psychiatry* **8**, 99 (2018).
58. Hemani, G., Bowden, J. & Smith, G.D. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Human Molecular Genetics* **27**, R195-R208 (2018).
59. Wang, Y. *et al.* Decay accelerating factor (CD55) protects neuronal cells from chemical hypoxia-induced injury. *Journal of Neuroinflammation* **7**(2010).
60. Bradshaw, E.M. *et al.* CD33 Alzheimer's disease locus: altered monocyte function and amyloid biology. *Nature Neuroscience* **16**, 848-U92 (2013).
61. Jiang, T. *et al.* CD33 in Alzheimer's Disease. *Molecular Neurobiology* **49**, 529-535 (2014).
62. Mayilyan, K.R., Weinberger, D.R. & Sim, R.B. The complement system in schizophrenia. *Drug News & Perspectives* **21**, 200-210 (2008).
63. Sekar, A. *et al.* Schizophrenia risk from complex variation of complement component 4. *Nature* **530**, 177-+ (2016).
64. Cropley, V. *et al.* Increased Complement Factors C3 and C4 in Schizophrenia and the Early Stages of Psychosis: Implications for Clinical Symptomatology and Cortical Thickness. *Schizophrenia Bulletin* **44**, S74-S74 (2018).
65. Sharif, K. *et al.* The link between schizophrenia and hypothyroidism: a population-based

study. *Immunologic Research* **66**, 663-667 (2018).

66. Santos, N.C. *et al.* Revisiting Thyroid Hormones in Schizophrenia. *Journal of Thyroid Research* (2012).
67. Haudenschild, D., Moseley, T., Rose, L. & Reddi, A.H. Soluble and transmembrane Isoforms of novel interleukin-17 receptor-like protein by RNA splicing and expression in prostate cancer. *Journal of Biological Chemistry* **277**, 4309-4316 (2002).
68. Ding, M.L. *et al.* Activation of Th17 cells in drug naive, first episode schizophrenia. *Progress in Neuro-Psychopharmacology & Biological Psychiatry* **51**, 78-82 (2014).
69. Wurfel, B.E. *et al.* Serum kynurenic acid is reduced in affective psychosis. *Translational Psychiatry* **7**(2017).
70. Reus, G.Z. *et al.* Kynurenine pathway dysfunction in the pathophysiology and treatment of depression: Evidences from animal and human studies. *Journal of Psychiatric Research* **68**, 316-328 (2015).
71. Schwieler, L. *et al.* Electroconvulsive therapy suppresses the neurotoxic branch of the kynurenine pathway in treatment-resistant depressed patients. *Journal of Neuroinflammation* **13**(2016).
72. Stuart, M.J., Singhal, G. & Baune, B.T. Systematic review of the neurobiological relevance of chemokines to psychiatric disorders. *Frontiers in Cellular Neuroscience* **9**(2015).
73. Hong, S. *et al.* Abnormalities in chemokine levels in schizophrenia and their clinical correlates. *Schizophrenia Research* **181**, 63-69 (2017).
74. Black, C. & Miller, B.J. Meta-Analysis of Cytokines and Chemokines in Suicidality: Distinguishing Suicidal Versus Nonsuicidal Patients. *Biological Psychiatry* **78**, 28-37 (2015).
75. Maes, M. *et al.* Increased Plasma-Concentrations of Interleukin-6, Soluble Interleukin-6, Soluble Interleukin-2 and Transferrin Receptor in Major Depression. *Journal of Affective Disorders* **34**, 301-309 (1995).
76. Maes, M. *et al.* Elevated serum interleukin-6 (IL-6) and IL-6 receptor concentrations in posttraumatic stress disorder following accidental man-made traumatic events. *Biological Psychiatry* **45**, 833-839 (1999).
77. Newton, T.L., Fernandez-Botran, R., Miller, J.J. & Burns, V.E. Interleukin-6 and soluble interleukin-6 receptor levels in posttraumatic stress disorder: Associations with lifetime diagnostic status and psychological context. *Biological Psychology* **99**, 150-159 (2014).
78. Conley, R.R., Ascher-Svanum, H., Zhu, B.J., Faries, D.E. & Kinon, B.J. The burden of depressive symptoms in the long-term treatment of patients with schizophrenia. *Schizophrenia Research* **90**, 186-197 (2007).
79. Seow, L.S.E. *et al.* A systematic review on comorbid post-traumatic stress disorder in schizophrenia. *Schizophrenia Research* **176**, 441-451 (2016).
80. Burgess, S. & Labrecque, J.A. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. *European Journal of Epidemiology* **33**,

947-952 (2018).

81. Verbanck, M., Chen, C.Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nature Genetics* **50**, 693-+ (2018).

82. Wardenaar, K.J. & de Jonge, P. Diagnostic heterogeneity in psychiatry: towards an empirical solution. *Bmc Medicine* **11**(2013).

83. Schmidt, A.F. & Dudbridge, F. Mendelian randomization with Egger pleiotropy correction and weakly informative Bayesian priors. *Int J Epidemiol* (2017).

84. Burgess, S., Zuber, V., Valdes-Marquez, E., Sun, B.B. & Hopewell, J.C. Mendelian randomization with fine-mapped genetic data: Choosing from large numbers of correlated instrumental variables. *Genetic Epidemiology* **41**, 714-725 (2017).

Table 1 MR approach adopted according to the number of instrument SNPs

Number of SNP in Harmonized Data (*n*)	**2MR Test Applied**
$n = 1$	Wald Ratio
$n = 2$	Inverse variance weighted (IVW)

$n \geq 3$	If (pleiotropy *p*-value < 0.05)
	Egger's Regression
	else
	Inverse variance weighted

Table 2 MR results with plasma proteins as exposure and psychiatric disorders/traits as outcome (joint analysis of KORA and INTERVAL; top 30 results shown)

Meta-analysis (KORA,INTERVAL)		KORA					INTERVAL					HEIDI Test		Joint analysis		Cis-only		Aggregate ES
Protein TargetFullName (Exposure)	Psy Disorder (Outcome)	nsnp	Het.p	Method	Beta	P	nsnp	Het.p	Method	Beta	P	KORA	INTERVAL	Simes_p	max_p	Beta.cis	P.cis	TE.fixed
Endothelial cell-selective adhesion molecule	SCZ2	1	NA	Wald	0.204	1.50E-08	1	NA	Wald	0.285	1.50E-08	PASS	PASS	1.50E-08	1.50E-08	0.232	2.80E-15	0.232
Myeloid cell surface antigen CD33	Alz_IGAP	3	0.226	IVW	0.083	3.67E-08	1	NA	Wald	0.101	6.66E-08	NA	FAIL	6.66E-08	6.66E-08	0.090	1.60E-14	0.090
Plasma protease C1 inhibitor	SCZ2	3	0.875	IVW	-0.052	3.03E-05	1	NA	Wald	-0.084	2.02E-05	NA	FAIL	3.03E-05	3.03E-05	-0.061	6.35E-09	-0.061
Cathepsin B	NeuroticFull	1	NA	Wald	0.029	1.77E-04	1	NA	Wald	0.040	2.14E-05	FAIL	PASS	4.28E-05	1.77E-04	0.034	2.24E-08	0.034
Cathepsin S	SCZ2	2	0.071	IVW	0.072	1.28E-04	1	NA	Wald	0.105	2.67E-05	NA	FAIL	5.34E-05	1.28E-04	0.084	2.33E-08	0.084
Elafin	BIP-2018	1	NA	Wald	-0.169	1.02E-04	1	NA	Wald	-0.169	1.52E-04	FAIL	PASS	1.52E-04	1.52E-04	-0.169	5.74E-08	-0.169
Contactin-5	NeuroticFull	1	NA	Wald	-0.016	8.65E-02	2	9.61E-01	IVW	-0.030	8.37E-05	NA	NA	1.67E-04	8.65E-02	-0.020	2.04E-02	-0.025
Cathepsin B	ASD	1	NA	Wald	0.118	1.05E-04	1	NA	Wald	0.139	2.34E-04	PASS	PASS	2.10E-04	2.34E-04	0.127	9.90E-08	0.127
Lysozyme C	ASD	2	0.436	IVW	0.058	1.20E-04	1	NA	Wald	0.051	8.30E-02	NA	NA	2.39E-04	8.30E-02	0.057	2.51E-05	0.057
Ectonucleoside triphosphate diphosphohydrolase 5	OCD	2	0.461	IVW	0.180	2.80E-04	1	NA	Wald	0.234	9.80E-04	NA	PASS	5.59E-04	9.80E-04	0.198	1.14E-06	0.198
Death-associated protein kinase 2	NoOfDepEpi	1	NA	Wald	0.058	2.81E-04	1	NA	Wald	0.007	5.38E-01	(no result)	NA	5.62E-04	5.38E-01	NA	NA	0.025
Interleukin-6 receptor subunit alpha	SCZ2	3	0.108	IVW	0.029	3.19E-04	1	NA	Wald	0.023	1.73E-02	NA	NA	6.39E-04	1.73E-02	0.026	1.77E-05	0.026
Agouti-related protein	BIP-2018	1	NA	Wald	0.115	2.99E-03	1	NA	Wald	0.246	3.64E-04	NA	PASS	7.29E-04	2.99E-03	0.150	2.60E-06	0.147
Hemojuvelin	SelfHarmSuicide	1	NA	Wald	-0.855	4.33E-04	1	NA	Wald	0.262	5.49E-01	PASS	NA	8.66E-04	5.49E-01	NA	NA	-0.592
Neutral ceramidase	ADHD-2017_Euro	2	0.059	IVW	-0.078	2.28E-03	1	NA	Wald	-0.092	4.36E-04	NA	PASS	8.72E-04	2.28E-03	-0.085	3.51E-06	-0.085
Haptoglobin	ASD	7	0.015	IVW	-0.021	2.58E-01	1	NA	Wald	-0.070	4.36E-04	NA	FAIL	8.72E-04	2.58E-01	-0.045	1.20E-03	-0.045

Interleukin-7 receptor subunit alpha	NeuroticFull	1	NA	Wald	-0.002	7.39E-01	1	NA	Wald	0.058	4.65E-04	NA	PASS	9.31E-04	7.39E-01	NA	NA	0.007
MAP kinase-activated protein kinase 2	NeuroticFull	1	NA	Wald	0.006	6.17E-01	1	NA	Wald	-0.036	4.65E-04	NA	PASS	9.31E-04	6.17E-01	NA	NA	-0.019
Interleukin-12 receptor subunit beta-1	NeuroticFull	1	NA	Wald	0.007	2.11E-01	1	NA	Wald	0.064	4.65E-04	NA	PASS	9.31E-04	2.11E-01	NA	NA	0.011
Low molecular weight phosphotyrosine protein phosphatase	Alz_IGAP	3	0.644	IVW	-0.043	1.49E-03	1	NA	Wald	-0.049	5.10E-04	NA	PASS	1.02E-03	1.49E-03	-0.046	2.63E-06	-0.046
Endothelial cell-selective adhesion molecule	ASD	1	NA	Wald	0.145	1.92E-03	1	NA	Wald	0.203	1.92E-03	NA	PASS	1.92E-03	1.92E-03	0.165	1.51E-05	0.165
WAP, Kazal, immunoglobulin, Kunitz and NTR domain-containing protein 2	BIP-2018	1	NA	Wald	0.080	1.87E-03	1	NA	Wald	0.068	2.10E-03	NA	PASS	2.10E-03	2.10E-03	0.071	2.32E-05	0.073
SPARC-like protein 1	AnxCaseCtrl	2	0.058	IVW	0.122	5.71E-03	2	7.04E-01	IVW	0.180	1.09E-03	NA	NA	2.18E-03	5.71E-03	0.146	3.24E-05	0.144
Annexin A1	NeuroticFull	1	NA	Wald	-0.013	3.17E-01	1	NA	Wald	-0.021	1.15E-03	NA	PASS	2.31E-03	3.17E-01	NA	NA	-0.020
Serine/threonine-protein kinase 17B	NeuroticFull	1	NA	Wald	-0.028	1.96E-02	1	NA	Wald	-0.021	1.15E-03	NA	PASS	2.31E-03	1.96E-02	NA	NA	-0.023
Caspase-3	NeuroticFull	1	NA	Wald	-0.003	8.41E-01	1	NA	Wald	-0.046	1.15E-03	NA	PASS	2.31E-03	8.41E-01	NA	NA	-0.022
Integrin alpha-I: beta-1 complex	NeuroticFull	1	NA	Wald	0.014	3.17E-01	1	NA	Wald	-0.035	1.15E-03	NA	PASS	2.31E-03	3.17E-01	NA	NA	-0.017
Annexin A2	NeuroticFull	2	0.888	IVW	0.001	7.91E-01	1	NA	Wald	-0.015	1.15E-03	NA	PASS	2.31E-03	7.91E-01	NA	NA	-0.007
Plasma protease C1 inhibitor	MDD2018	3	0.901	IVW	-0.028	2.52E-03	1	NA	Wald	-0.045	2.24E-03	NA	FAIL	2.52E-03	2.52E-03	-0.033	2.82E-05	-0.033
Beta-endorphin	MDD-CON	1	NA	Wald	0.299	1.32E-03	1	NA	Wald	0.026	6.49E-01	NA	NA	2.63E-03	6.49E-01	NA	NA	0.101

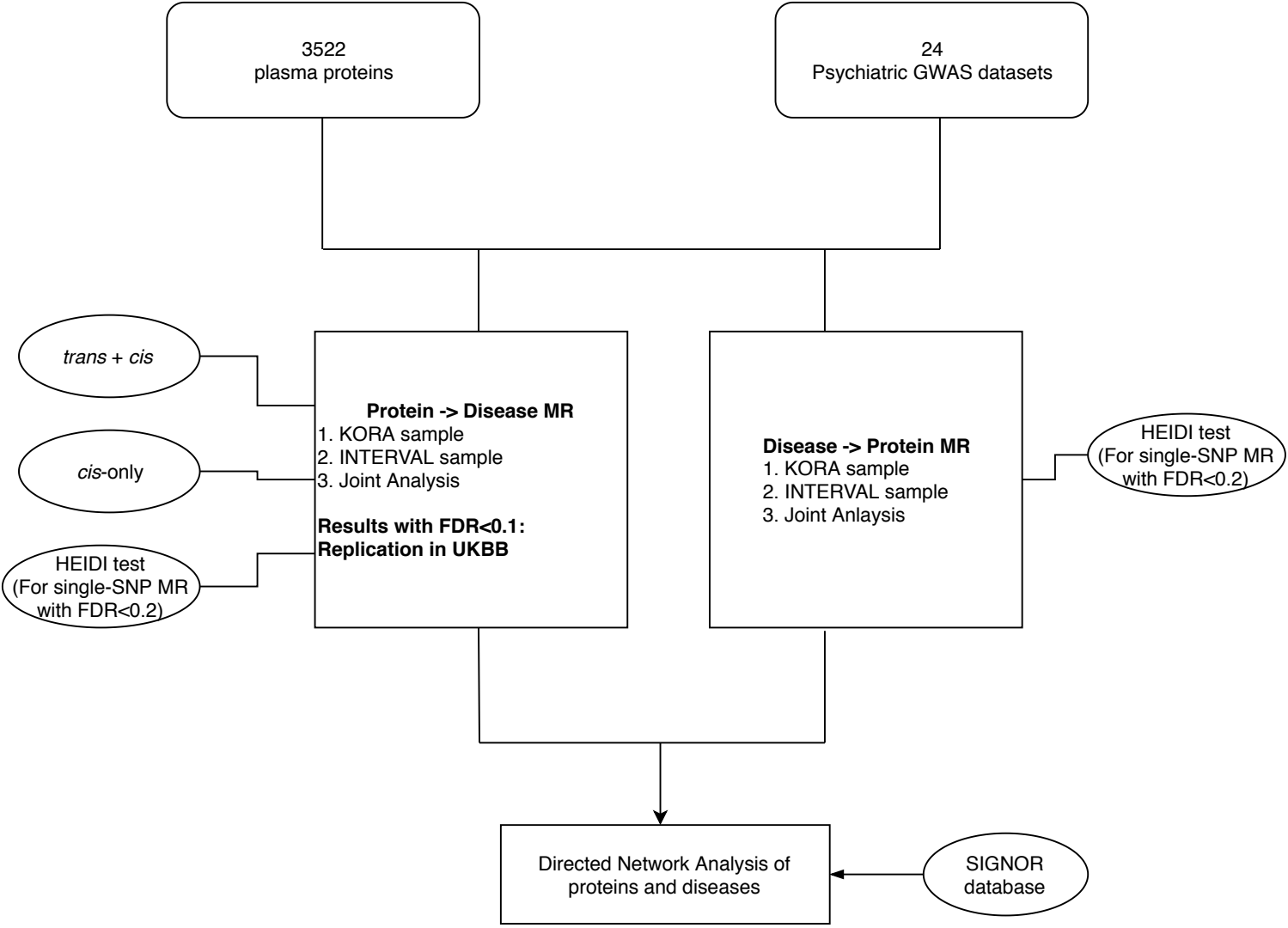
The above are sorted by the joint analysis results (by Simes test). Only the top 30 results are shown. Beta, MR causal estimate; Het.p, heterogeneity p-value; Simes_p, p-value from Simes' test; MAX_p, maximum p-value across two samples (test whether the hypothesis is non-null in every sample); Beta.cis and P.cis, MR estimate and corresponding p-value from cis-only analysis; TE.fixed, aggregate effect size based on fixed-effects model. Please refer to Table S1j for abbreviations of each disorder. The HEIDI test was only applied when nsnp = 1 & FDR-adjusted p-value<0.2 in the primary MR analysis.

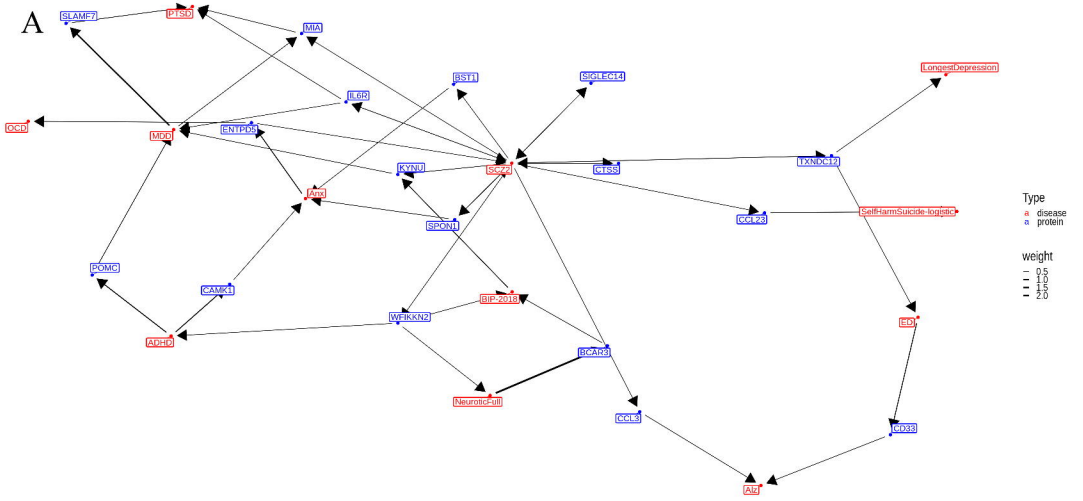
Table 3 MR results with psychiatric disorders/traits as exposure and proteins as outcome (joint analysis of KORA and INTERVAL)

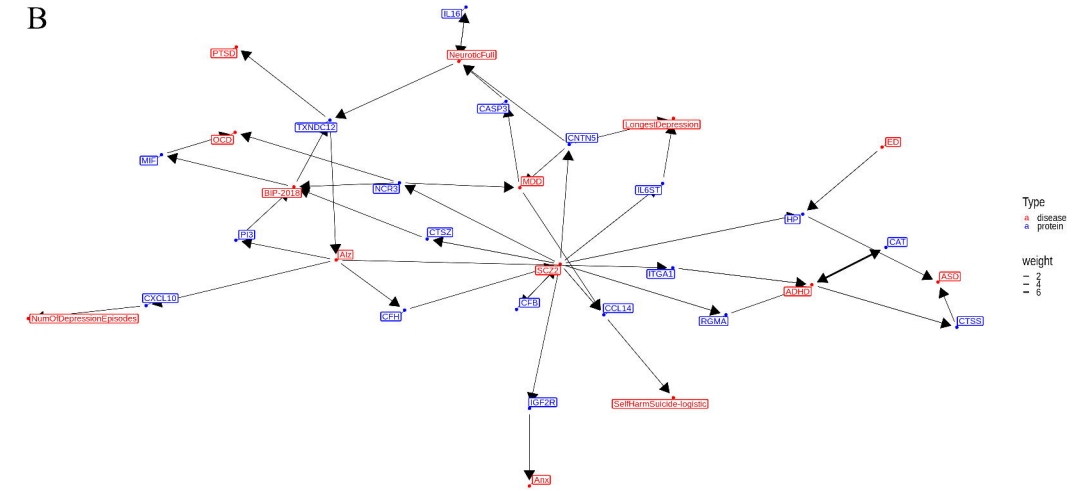
Psy disorder	Protein	KORA						Cambridge						Meta-analysis			
		nsnp	Pleio.p	Method	beta	se	pval	nsnp	Pleio.p	Method	beta	se	pval	Effect	SE	pval	qval.Storey
SCZ2	MHC class I polypeptide-related sequence B	86	0.001	Egger	1.265	0.311	4.75E-05	132	0.001	Egger	0.918	0.170	6.91E-08	0.998	0.149	2.33E-11	1.61E-07
Alz_IGAP	FACT complex subunit SSRP1	18	0.420	IVW	-0.037	0.065	5.71E-01	44	0.000	Egger	-0.117	0.020	2.51E-09	-0.111	0.019	4.35E-09	1.50E-05
SCZ2	cAMP-specific 3',5'-cyclic phosphodiesterase 4D	86	0.742	IVW	0.153	0.082	6.30E-02	132	0.271	IVW	0.434	0.073	2.23E-09	0.311	0.054	1.09E-08	1.92E-05
SCZ2	Interferon lambda-2	86	0.700	IVW	0.120	0.067	7.25E-02	132	0.703	IVW	0.168	0.031	4.66E-08	0.160	0.028	1.11E-08	1.92E-05
Alz_IGAP	Apolipoprotein E (isoform E2)	18	0.442	IVW	0.306	0.083	2.30E-04	44	0.092	IVW	0.126	0.029	9.48E-06	0.145	0.027	7.29E-08	1.01E-04
Alz_IGAP	Apolipoprotein B	18	0.262	IVW	0.137	0.075	6.94E-02	44	0.245	IVW	0.166	0.033	6.64E-07	0.161	0.031	1.29E-07	1.49E-04
Alz_IGAP	C-reactive protein	18	0.022	Egger	-0.167	0.064	9.26E-03	44	0.364	IVW	-0.107	0.023	3.66E-06	-0.114	0.022	1.63E-07	1.61E-04
SCZ2	Trypsin-3	86	0.597	IVW	-0.218	0.091	1.65E-02	132	0.456	IVW	-0.335	0.072	3.00E-06	-0.290	0.056	2.58E-07	2.22E-04
MDD2018	MHC class I polypeptide-related sequence B	3	0.086	IVW	2.468	0.562	1.13E-05	5	0.002	Egger	7.958	2.230	3.59E-04	2.796	0.545	2.89E-07	2.22E-04
Alz_IGAP	Chordin-like protein 1	18	0.316	IVW	-0.120	0.046	8.68E-03	44	0.008	Egger	-0.080	0.019	2.62E-05	-0.086	0.018	9.97E-07	6.89E-04
Alz_IGAP	Low-density lipoprotein receptor-related protein 8	18	0.501	IVW	-0.059	0.056	2.94E-01	44	0.025	Egger	-0.093	0.020	3.29E-06	-0.089	0.019	2.21E-06	1.27E-03
Alz_IGAP	Interleukin-22 receptor subunit alpha-1	18	0.805	IVW	-0.038	0.054	4.90E-01	44	0.117	IVW	0.146	0.026	2.48E-08	0.112	0.024	2.33E-06	1.27E-03
Alz_IGAP	N-acylethanolamine-hydrolyzing acid amidase	18	0.354	IVW	-0.109	0.054	4.34E-02	44	0.040	Egger	-0.088	0.021	1.87E-05	-0.091	0.019	2.39E-06	1.27E-03
Alz_IGAP	Granulocyte colony-stimulating factor receptor	18	0.714	IVW	0.097	0.061	1.10E-01	44	0.026	Egger	0.085	0.019	1.00E-05	0.086	0.018	2.67E-06	1.29E-03
SCZ2	Interleukin-5 receptor subunit alpha	86	0.558	IVW	0.073	0.091	4.24E-01	132	0.528	IVW	0.149	0.032	2.81E-06	0.141	0.030	2.79E-06	1.29E-03
SCZ2	Fibroblast growth factor 20	86	0.671	IVW	0.078	0.071	2.74E-01	132	0.969	IVW	0.143	0.032	7.85E-06	0.132	0.029	6.04E-06	2.61E-03
Alz_IGAP	Diablo homolog, mitochondrial	18	0.816	IVW	-0.085	0.053	1.10E-01	44	0.031	Egger	-0.079	0.019	2.86E-05	-0.079	0.018	7.55E-06	2.77E-03
Alz_IGAP	C-C motif chemokine 22	18	0.203	IVW	0.016	0.066	8.06E-01	44	0.007	Egger	-0.098	0.021	2.00E-06	-0.088	0.020	8.03E-06	2.77E-03
SCZ2	Interleukin-19	86	0.548	IVW	0.110	0.074	1.39E-01	132	0.260	IVW	0.167	0.039	2.00E-05	0.155	0.035	8.13E-06	2.77E-03
SCZ2	Secreted frizzled-related protein 1	86	0.053	IVW	0.228	0.069	9.81E-04	132	0.416	IVW	0.108	0.032	6.83E-04	0.129	0.029	8.14E-06	2.77E-03
SCZ2	Thyroid Stimulating Hormone	86	0.075	IVW	0.088	0.068	1.93E-01	132	0.538	IVW	0.156	0.036	1.37E-05	0.141	0.032	8.47E-06	2.77E-03
SCZ2	WAP, Kazal, immunoglobulin, Kunitz and NTR domain-containing protein 2	86	0.162	IVW	0.029	0.063	6.52E-01	132	0.095	IVW	0.161	0.034	1.65E-06	0.132	0.030	8.83E-06	2.77E-03
SCZ2	Bone morphogenetic protein 7	86	0.258	IVW	0.028	0.080	7.26E-01	132	0.301	IVW	0.125	0.028	6.84E-06	0.115	0.026	1.27E-05	3.81E-03

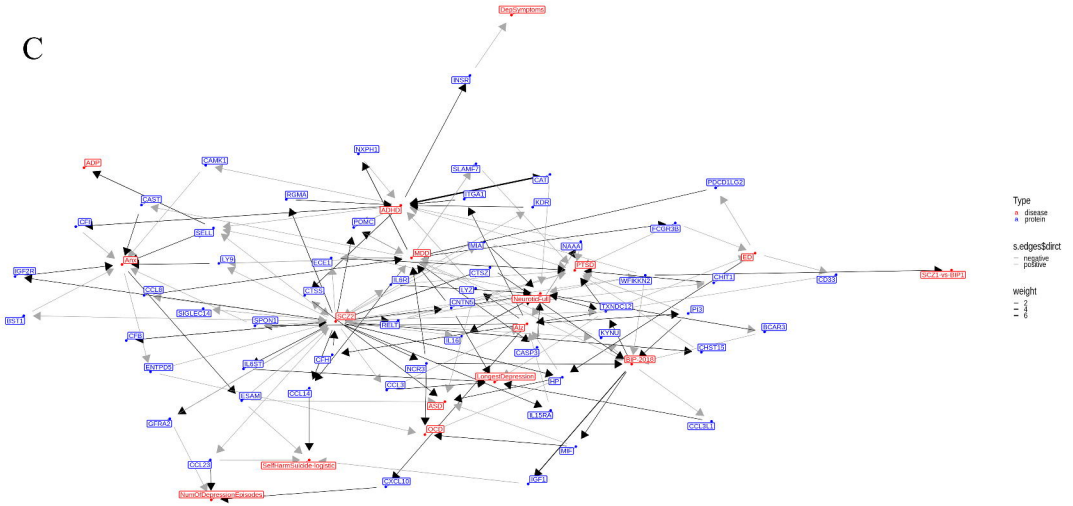
Alz_IGAP	Pleiotrophin	18	0.384	IVW	-0.068	0.048	1.58E-01	44	0.037	Egger	-0.082	0.020	4.04E-05	-0.080	0.018	1.48E-05	4.26E-03
SCZ2	Complement C4	86	0.003	Egger	0.765	0.309	1.32E-02	132	0.016	Egger	0.528	0.147	3.27E-04	0.572	0.133	1.64E-05	4.27E-03
Alz_IGAP	Integrin alpha-I: beta-1 complex	18	0.300	IVW	-0.066	0.053	2.15E-01	44	0.04825	Egger	-0.082	0.020	3.54E-05	-0.080	0.019	1.65E-05	4.27E-03
SCZ2	Interleukin-17 receptor C	86	0.012	Egger	-0.727	0.280	9.58E-03	132	0.32223	IVW	-0.124	0.031	5.18E-05	-0.131	0.031	1.67E-05	4.27E-03
Alz_IGAP	Transcription factor IIIB 90 kDa subunit	18	0.367	IVW	0.105	0.054	5.29E-02	44	0.05498	IVW	0.082	0.021	1.29E-04	0.085	0.020	1.94E-05	4.69E-03
SCZ2	Low affinity immunoglobulin epsilon Fc receptor	86	0.182	IVW	-0.160	0.075	3.27E-02	132	0.50051	IVW	-0.166	0.045	2.19E-04	-0.164	0.038	1.97E-05	4.69E-03
SCZ2	Fibroblast growth factor receptor 4	86	0.003	Egger	0.780	0.287	6.53E-03	132	0.7263	IVW	0.129	0.033	8.93E-05	0.138	0.033	2.63E-05	6.04E-03

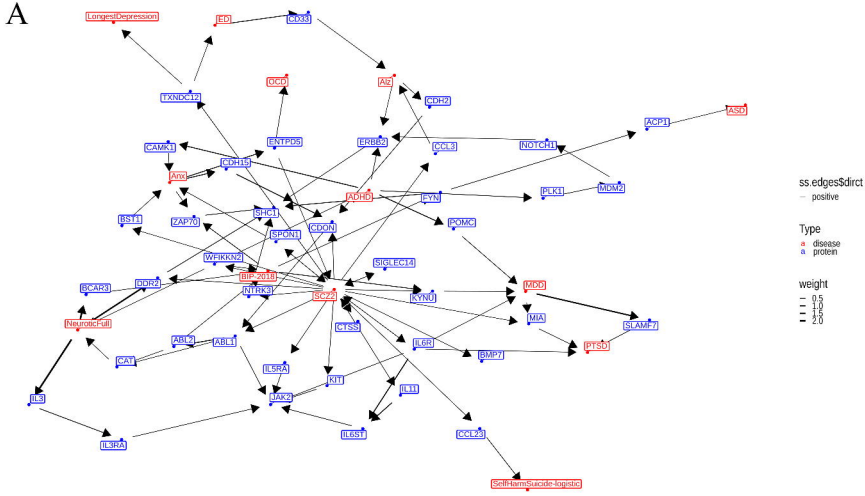
Please also refer to legends of Table 2. Please refer to Table S1j for abbreviations of each disorder. qval.Stroey: q-value based on the Storey method.

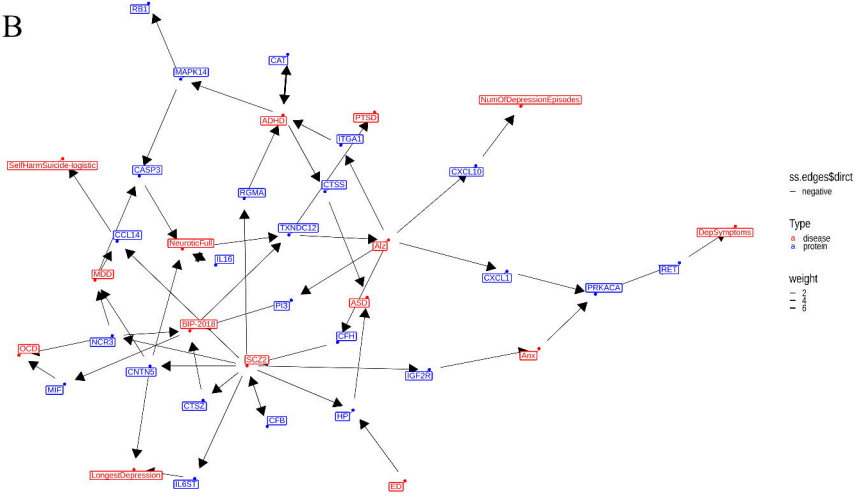

3522
plasma proteins
24
Psychiatric GWAS datasets
trans + cis
cis-only
HEIDI test
(For single-SNP MR
with FDR<0.2)
Protein -> Disease MR
1. KORA sample
2. INTERVAL sample
3. Joint Analysis
Results with FDR<0.1:
Replication in UKBB
Disease -> Protein MR
1. KORA sample
2. INTERVAL sample
3. Joint Anlaysis
HEIDI test
(For single-SNP MR
with FDR<0.2)
Directed Network Analysis of
proteins and diseases
SIGNOR
database


A
SLAMF7
PTSD
MIA
BST1
SIGLEC14
LongestDepression
IL6R
OCD
MDD
ENTPD8
KYNU
SCZ2
CTSS
TXNDC17
Alzh
CCL23
SelfHarmSuicide-logistic
SPON1
POMC
CAMK1
BIP-2018
WFIKKN2
ED
ADHD
BCAR3
NeuroticFull
CCL3
MAP3K8
Alz
Type
disease
protein
weight
0.5
1.0
1.5
2.0


B
IL1B
PTSD
NeuroticFull
CASP3
TXNDC17
LongestDepression
OCD
CNTN5
ED
MIF
BIP-2018
NCR3
MDD
IL6ST
LIF
PS
CTSZ
CAT
Alz
SCZ2
ITGA1
ASD
CXCL10
ADHD
NumOfDepressionEpisodes
CFH
CFB
CCL14
RGMA
CTSS
IGF2R
SelfHarmSuicide-logistic
Anx
Type
disease
protein
weight
2
4
6


C
DepSymptoms
RGS8
ADP
CAMK1
NXPH1
SLAMF7
CAT
PCDH10OS
CASP3
RGMA
ITGA1
ADR2
CFI
SFRP1
POMC
PCDH9
ADP
LY9
MDD
MIA
NXXA1
PTSD
ED
IGF2R
CCL8
CTSZ
SCZ2_vs_BIP1
CCL8
IL1B
LY7
WFIKKN2
CHIT1
CSS3
CNTN5
NeuroticFull
TXNDC17
PS
IBST1
SIGLEC14
SPON1
SCZ2
RELM
Alz
KYNU
BCAR3
CFB
IL1B
CASP3
CHST10
ENTPD8
IL6R
CFP
NCR3
LongestDepression
CCL8
CFP
ESAM
CCL13
Anx
IL1RA
CCL8L1
CrfrA2
ASD
MIF
CCL22
SelfHarmSuicide-logistic
CCL16
IGF1
NumOfDepressionEpisodes
Type
disease
protein
s.edges$direct
negative
positive
weight
2
4
6


A
LongestDepression
ED
CD33
OCD
Alz
TXNDC12
CDH2
ASD
ACP1
CAMK1
ENTPD5
ERBB2
CCL3
NOTCH1
CDH15
Anx
ADHD
FYN
PLK1
MDM2
BST1
ZAP70
SHC1
CDON
POMC
SPON1
WFIKKN2
SIGLEC14
BIP-2018
BCAR3
DDR2
NTRK3
SCZ2
KYNU
MDD
NeuroticFull
CTSS
MIA
SLAMF7
ABL2
ABL1
IL6R
PTSD
CAT
BMP7
IL5RA
KIT
IL3
JAK2
IL11
IL6ST
CCL23
IL3RA
SelfHarmSuicide-logistic
ss.edges$dirct
positive
Type
disease
protein
weight
0.5
1.0
1.5
2.0


B
RB1
CAT
MAPK14
NumOfDepressionEpisodes
ADHD
PTSD
ITGA1
SelfHarmSuicide-logistic
CASP3
CXCL10
CTSS
RGMA
TXNDC12
CCL14
NeuroticFull
Alz
DepSymptoms
IL16
MDD
CXCL1
RET
PRKACA
PI3
ASD
BIP-2018
CFH
OCD
NCR3
SCZ2
Anx
CNTN5
IGF2R
MIF
CTS2
CFB
HP
LongestDepression
IL6ST
ED
ss.edges$dirct
negative
Type
disease
protein
weight
2
4
6