Genetic estimates of correlation and causality between blood-based biomarkers and psychiatric disorders

William R. Reay^{1,2}, Dylan J. Kiltschewskij^{1,2}, Michael P. Geaghan^{1,2}, Joshua R. Atkins¹,
Vaughan J. Carr^{3,4,5}, Melissa J. Green^{3,4}, Murray J. Cairns^{1,2} *

¹School of Biomedical Sciences and Pharmacy, The University of Newcastle, Callaghan,
NSW, Australia

²Centre for Brain and Mental Health Research, Hunter Medical Research Institute,
Newcastle, NSW, Australia

³School of Psychiatry, University of New South Wales, Randwick, NSW, Australia

⁴Neuroscience Research Australia, Sydney, NSW, Australia

⁵Department of Psychiatry, Monash University, Melbourne, VIC, Australia

To whom correspondence should be addressed: Professor Murray Cairns

(Murray.Cairns@newcastle.edu.au)

Running title: Biochemical traits and psychiatric disorders

Keywords: genome-wide association study, genetic correlation, causal inference,
biochemical traits, pleiotropy, psychiatric illness

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

ABSTRACT

There is a long-standing interest in exploring the relationship between blood-based biomarkers of biological exposures and psychiatric disorders, despite their causal role being difficult to resolve in observational studies. In this study, we leverage genome-wide association study data for a large panel of heritable biochemical traits measured from serum to refine our understanding of causal effect in biochemical-psychiatric trait pairings. In accordance with expectation we observed widespread evidence of positive and negative genetic correlation between psychiatric disorders and biochemical traits. We then implemented causal inference to distinguish causation from correlation and found strong evidence that C-reactive protein (CRP) exerts a causal effect on psychiatric disorders, along with other putatively causal relationships involving urate and glucose. Strikingly, these analyses suggested CRP has a protective effect on three disorders including anorexia nervosa, obsessive-compulsive disorder, and schizophrenia, whilst being a risk factor for major depressive disorder. Multivariable models that conditioned CRP effects on interleukin-6 signalling and body mass index suggested that CRP-schizophrenia relationship was not likely mediated by those factors. Collectively, these data suggest that there are shared pathways that influence both biochemical traits and psychiatric illness, including factors such as CRP that are likely to constitute a causal effect and could be targets for therapeutic intervention and precision medicine.

MAIN TEXT

INTRODUCTION

Psychiatric disorders arise from a complex interplay between genetic and environmental risk factors. Indeed, twin-based and genome-wide association (GWAS) studies' heritability estimates have demonstrated the importance of genetic risk to the spectrum of psychiatric illness (1–3). In particular, GWAS have been successful in identifying regions of the genome associated with psychiatric disorders, as well as revealing both overlapping and distinct features amongst the genetic architecture of these traits (2–6). For example, our group previously demonstrated that several genes associated with schizophrenia were shared with other psychiatric disorders, along with genes that appeared more specifically linked to schizophrenia (4). The challenge for psychiatric genetics from here onwards is to integrate and expand these data such that the biological insights gained may be directly relevant for psychiatric practice.

GWAS has proven valuable beyond just gene discovery in psychiatry, in that it allows the study of relationships between sets of traits in terms of genetic correlation (7), as well as GWAS informed methods for causal inference (8,9). An area of continued interest is the interplay between circulating biochemical factors and the pathophysiology of psychiatric disorders (10–13). These studies have endeavoured to find biochemical traits readily detectable in blood, that, in theory, could be diagnostic or prognostic biomarkers for a given psychiatric disorder. Many of these hypotheses stem from the idea that peripheral biochemical traits may exert an effect on the brain, directly or indirectly through their effect on other mediators, and that the manifestation of mental illness is in part attributed to these factors which primarily act in the periphery (11,12,14–16). Identifying these biochemical-psychiatric relationships would be clinically valuable as many of these traits can be modulated by existing drugs and/or lifestyle

interventions. However, progress in this field has been hampered by small sample size studies, along with the fact that the majority of these studies are observational in nature, and thus, likely subject to at least some confounding. Genetics offers an attractive prospect for studying biochemical traits in psychiatry as many such measures are heavily influenced by genetic factors, with germline genetic variants fixed at birth and immune to reverse causation in most instances. In this study, we attempt to harmonise inter-study variability by using a panel of large sample size ($N > 300,000$) biochemical GWAS from a single cohort (UK biobank) to investigate genetic overlap with different psychiatric disorders, along with putative causal effects. We found that the majority of biochemical traits tested were genetically correlated with at least one psychiatric trait, with evidence of convergent and divergent correlation profiles amongst the different disorders. Interestingly, we also demonstrated evidence that there may be a causal relationship on psychiatric phenotypes through circulating C-reactive protein (CRP), glucose, and urate – which may have direct implications for clinical practice.

METHODS AND MATERIALS

Psychiatric genome-wide association studies

GWAS summary statistics for nine European ancestry cohorts for the following disorders were obtained from the psychiatric genomics consortium (PGC): attention/deficit hyperactivity disorder (ADHD) (17), anorexia nervosa (AN) (18), autism spectrum disorder (ASD) (19), bipolar disorder (BIP) (20), major depressive disorder (MDD) (21), obsessive compulsive disorder (OCD), post-traumatic stress disorder (PTSD) (22), schizophrenia (SZ) (23), and Tourette's syndrome (TS) (24). Given that cognitive symptoms are pervasively associated with psychiatric illness, we also included a GWAS of general cognitive ability (25). Further information regarding these studies is provided in the supplementary text.

Blood-based biomarker genome-wide association studies

We obtained GWAS summary statistics for a series of blood-based biochemical traits from the large UK biobank (UKBB) sample performed by the Neale group (<http://www.nealelab.is/uk-biobank>). The key advantage of these data is its large sample size ($N > 300,000$) and that the biochemical traits analysed were obtained from a single large cohort. Specifically, we utilised a panel of 50 biochemical GWAS which had high or medium confidence estimates of SNP heritability that were significantly different from zero as outlined in supplementary table 1 and supplementary text. These traits included lipids, micronutrients, hormones, metabolites, and enzymes.

Genetic correlation

Genetic correlation between each psychiatric and biochemical trait was estimated using linkage disequilibrium score regression (LDSR) (7), with summary statistic cleaned ('munged') to contain around one million HapMap 3 SNPs outside the major histocompatibility complex

(MHC) with minor allele frequency > 0.05 for consistency (<https://github.com/bulik/ldsc>). Briefly, LDSR estimates genetic covariance by regressing SNP-wise χ^2 , the product of the marginal SNP effects from both traits ($Z_1 Z_2$), on its LD score, which is an estimate of total LD existing with that SNP. Trait heritabilities are utilised to normalise the genetic covariance to obtain genetic correlation (r_g). A key advantage of LDSR is that sample overlap only affects the LDSR intercept and not the slope, meaning we can accurately estimate r_g between UKBB biochemical GWAS and psychiatric GWAS with UKBB samples included. We used the Bonferroni method to correct for the fifty traits tested. One biochemical trait was excluded from further analysis, Apolipoprotein B, as it exhibited negative heritability within the block jackknifing procedure to estimate the r_g standard error in some instances. The resulting 49 x 10 matrix of LDSR r_g , divided by its standard error to obtain Z , was subjected to a latent clustering method, finite Gaussian mixture modelling (GMM), with the mclust R package version 5.4.6 (26). We selected the most parsimonious clustering configuration based on parametrisation of the covariance matrix utilising the largest Bayesian Information Criterion (BIC) value.

Biochemical polygenic scoring in a severe cognitive deficit subtype of schizophrenia

We sought to further investigate biochemical traits displaying psychiatric genetic correlation and examine their relevance to the clinical dimensions of psychiatric disorders. Specifically, we considered the heterogeneity of cognitive impairment observed in schizophrenia, wherein deficits often manifest before the first psychotic episode and are highly variable in their presentation throughout clinical course (27,28). We considered biochemical traits that were genetically correlated with either schizophrenia or cognition after the application of multiple testing correction and interrogated their relationship with severe cognitive deficit in a cohort of schizophrenia cases from the Australian Schizophrenia Research Bank (ASRB) cohort. The

use of these data was approved by the University of Newcastle Human Research Ethics Committee (HREC) and the ASRB (28–31). Previously, Green *et al.* utilised multidimensional Grade of Membership (GoM) clustering with nine cognitive measures to derive subgroups of cognitive performance in the ASRB (28). The most parsimonious configuration were two clusters of schizophrenia cases termed cognitive deficit (CD), with more pervasive cognitive impairment, and cognitively spared (CS), displaying intermediate cognitive performance relative to CD cases and healthy controls. Polygenic scores (PGS) were constructed for the 25 biochemical traits correlated with either schizophrenia or cognitive ability in a genotyped subset of the ASRB with schizophrenia cases subtyped as CD or CS (N = 391, Supplementary materials). The full details of this cohort and the generation of PGS is described in the supplementary text. We tested the association of each biochemical PGS with CD status using binomial logistic regression covaried for sex and the first three SNP-derived principal components (Supplementary Materials). The variance explained (Nagelkerke’s R^2) in the full model with the PGS versus the null (covariates and intercept only) model, was converted to the liability scale assuming a population prevalence for CD of 0.33% (32). The population prevalence for CD is somewhat arbitrary, however, given the population prevalence of schizophrenia is around 0.7%, and 43% of this portion of the schizophrenia cases in the ASRB cohort was subtyped as CD, we believe this was an appropriate value to select. A polygenic score for general cognitive ability and a schizophrenia polygenic risk score (PRS) was also derived in this cohort for comparison (Supplementary Materials).

Latent causal variable models

Genetic correlation may reveal important insights into shared biology between two traits, however, this should not be interpreted as implying a causal relationship in either direction. To evaluate evidence for a causal relationship, we implemented the latent causal variable (LCV)

model to estimate genetic causality between traits, as outlined extensively elsewhere (9,33). The LCV framework leverages the bivariate genome-wide distribution of marginal SNP effects on both traits to estimate *partial genetic causality*. Specifically, the LCV assumes a latent variable, L , mediates the genetic correlation between the traits, and tests the strength of the correlation of each trait with L . The mixed fourth moments (cokurtosis) of marginal effect sizes for each trait (Z_1, Z_2) are compared to evaluate the proportionality of effects on either trait. The main output of the LCV model is the posterior mean genetic causality proportion ($\widehat{GCP}$), whereby $\widehat{GCP} > 0$ implies partial genetic causality of trait one on trait two, and *vice versa*. In other words, given $\widehat{GCP} > 0$, then trait one SNP effects (Z_1^2) tend to be proportionally large on trait two (Z_1Z_2), such that $|E(Z_1^2 Z_1Z_2)| \geq |E(Z_2^2 Z_1Z_2)|$. As in LDSR, the LDSR intercept is utilised here to guard against inflation due to sample overlap. We defined *partial genetic causality* using the recommended threshold of a significantly non-zero $|\widehat{GCP}| > 0.6$, as this was shown by O'Connor and Price in simulations to guard against false positives (9). An LCV model was constructed for all genetically correlated psychiatric-biochemical trait pairs. Weak GCP estimates close to zero for genetically correlated traits imply that their relationship is potentially mediated by horizontal pleiotropy, whereby there are shared pathways, but the two traits do not likely exhibit vertical pleiotropy by acting within the same pathway. We also attempted to replicate the observed $\widehat{GCP}$ using a non-UKBB biochemical GWAS (34–37). It should be noted that the posterior mean $\widehat{GCP}$ is not an estimate of the magnitude of any potential causal relationship and should not be interpreted as such – rather it evaluates the strength of evidence for a putative causal relationship using genome-wide SNP effects. The scripts to construct an LCV model are available from (<https://github.com/lukejconnor/LCV>).

Mendelian randomisation

C-reactive protein (CRP) exhibited strong evidence of partial genetic causality on multiple psychiatric disorders, and thus, we sought to estimate the magnitude of this causal relationship using univariable and multivariable Mendelian randomisation (MR). A detailed description of the MR methodology in this study is provided in the supplementary materials. MR is different from the LCV approach in that it leverages independent SNPs strongly associated with trait one (the exposure trait) as instrumental variables (IVs) to estimate the effect of the exposure on an outcome. The theoretical justification for using SNPs as IVs has been discussed extensively previously (8,38). We estimated the total effect of CRP on each disorder using independent genome-wide significant SNPs from a smaller non-UKBB GWAS as sample overlap between exposure and outcome can bias MR estimates (Supplementary Materials) (34). The F -statistic for IVs for this CRP GWAS was sufficiently strong (F -statistic > 10), as reported previously (34).

Our primary model was an inverse-variance weighted effect (IVW) estimator with multiplicative random effects, which assumes all IVs are valid and is less biased by heterogeneity than a fixed-effects IVW estimator (8). Whilst the IVW estimator is generally considered the most well-powered approach, the assumption that all IVs are valid is probably unrealistic in practice. As a result, we implemented a series of models which make different underlying assumptions regarding IV validity. Specifically, median based estimators which assume the majority of IVs are valid (39); a weighted mode estimator and a contamination mixture model, that both assume out of groups of IVs having the same asymptotic estimate, the largest group will be comprised of valid IVs (plurality valid) (40,41); and MR-Egger, which includes a non-zero intercept as a test of the average pleiotropic effect and assumes that there is no significant correlation between direct IV effects on the outcome and genetic association

of IVs with the exposure (Instrument Strength Independent of Direct Effect (InSIDE) assumption) (42). As recommended by Bowden *et al.*, we ensured that the I^2 statistic of IV-exposure effects exceeded 0.9, as this assesses the relative strength of the no-measurement error assumption, and thus, the suitability of using an MR-Egger model (43). We also tested the effect of using robust regression and penalised estimates for heterogeneity on the IVW, median, and Egger regression estimates (44). Evidence of horizontal pleiotropy and outliers were further investigated by testing heterogeneity in the IV/exposure-outcome effects (45), performing a leave-one-IV out analysis (38), testing for a non-zero MR Egger intercept (42), and an MR PRESSO test of global pleiotropy (also related to heterogeneity) (46). Furthermore, we also tested whether there was evidence of a causal effect in the reverse direction (psychiatric disorder as exposure), although MR estimates with binary exposures should be treated cautiously, as described elsewhere (47,48). Given only approximately 2,000,000 SNPs were available in the non-UKBB CRP GWAS utilised, we utilised the more deeply imputed UKBB CRP GWAS as the outcome here, although these results could therefore be inflated by sample overlap for AN and MDD.

Multivariable MR (MVMR) was then performed to evaluate the direct effect of CRP on each psychiatric outcome tested when conditioned on BMI and interleukin six (IL-6) signalling, which are both postulated to be closely functionally related to circulating CRP (Supplementary Text) (49,50). MVMR assumes that IVs are strongly associated with at least one exposure, and therefore, SNPs were chosen which were genome-wide significant for at least one phenotype. We constructed an IL-6 and BMI multivariable model separately – that is CRP conditioned on circulating IL-6 and its receptor (IL6R), and CRP conditioned on BMI, as well as BMI and IL-6R. The strength of the IVs in each multivariable model was assessed using a two-sample conditional F statistic – which tests whether the IVs strongly predict each exposure, conditional

on the other exposures in the model (51,52). When an F statistic > 10 could not be achieved, we relaxed the SNP inclusion threshold to represent suggestive significance in the GWAS ($P_{\text{GWAS}} < 1 \times 10^{-5}$, Supplementary Text). We compared direct estimates for each exposure using four different MVMR models – an IVW MVMR estimator, a median based MVMR estimator, an Egger regression based MVMR estimator, and a regularisation approach whereby LASSO type penalisation is applied to shrink intercept terms corresponding to IVs predicted as valid (MVMR-LASSO) (53). Given that CRP was genetically correlated with several psychiatric GWAS, genetic correlation may result in bias in MR estimates (9). However, MR is a valuable extension to the LCV model as it allows for specification of several different assumptions about IV validity and facilitates the estimate of total (univariable MR) and direct (MVMR) effects. The MR analyses were performed using the following packages in R version 3.6.0 – TwoSampleMR v 0.5.5 (54), MendelianRandomization v 0.5 (55), MRPRESSO v 1.0 (46), and MVMR v 0.3 (51).

Genetic overlap between C-reactive protein and schizophrenia

We investigated whether any of the lead SNPs (genome-wide significant) reported in the PGC3 SZ GWAS were also associated with CRP ($P < 5 \times 10^{-8}$). For overlapping genome-wide significant signals, we tested whether there was a shared causal variant or different causal variants underlying these loci, assuming a single causal variant, using the *coloc* package colocalisation method (56). Moreover, we utilised ρ -HESS (<https://github.com/huwenboshi/hess>) to estimate local genetic covariance between SZ and CRP, as opposed to a genome-wide estimate by LDSR, with local genetic covariance ($\rho_{g,local}$) calculated for munged HapMap3 SNPs in around 1,600 approximately independent LD blocks outside the MHC (57,58). Genes within the five LD blocks with the most statistically significant $\rho_{g,local}$ were subjected to pathway analysis using g:Profiler

(<https://biit.cs.ut.ee/gprofiler/gost>) (59). We derived an estimate of genetic correlation ($r_{g,local}$) by dividing $\rho_{g,local}$ by the product of the square roots of CRP and SZ local heritability per LD block, respectively.

Downstream effects of C-reactive protein

We investigated the downstream effect of CRP on circulating levels of 3284 proteins in blood using MR (60). We used the larger UKBB CRP GWAS to select IVs to maximise power. The principal MR model was the IVW estimator with multiplicative random effects to maximise power. Proteins which demonstrated at least nominal association with CRP levels after multiple testing correction ($FDR < 0.1$) were investigated for protein-protein interaction and overrepresentation in biological pathways using STRING v 11.0 (61).

RESULTS

Widespread genetic correlation between blood-based biomarkers and psychiatric traits

We tested the genetic correlation between a panel of blood-based biomarkers and 10 psychiatric GWAS using LDSR. Interestingly, we found that 61% (N = 30) of the biochemical traits tested were significantly correlated with at least one psychiatric trait after multiple-testing correction, with every psychiatric trait exhibiting a significantly non-zero biochemical correlation after correction except for Tourette's syndrome (Figure 1a, Supplementary Tables 2 – 11) The most significantly correlated biomarker for each trait is outlined in table 1.

Table 1: The most significant genetic correlation between each trait and a biochemical GWAS

Biochemical trait	Psychiatric trait ¹	r_g^2	SE	P
High light scatter reticulocyte percentage	ADHD	0.256	0.033	4.58×10^{-15}
High light scatter reticulocyte count	OCD	-0.212	0.046	3.37×10^{-6}
C-reactive protein	AN	-0.286	0.038	9.76×10^{-14}
Vitamin D	ASD	-0.177	0.047	2×10^{-4}
Creatinine	BIP	-0.106	0.027	7.23×10^{-5}
Leukocyte count	Cognition	-0.198	0.024	2.66×10^{-16}
Leukocyte count	MDD	0.155	0.025	9.77×10^{-10}
Leukocyte count	PTSD	0.225	0.043	1.37×10^{-7}
Lymphocyte count	SZ	0.075	0.017	1.52×10^{-5}
Lymphocyte count	TS	-0.093	0.039	0.017

¹Psychiatric GWAS: attention/deficit hyperactivity disorder (ADHD), anorexia nervosa (AN), autism spectrum disorder (ASD), bipolar disorder (BIP), major depressive disorder (MDD), obsessive compulsive disorder (OCD), post-traumatic stress disorder (PTSD), schizophrenia (SZ), and Tourette's syndrome (TS).

2r_g is the estimate of genetic correlation from LDSR

There was clear evidence of biomarkers with divergent genetic correlations between different psychiatric phenotypes – for instance, CRP, alanine aminotransferase (ALT), and sex-hormone binding globulin (SHBG, Figure 1b). In the case of CRP, it displayed a negative correlation after correction with OCD, AN, and cognition, as well as a trend towards a negative correlation with schizophrenia ($P = 1.7 \times 10^{-3}$), whilst its correlation with PTSD, MDD, and ADHD was strongly positive. These data provide some support to recent cohort studies, including lower CRP observed in patients with eating disorder (62), whilst elevated CRP was associated with ADHD (63). Notably, this contradicts previous observational estimates of elevated CRP in schizophrenia (64). Moreover, there were 14 biochemical traits which were only correlated after multiple testing correction with two or fewer psychiatric GWAS. A few such of examples of these more specific correlations included albumin and ADHD ($r_g = -0.151$), mean corpuscular volume and AN ($r_g = 0.087$), mean spherized cell volume and SZ ($r_g = 0.06$), and creatinine with BIP and SZ (SZ: $r_g = -0.07$, BIP: $r_g = -0.106$).

We further investigated the relationship between the profile of 49 LDSR biochemical Z scores for each psychiatric trait. We observed strong positive and negative correlations between the trait-wise LDSR biochemical Z for each trait, which we term the *biochemical correlation profile* (Figure 1c). For instance, the ADHD biochemical correlation profile demonstrates large magnitude positive correlations with ASD, MDD, and PTSD but negative correlations with OCD, AN, and Cognition. This can be interpreted as traits which tend to be positively correlated with ADHD are also positively correlated with ASD, MDD, and PTSD and *vice versa* for OCD, AN, and Cognition. These relationships were further interrogated by subjecting the 10 biochemical correlation profiles to finite Gaussian mixture modelling (GMM, Figure

1d). We observed five components (clusters) with diagonal distribution, variable volume, and equal shape as the most parsimonious parameterisation of the covariance matrix (Supplementary Table 12). The biochemical correlation profile LDSR Z within each cluster are plotted in figure 1e. Briefly, the first component was composed of a series of traits with discordant correlations between the disorders, such as CRP, ALT, and glycaeted haemoglobin (HbA1c), whilst the second and third component were a diverse set of biomarkers with more similar LDSR Z across the psychiatric disorders tested. Component four was notable as it was solely composed of reticulocyte (immature erythrocytes) related traits, which analogous to component one was quite discordant in its correlation profiles. The fifth and final component was composed of other erythrocytic related traits, however, the differences between disorders were less marked than component four. Taken together, this demonstrates that groups of biomarkers tend to have similar relationships with different psychiatric traits.

Figure 1. Genetic correlation between blood-based biomarkers and psychiatric GWAS. (a) Heatmap of LDSR correlation Z scores (r_g /standard error) between each psychiatric trait and 49 biochemical GWAS. The psychiatric and biochemical traits are grouped on the x and y axes, respectively, by hierarchical clustering using Pearson's distance. (b) Examples of biochemical traits with evidence of discordant genetic correlations amongst the different psychiatric phenotypes. C-reactive protein (CRP), alanine aminotransferase (ALT), and sex-hormone binding globulin (SHBG) are presented for illustration. The forest plot denotes the LDSR r_g with its standard error representing the confidence bars. Traits highlighted in blue, orange, and red, for CRP, ALT, and SHBG, respectively, were significantly correlated after the application of multiple testing correction. (c) Correlation matrix (Pearson) of LDSR Z between each trait, correlation estimates that survive correction for the number of tests performed are highlighted. (d) Components of biochemical LDSR Z scores derived using finite Gaussian mixture modelling (GMM) – the optimal parametrisation of the variance-covariance matrix was five components with diagonal distribution, variable volume, and equal shape (VEI). The components are plotted relative to their contribution to the first and second principal component of the LDSR Z matrix. (e) Box-and-whisker plots of the LDSR Z for each disorder composed of traits assigned to each of the five components derived from the GMM procedure.

Genetically proxied biochemical measures were associated with a severe cognitive deficit schizophrenia subtype

We tested the association between 25 genetically proxied biomarkers (polygenic scores – PGS) that were correlated with either the schizophrenia or general cognitive ability GWAS and a severe cognitive deficit subtype of schizophrenia (Cognitive Deficit - CD) and a subset of cases with less marked impairment (Cognitively Spared – CS). There were two biochemical PGS that displayed a relatively significant association with CD status – haematocrit percentage and immature reticulocyte fraction (Figure 2, Supplementary Table 13). Each standard deviation increase in the immature reticulocyte fraction PGS was associated with a 35.7% (95% CI: 14.5%, 56.9%, $P = 4.72 \times 10^{-3}$, $q = 0.07$) increase in the odds of severe cognitive deficit. Conversely, genetically proxied haematocrit percentage exerted a protective effect – OR = 0.744 [95% CI: 0.533, 0.954], $P = 5.82 \times 10^{-3}$, $q = 0.07$. We emphasise that these signals only survive multiple testing correction using a lenient false-discovery rate (FDR) threshold of 10%, however, given the small sample size of this cohort, we believe that these findings remain noteworthy. Immature reticulocyte fraction was negatively correlated with cognition, with a trend towards a negative relationship with schizophrenia as well. Interestingly, haematocrit percentage was also negatively correlated with cognition, however, there was a depletion of haematocrit percentage alleles amongst CD vs CS schizophrenia cases, suggesting a more complex relationship may be present. Both the haematocrit and reticulocyte fraction PGS explained around 1% of phenotypic variance in CD on the liability scale, which is similar to a PGS for general cognitive ability (Figure 2). A model constructed with all nominally CD-associated PGS ($P < 0.05$) explained almost 3% in phenotypic variance. There was a non-significant trend of enrichment of schizophrenia PRS ($P = 0.054$) in CD. Whilst these values are modest, it does suggest that the genetic architecture of biochemical traits is correlated with the severity of cognitive impairment in schizophrenia.

Figure 2. The association between biochemical polygenic scores and a severe cognitive deficit subtype of schizophrenia. Left panel is a forest plot of the association (odds ratio with 95% confidence interval) between a standard deviation increase in each polygenic score (PGS) a severe cognitive deficit (CD), relative to schizophrenia cases with less marked impairment (CS). The right panel is the variance explained on the liability scale (Nagelkerke's R^2 , assuming an 0.33% CD population prevalence) for biochemical and a general cognitive ability PGS.

Strong evidence of partial genetic causality between blood-based biomarkers and neuropsychiatric illness

A latent causal variable (LCV) model was constructed between each significantly correlated biochemical-psychiatric trait pair to estimate partial genetic causality ($|\widehat{GCP}| > 0.6$ as strong evidence of partial genetic causality). We can then infer the consequence of the partial genetic causality of one trait on another using the sign of the genetic correlation. There were five instances where we found strong evidence of a potential causal relationship, with all of them suggesting an effect of the biochemical measure on the psychiatric trait, rather than *vice versa* (Table 2, Supplementary table 14). These were as follows: glucose on ADHD ($\widehat{GCP} = 0.64$), CRP on AN ($\widehat{GCP} = 0.92$), urate on cognition ($\widehat{GCP} = 0.88$), CRP on MDD ($\widehat{GCP} = 0.62$), and

CRP on OCD ($\widehat{GCP} = 0.75$). It is important to emphasise that these posterior mean $\widehat{GCP}$ are not magnitudes of causal effect and only imply that there is a causal relationship between the biochemical traits and the psychiatric phenotype. Given the sign of the genetic correlation, we can likely infer that glucose may increase the risk of ADHD and urate could have a deleterious effect on general cognitive function. As outlined in a previous section, CRP has highly divergent correlations, and these data coupled with the LDSR further support a protective effect on AN and OCD, whilst it is likely risk increasing for MDD. SZ did not quite survive multiple-testing correction for a genetic correlation with CRP, however, given previous evidence of a protective effect of CRP on SZ from MR studies (65–67), we also constructed an LCV model between CRP and SZ and found moderate support for this relationship ($\widehat{GCP} = 0.56$, $SE = 0.23$, $P = 5.11 \times 10^{-6}$). Given the genetic correlation is only small it is less likely that previous MR studies were unduly biased by genetic correlation. In addition, we also observed an unusual phenomenon in the HbA1c and PTSD model, whereby the posterior mean $\widehat{GCP}$ was strongly positive ($\widehat{GCP} = 0.76$), implying an effect of HbA1c on PTSD, whilst its Z score was negative ($Z = -6.51$), which signifies the opposite. As described in the supplementary material and supplementary figure 1, we found that these conflicting data was likely attributable to a rare violation of the LCV model assumptions, whereby the mixed fourth moments had opposite signs to each other and the genetic correlation. This could be explained by certain SNPs having highly divergent effects from the rest of the genome-wide signal. Moreover, there were two other trait pairs that trended towards partial genetic causation ($|\widehat{GCP}| > 0.5$) but did not exceed the stringent 0.6 threshold. These were SHBG on cognition ($\widehat{GCP} = 0.55$) and triglycerides on OCD ($\widehat{GCP} = 0.55$). We then sought to replicate the LCV findings by utilising different previously published GWAS for glucose, urate, CRP, and HbA1c. Despite smaller sample sizes, we found relatively consistent GCP estimates which supported the above models using the UKBB GWAS (supplementary table 14). The exception to this was HbA1c, which did not

show strong evidence of partial genetic causality ($\widehat{GCP} = 0.13$) using a smaller sample size GWAS by Wheeler *et al.* (37).

Table 2: Strong evidence of partial genetic causality of a biochemical measure on a psychiatric trait

Biochemical trait	Psychiatric trait ¹	GCP^2	SE	P^3	r_g
Glucose	ADHD	0.64	0.28	0.035	0.134
C-reactive protein	AN	0.92	0.07	5.95×10^{-56}	-0.286
C-reactive protein	MDD	0.62	0.21	1.57×10^{-12}	0.154
C-reactive protein	OCD	0.75	0.16	2.07×10^{-18}	-0.201
Urate	Cognition	0.88	0.09	5.27×10^{-114}	-0.1047

¹Psychiatric GWAS: attention/deficit hyperactivity disorder (ADHD), anorexia nervosa (AN), major depressive disorder (MDD), and obsessive-compulsive disorder (OCD).

²The posterior mean genetic causality proportion (GCP), with its accompanying posterior standard error in the adjacent column.

³ P value which tests whether the GCP estimate is significantly different from zero.

C-reactive protein levels exert a direct protective effect on schizophrenia conditioned on body mass index and interleukin-6 signalling

CRP displayed strong evidence of partial genetic causality on three psychiatric disorders, and thus, we sought to further analyse these relationships by estimating total effects and direct impact of CRP using univariable and multivariable Mendelian randomisation (MR), respectively (Supplementary Tables 16-22). We included SZ in these analyses, due to evidence from previous SZ GWAS that CRP exerts a protective effect on schizophrenia liability, as discussed in the previous section. The CRP GWAS utilised here was drawn from a non-UKBB cohort such that there was no sample overlap with the AN and MDD GWAS (Supplementary

Materials) (34). Using our primary model (IVW estimator with multiplicative random effects), we found evidence that a natural log transformed mg/L increase in CRP was associated with a statistically significant reduction in the odds of SZ (OR = 0.91 [95% CI: 0.85,0.98], $P = 0.01$) and AN (OR = 0.91 [95% CI: 0.83,0.99], $P = 0.03$), which supports the LDSR inferred direction of the LCV relationship. Using a less conservative IVW estimator with fixed effects yielded a more precise estimate in both instances, as expected (CRP $\rightarrow$ SZ $P = 6.29 \times 10^{-5}$, CRP $\rightarrow$ AN $P = 9.79 \times 10^{-3}$). There was a trend towards an odds increasing effect of CRP on MDD ($P = 0.19$), whilst there was no indication of a reliable effect in the OCD model ($P = 0.82$). It should be noted that as MR exploits independent (relative linkage equilibrium) genome-wide significant variants as IVs, as opposed to using genome-wide marginal effects in the LCV approach, and thus, a non-significant estimate from MR does not necessarily preclude the existence of a causal relationship, although LCV models with corresponding MR support would perhaps be viewed as stronger evidence. We discuss the sensitivity analyses for each of these univariable models in detail in the supplementary text. Briefly, the five MR tests deployed with different assumptions regarding IV validity (plurality valid, majority valid, and InSIDE assumption) had very similar point estimates (OR range: 0.88-0.91) and were all statistically significant for CRP $\rightarrow$ SZ with the exception of the simple median ($P = 0.07$). The CRP $\rightarrow$ AN estimate across the different models were also directionally consistent, however, they were not statistically significant (except for some contamination mixture models with different prespecified standard deviation of invalid IVs), and thus, the total effect of CRP on AN by MR has comparatively weaker evidence compared to SZ. The effect of using robust, penalised, or robust penalised weights in the median, IVW, and Egger models was not marked for each of the CRP psychiatric models (Figure 3a, Supplementary Table 21). In the CRP $\rightarrow$ SZ and CRP $\rightarrow$ AN model, the Egger intercept was not significantly different than zero (although there was a trend in the AN model, $P = 0.07$), and thus, there was no strong statistical evidence of

confounder pleiotropy using this metric. However, there was evidence of heterogeneity between the IV ratio estimates for AN and SZ: the MR PRESSO global test of pleiotropy was significant and Cochran's Q statistic significant for AN and SZ, although both causal estimates remained statistically significant in the MR-PRESSO outlier corrected estimates (Supplementary table 19). Given the biological complexity of these phenotypes, heterogeneity does not necessarily imply confounding pleiotropy. Moreover, using a leave-one-out-analysis, we found evidence of one outlier IV in the SZ model, whilst there were three outlier SNPs in the AN model (Supplementary Table 20), although the effects of removing these IVs were relatively small. The outlier IV in the SZ model was also proximal to the CRP gene itself (IVW $P = 0.07$ when removed), meaning it is likely to influence SZ through CRP, rather than being indicative of confounding pleiotropy. Given the CRP estimates on AN and SZ remained significant removing outliers by MR PRESSO or through iterative single IV exclusion, it is less likely that horizontal pleiotropy fully explains these signals. There was also no evidence using a reverse MR model, with the psychiatric disorder as the exposure, that there were bidirectional effects, although these models are inherently underpowered and are best treated as a test of the null hypothesis only (Supplementary Table 22).

Figure 3. Total and direct estimated effect of C-reactive protein on psychiatric illness. (a) Total effect of CRP on each psychiatric outcome considered. Each point represents the IV-exposure vs IV-outcome effect. Trend lines are indicative of the slope of each MR method utilised. The outcomes plotted from left to right are AN,

MDD, OCD, and SZ. **(b)** Direct effect of CRP in multivariable MR models – the three models contained the following phenotypes as additional exposures: i) circulating interleukin 6 (IL6) and its receptor (IL6R), body mass index (BMI), and IL6R and BMI.

We sought to investigate the direct effect of circulating CRP on the above psychiatric phenotypes using multivariable Mendelian randomisation (MVMR) conditioning on IL-6 signalling and BMI. We found that CRP exhibited a robust direct protective effect on schizophrenia conditioned on IL-6 signalling and BMI (Supplementary table 23, Figure 3b). For instance, using a multivariable IVW estimator, the direct effect of CRP on schizophrenia conditioned on BMI and IL-6R was analogous to the univariable IVW total estimate (OR = 0.88 [95% CI: 0.79, 0.98], $P = 0.01$). There was no evidence of an effect of BMI or IL-6 signalling on schizophrenia conditioned on CRP. Similarly, the effect size of the CRP → AN MVMR models remained similar to that estimated in the univariable constructs (Figure 3b), however, these estimates were only statistically significant in a subset of the models (Supplementary table 23). As a result, the MR evidence for the direct protective effect of CRP on AN is comparatively weaker than the CRP → SZ model, as was seen in the univariable estimates. We also observed some evidence to suggest that IL-6 abundance exerts a protective effect on AN conditional on CRP and IL-6R, which was nominally non-zero in every MVMR model except for the median estimator (OR = 0.96 [95% CI: 0.93, 0.99] per unit increase in IL-6, $P = 0.037$ – multivariable IVW-multiplicative random effects). Interestingly, whilst evidence for a direct effect of CRP on MDD conditioned on each variable set was weak (Figure 3b, Supplementary Table 23), there was consistent evidence that elevated IL-6R was associated with increased odds of MDD conditioned on CRP and IL-6 or CRP and BMI. For example, each unit increase in blood IL-6R protein expression was estimated to increase the odds of

MDD by 2.7% [95% CI: 0.6%, 4.9%] conditioned on CRP and BMI. Finally, there was no evidence in the MVMR models to support the protective effect of CRP on OCD, however, there was evidence that increased BMI decreases the odds of OCD conditioned on CRP and IL-6 signalling (Supplementary Table 23). This BMI → OCD effect estimate was quite large, albeit with wide confidence intervals – OCD OR per SD increase in BMI conditioned on CRP and IL-6R = 0.52 [95% CI: 0.14, 0.91], $P = 1 \times 10^{-3}$.

C-reactive protein displays overlapping association signals with schizophrenia and may have downstream impacts on the brain

There were five schizophrenia GWAS lead SNPs ($P < 5 \times 10^{-8}$), that also obtained genome-wide significance in the UKBB CRP GWAS. Colocalisation analyses using the European only subset of the SZ GWAS demonstrated strong evidence for the association of three of these loci with both SZ and CRP, however, there was likely a different underlying causal variant (Supplementary Table 24, Supplementary Text, Supplementary Figure 2). Local genetic correlation estimates with ρ -HESS demonstrated that 13 of the LD block partitions of the human genome displayed non-zero local covariance between CRP and SZ (Supplementary Table 25, Supplementary Figure 3) after Bonferroni correction. Interestingly, five of these LD blocks with strong evidence of local genetic covariance were positive, and thus, for these regions of the genome SZ and CRP were positively correlated in contrast to the genome-wide estimate of nominal negative correlation. We investigated the pathways overrepresented for genes physically mapped to the LD blocks of the top five most significant $r_{g,local}$ estimates (Supplementary Table 27). Interestingly, we found that *retinol metabolism* was the most significantly overrepresented gene-set after multiple testing correction, which is notable given evidence that retinoids play a role in the pathogenesis of SZ (30,68).

We also sought to investigate the downstream consequences of raised CRP. We estimated the effect of raised CRP on the expression of 3284 proteins in blood using univariable MR. Using a liberal FDR cut-off of 10%, we found that 95 proteins were putatively causally influenced by elevated CRP, with 45 of these proteins surviving a stricter FDR threshold of 5% (Supplementary Table 28). We emphasise that these analyses are exploratory in nature and we treat the effect sizes of the CRP effect on each protein largely as a test of the null hypothesis that the two are not associated. The vast majority of proteins prioritised using the IVW estimator were directionally consistent in the median, mode, and Egger sensitivity analyses (Supplementary Table 29). However, 20 of the $FDR < 0.1$ proteins were suggested to act in the opposite direction and causally influence CRP as the estimated variance explained by the IVs was significantly larger in the outcome than the CRP exposure, although this does not rule out bidirectional effects (Supplementary Table 30). We found that these proteins putatively influenced by CRP were enriched in several pathways including glycaemic signalling and lymphocyte biology (Supplementary Tables 31-33). Notably, there were neuronal pathways overrepresented with CRP-associated proteins – including, *axon guidance*, *dopaminergic synapse*, *neurogenesis*, *glial cell differentiation*, and *cholinergic synapse*. For example, dopaminergic and cholinergic synapse overrepresentations were driven by the three genes (*AKT1*, *AKT2*, and *AKT3*) that encode the RAC-alpha/beta/gamma serine/threonine-protein kinase complex, which was measured as a single entity in the protein study. Collectively, these data provide preliminary evidence that CRP may influence the expression of neuronally relevant proteins, and subsequent studies should seek to investigate these associations and their significance for different psychiatric phenotypes.

DISCUSSION

In this study, we investigated genetic correlation and causality between a diverse panel of biochemical traits and psychiatric disorders, as well as general cognitive ability and a cognitive deficit subtype of schizophrenia. Our data demonstrated that there is clear evidence of genetic overlap between blood-based measures and psychiatric phenotypes, as quantified by LDSR, which is likely indicative of shared variants and pathways that predispose to these traits. Interestingly, the distribution of biochemical-psychiatric correlations demonstrated traits often exhibited highly divergent correlations (different signs), as well as clusters of biochemical measures that tended to have similar psychiatric correlation profiles. For instance, we found that five reticulocyte traits clustered together (Figure 1e), and they tended to have opposing psychiatric correlations (strong positive correlation with ADHD, MDD, and PTSD, whilst negative correlations with AN, OCD, SZ, and cognition). The genetic architecture of reticulocyte related traits remains relatively uncharacterised, however, it is a highly polygenic trait that also has demonstrated genome-wide significant associations with rare non-synonymous variation in genes such as *SPTA1*, *E2F4*, and *IFRD2* (69,70). Future study should further investigate how the genetic factors which contribute to reticulocyte biology may also influence psychiatric traits. Given the LCV posterior mean GCP estimates between the genetically correlated reticulocyte traits and each psychiatric phenotype were low, it suggests the existence of horizontal rather than vertical pleiotropy.

A key advantage of our study is that we extended the findings from the LDSR models to estimate which biochemical-psychiatric trait pairs may represent causal relationships. We caution that all of these findings require validation in well-powered, replicated, randomised controlled trials to confirm that the causal effects do indeed exist. The putative effect of urate and glucose on cognition and ADHD, respectively, may have direct implications for drug

repurposing given that compounds which modulate these traits are readily available. Both of these relationships are also supported by previous observational data (71–73). Indeed, data from an incident cohort study which treated individuals with urate lowering compounds allopurinol and febuxostat demonstrated evidence of a risk decreasing effect on dementia, supporting the deleterious effect of urate on general cognitive function (74).

The inferred effect of CRP on different psychiatric disorders was particularly interesting given that there was evidence of an odds-decreasing effect on AN, OCD, and SZ, whilst the opposite is true for MDD. CRP is traditionally conceptualised as a biomarker of chronic inflammation; however, its biology is likely somewhat more complex given it is also implicated to play a direct role in pathogen response (75,76). Moreover, as reviewed by Del Giudice and Gangestad, CRP in its hepatically secreted pentameric isoform demonstrates some anti-inflammatory effects and may be a marker of other non-inflammatory states (77). In our study, we also provide evidence that CRP levels may also exert an effect on proteins with neurological significance, including proteins enriched in pathways relevant to psychiatric illness such as axon guidance. Moreover, we show for the first time the previously documented protective effect of CRP on SZ through MR is not likely attributable to the effect of BMI and IL-6 signalling, which are closely related variables to CRP. Data from our study and previous examinations of the relationship between CRP and SZ through MR seemingly contradict previous observational evidence that CRP is elevated in SZ (64,78). If we assume that there is a causal effect, there are a number of explanations that could account for this, although all require further investigation. Firstly, previous observational studies that directly measured CRP in case/control cohorts could be confounded due to a variety of variables including lifestyle and general health, or even be caused by factors related to psychosis and/or schizophrenia itself. Whilst there was no evidence in this study using the LCV model or reverse

MR that SZ causally influences CRP levels, local significant estimates of positive and negative genetic correlation between SZ and CRP observed for several LD blocks suggests a complex interrelationship between these phenotypes may exist that warrants further study. Secondly, CRP itself may influence factors that impact the brain, directly or indirectly, and these factors may play a role in the pathogenesis of psychiatric illness. For instance, CRP was postulated in this study to causally upregulate the expression of the RAC-alpha/beta/gamma serine/threonine-protein kinase complex (*AKT1*, *AKT2*, and *AKT3*), with impaired signalling by these serine/threonine kinases implicated to dysregulate dopaminergic neurotransmission and downregulation of genetically predicted neuronal *AKT3* expression associated with SZ via a transcriptome-wide association study (23,79). Finally, given that infection has been associated with liability of SZ, the role of CRP in pathogen defence may contribute to its putative protective properties. Further study on the neurobiological consequences of CRP signalling and its role in SZ is warranted, particularly to reconcile the discrepancies between observational studies and MR.

The putative causal effect of CRP on AN was demonstrated in this study is, to our knowledge, a novel finding; however, it does support data from a recent longitudinal study which demonstrated that elevated CRP was associated with a protective effect on eating disorders (62), along with decreased measured CRP observed specifically in AN (80). It should be noted that whilst the LCV data supported strong partial genetic causality of CRP on AN, the MR evidence was less statistically significant, with some evidence in the MVMR that IL-6 signalling may exert a protective effect on AN conditioned-on CRP. Inhibition of the IL-6 pathway has been associated with weight gain which may be protective for AN (81). Furthermore, CRP likely is intertwined with other metabolic factors, including insulin signalling, which our group has previously shown through MR is also putatively protective for

AN (48). In MDD, we found some evidence to support that upregulation of the IL-6 receptor may be risk increasing conditioned on CRP and BMI, supporting preliminary data that blockade of IL-6R by agents like tocilizumab may improve depressive symptoms (82). However, data related to the anti-depressant qualities of tocilizumab are conflicting (83), and randomised control trials are warranted to further investigate repurposing opportunities for IL-6 inhibition in MDD. Finally, BMI demonstrated a quite robust protective effect on OCD conditioned on CRP and IL-6 signalling, in accordance with observational data that OCD is associated with reduced odds of obesity (84).

There are a number of important limitations that are central to the interpretation of the data in this study. Genetic correlations from LDSR likely reflect a shared underlying genetic architecture, however, this could be mediated by the relationship of the same genetic variants to another variable or variables (7). Despite this limitation, the existence of horizontal pleiotropy between traits is still informative as identifying genes which effect both psychiatric and biochemical traits, and further insight into the mechanisms, would likely refine our understanding of both traits. Moreover, the genetically informed causal inference approaches we implement in this study are subject to limitations regarding the data they are performed with and any biases therein, including potential effects of population stratification (85), selection bias (86), and the assumption of acyclicity. The LCV model is also fixed to be bivariate in nature, and thus, the effects of multiple mediators cannot be taken into account. We address this by constructing multivariable MR models such that direct effects are estimated conditioned on likely confounders, however, our selection of confounders is not exhaustive, and other unidentified factors may influence our findings. The UKBB sample is also composed of older individuals over the age of 40, and thus, more developmentally sensitive effects on the biochemical traits in question could not be assessed. Genetic variants are also sometimes

claimed to represent a lifetime effect on a particular variable, although caution is required in this inference given that factors like age may modulate the effect of a variant (87,88). We assert that although GWAS informed causal inference has a number of caveats and limitations it enables an important opportunity to prioritise biochemical traits which are putatively clinically relevant in psychiatry and inform future study into these traits.

ACKNOWLEDGEMENTS

The authors wish to acknowledge Dr Luke O'Connor for his insightful correspondence in relation to some of the LCV data in this study. This work was supported by a National Health and Medical Research Council (NHMRC) grant (1188493). M.J.C. is supported by an NHMRC Senior Research Fellowship (1121474), URL: <https://www.nhmrc.gov.au/>. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the paper.

DISCLOSURES

The authors declare no competing financial interests

AUTHOR CONTRIBUTIONS

W.R.R. conceived and designed the study with input from M.J.C. W.R.R. performed the primary analyses. M.P.G and D.J.K performed the rho-HESS analyses. J.R.A performed the SNP imputation in the ASRB cohort. M.J.C, V.J.C, and M.J.G contributed the ASRB cohort, as well as to critical interpretation of the results and drafting of the manuscript.

REFERNECES

1. Polderman TJC, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, Posthuma D (2015): Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet* 47: 702–709.
2. Brainstorm Consortium, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, *et al.* (2018): Analysis of shared heritability in common disorders of the brain. *Science* 360. <https://doi.org/10.1126/science.aap8757>
3. Cross-Disorder Group of the Psychiatric Genomics Consortium. Electronic address: plee0@mg.harvard.edu, Cross-Disorder Group of the Psychiatric Genomics Consortium (2019): Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* 179: 1469-1482.e11.
4. Reay WR, Cairns MJ (2020): Pairwise common variant meta-analyses of schizophrenia with other psychiatric disorders reveals shared and distinct gene and gene-set associations. *Transl Psychiatry* 10: 134.
5. Gamazon ER, Zwinderman AH, Cox NJ, Denys D, Derks EM (2019): Multi-tissue transcriptome analyses identify genetic mechanisms underlying neuropsychiatric traits. *Nat Genet* 51: 933–940.
6. Byrne EM, Zhu Z, Qi T, Skene NG, Bryois J, Pardinas AF, *et al.* (2020): Conditional GWAS analysis to identify disorder-specific SNPs for psychiatric disorders. *Mol Psychiatry*. <https://doi.org/10.1038/s41380-020-0705-9>
7. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, *et al.* (2015): An atlas of genetic correlations across human diseases and traits. *Nat Genet* 47: 1236–1241.
8. Burgess S, Butterworth A, Thompson SG (2013): Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 37: 658–665.

9. O'Connor LJ, Price AL (2018): Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat Genet* 50: 1728–1734.
10. Chan MK, Guest PC, Levin Y, Umrana Y, Schwarz E, Bahn S, Rahmoune H (2011): Converging evidence of blood-based biomarkers for schizophrenia. *International Review of Neurobiology*, vol. 101. Elsevier, pp 95–144.
11. Lai C-Y, Scarr E, Udawela M, Everall I, Chen WJ, Dean B (2016): Biomarkers in schizophrenia: A focus on blood based diagnostics and theranostics. *World J Psychiatry* 6: 102–117.
12. Carvalho AF, Solmi M, Sanches M, Machado MO, Stubbs B, Ajnakina O, *et al.* (2020): Evidence-based umbrella review of 162 peripheral biomarkers for major mental disorders. *Transl Psychiatry* 10: 152.
13. Meduna LJ (1942): BIOCHEMICAL DISTURBANCES IN MENTAL DISORDERS: I. ANTI-INSULIN EFFECT OF BLOOD IN CASES OF SCHIZOPHRENIA. *Arch NeurPsych* 47: 38.
14. Kim JH, Kim JY, Lee J, Jeong GH, Lee E, Lee S, *et al.* (2020): Environmental risk factors, protective factors, and peripheral biomarkers for ADHD: an umbrella review. *The Lancet Psychiatry* 7: 955–970.
15. Masi A, Quintana DS, Glozier N, Lloyd AR, Hickie IB, Guastella AJ (2015): Cytokine aberrations in autism spectrum disorder: a systematic review and meta-analysis. *Mol Psychiatry* 20: 440–446.
16. Brown NC, Andreazza AC, Young LT (2014): An updated meta-analysis of oxidative stress markers in bipolar disorder. *Psychiatry Res* 218: 61–68.
17. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, *et al.* (2019): Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* 51: 63–75.

18. Watson HJ, Yilmaz Z, Thornton LM, Hübel C, Coleman JRI, Gaspar HA, *et al.* (2019):
Genome-wide association study identifies eight risk loci and implicates metabo-
psychiatric origins for anorexia nervosa. *Nat Genet* 51: 1207–1214.
19. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, *et al.* (2019):
Identification of common genetic risk variants for autism spectrum disorder. *Nat*
Genet 51: 431–444.
20. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V, *et al.* (2019):
Genome-wide association study identifies 30 loci associated with bipolar disorder.
Nat Genet 51: 793–803.
21. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, *et al.*
(2018): Genome-wide association analyses identify 44 risk variants and refine the
genetic architecture of major depression. *Nat Genet* 50: 668–681.
22. Nievergelt CM, Maihofer AX, Klengel T, Atkinson EG, Chen C-Y, Choi KW, *et al.*
(2019): International meta-analysis of PTSD genome-wide association studies
identifies sex- and ancestry-specific genetic risk loci. *Nat Commun* 10: 4558.
23. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Ripke S,
Walters JT, O'Donovan MC (2020): *Mapping Genomic Loci Prioritises Genes and*
Implicates Synaptic Biology in Schizophrenia. Genetic and Genomic Medicine.
<https://doi.org/10.1101/2020.09.12.20192922>
24. Yu D, Sul JH, Tsetsos F, Nawaz MS, Huang AY, Zelaya I, *et al.* (2019): Interrogating the
Genetic Determinants of Tourette's Syndrome and Other Tic Disorders Through
Genome-Wide Association Studies. *Am J Psychiatry* 176: 217–227.
25. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, de Leeuw CA, *et al.* (2018):
Genome-wide association meta-analysis in 269,867 individuals identifies new genetic
and functional links to intelligence. *Nat Genet* 50: 912–919.

26. Scrucca L, Fop M, Murphy TB, Raftery AE (2016): mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *R J* 8: 289–317.
27. Hallmayer JF, Kalaydjieva L, Badcock J, Dragovic M, Howell S, Michie PT, *et al.* (2005): Genetic evidence for a distinct subtype of schizophrenia characterized by pervasive cognitive deficit. *Am J Hum Genet* 77: 468–476.
28. Green MJ, Cairns MJ, Wu J, Dragovic M, Jablensky A, Tooney PA, *et al.* (2013): Genome-wide supported variant MIR137 and severe negative symptoms predict membership of an impaired cognitive subtype of schizophrenia. *Mol Psychiatry* 18: 774–780.
29. Loughland C, Draganic D, McCabe K, Richards J, Nasir A, Allen J, *et al.* (2010): Australian Schizophrenia Research Bank: a database of comprehensive clinical, endophenotypic and genetic data for aetiological studies of schizophrenia. *Aust N Z J Psychiatry* 44: 1029–1035.
30. Reay WR, Atkins JR, Quidé Y, Carr VJ, Green MJ, Cairns MJ (2020): Polygenic disruption of retinoid signalling in schizophrenia and a severe cognitive deficit subtype. *Mol Psychiatry* 25: 719–731.
31. Reay WR, Atkins JR, Carr VJ, Green MJ, Cairns MJ (2020): Pharmacological enrichment of polygenic risk for precision medicine in complex disorders. *Sci Rep* 10: 879.
32. Lee SH, Goddard ME, Wray NR, Visscher PM (2012): A better coefficient of determination for genetic profile analysis. *Genet Epidemiol* 36: 214–224.
33. Reay WR, Shair SE, Geaghan MP, Riveros C, Holiday EG, McEvoy MA, *et al.* (2020): *Genetically Informed Precision Drug Repurposing for Lung Function and Implications for Respiratory Infection*. Respiratory Medicine. <https://doi.org/10.1101/2020.06.25.20139816>

34. Ligthart S, Vaez A, Vösa U, Stathopoulou MG, de Vries PS, Prins BP, *et al.* (2018):
Genome Analyses of >200,000 Individuals Identify 58 Loci for Chronic Inflammation
and Highlight Pathways that Link Inflammation and Complex Disorders. *Am J Hum*
Genet 103: 691–706.
35. Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC), Lagou V,
Mägi R, Hottenga J-J, Grallert H, Perry JRB, *et al.* (2021): Sex-dimorphic genetic
effects and novel loci for fasting glucose and insulin variability. *Nat Commun* 12: 24.
36. Tin A, Marten J, Halperin Kuhns VL, Li Y, Wuttke M, Kirsten H, *et al.* (2019): Target
genes, variants, tissues and transcriptional pathways influencing human serum urate
levels. *Nat Genet* 51: 1459–1474.
37. Wheeler E, Leong A, Liu C-T, Hivert M-F, Strawbridge RJ, Podmore C, *et al.* (2017):
Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk
and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-
analysis. *PLoS Med* 14: e1002383.
38. Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG (2017): Sensitivity Analyses for
Robust Causal Inference from Mendelian Randomization Analyses with Multiple
Genetic Variants. *Epidemiology* 28: 30–42.
39. Bowden J, Davey Smith G, Haycock PC, Burgess S (2016): Consistent Estimation in
Mendelian Randomization with Some Invalid Instruments Using a Weighted Median
Estimator. *Genet Epidemiol* 40: 304–314.
40. Burgess S, Foley CN, Allara E, Staley JR, Howson JMM (2020): A robust and efficient
method for Mendelian randomization with hundreds of genetic variants. *Nat Commun*
11: 376.

41. Hartwig FP, Davey Smith G, Bowden J (2017): Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol* 46: 1985–1998.
42. Bowden J, Davey Smith G, Burgess S (2015): Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 44: 512–525.
43. Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan NA, Thompson JR (2016): Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I² statistic. *Int J Epidemiol* 45: 1961–1974.
44. Rees JMB, Wood AM, Dudbridge F, Burgess S (2019): Robust methods in Mendelian randomization via penalization of heterogeneous causal estimates. *PLoS One* 14: e0222362.
45. Bowden J, Hemani G, Davey Smith G (2018): Invited Commentary: Detecting Individual and Global Horizontal Pleiotropy in Mendelian Randomization-A Job for the Humble Heterogeneity Statistic? *Am J Epidemiol* 187: 2681–2685.
46. Verbanck M, Chen C-Y, Neale B, Do R (2018): Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet* 50: 693–698.
47. Burgess S, Labrecque JA (2018): Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. *Eur J Epidemiol* 33: 947–952.
48. Adams DM, Reay WR, Geaghan MP, Cairns MJ (2020): Investigation of glycaemic traits in psychiatric disorders using Mendelian randomisation revealed a causal relationship

- with anorexia nervosa. *Neuropsychopharmacology*. <https://doi.org/10.1038/s41386-020-00847-w>
49. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, *et al.* (2015): Genetic studies of body mass index yield new insights for obesity biology. *Nature* 518: 197–206.
50. Folkersen L, Fauman E, Sabater-Lleal M, Strawbridge RJ, Frånberg M, Sennblad B, *et al.* (2017): Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. *PLoS Genet* 13: e1006706.
51. Sanderson E, Davey Smith G, Windmeijer F, Bowden J (2019): An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *International Journal of Epidemiology* 48: 713–727.
52. Sanderson E, Spiller W, Bowden J (2020): *Testing and Correcting for Weak and Pleiotropic Instruments in Two-Sample Multivariable Mendelian Randomisation*. Genetics. <https://doi.org/10.1101/2020.04.02.021980>
53. Grant AJ, Burgess S (2020): Pleiotropy robust methods for multivariable Mendelian randomization. *arXiv:2008.11997 [stat]*. Retrieved March 1, 2021, from <http://arxiv.org/abs/2008.11997>
54. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, *et al.* (2018): The MR-Base platform supports systematic causal inference across the human phenome. *eLife* 7: e34408.
55. Yavorska OO, Burgess S (2017): MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol* 46: 1734–1739.

56. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V (2014): Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* 10: e1004383.
57. Shi H, Mancuso N, Spendlove S, Pasaniuc B (2017): Local Genetic Correlation Gives Insights into the Shared Genetic Architecture of Complex Traits. *Am J Hum Genet* 101: 737–751.
58. Berisa T, Pickrell JK (2016): Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* 32: 283–285.
59. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, Vilo J (2019): g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research* 47: W191–W198.
60. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, *et al.* (2018): Genomic atlas of the human plasma proteome. *Nature* 558: 73–79.
61. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, *et al.* (2019): STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 47: D607–D613.
62. Solmi F, Bulik CM, De Stavola BL, Dalman C, Khandaker GM, Lewis G (2020): Longitudinal associations between circulating interleukin-6 and C-reactive protein in childhood, and eating disorders and disordered eating in adolescence. *Brain, Behavior, and Immunity* 89: 491–500.
63. Chang JP-C, Mondelli V, Satyanarayanan SK, Chiang Y-J, Chen H-T, Su K-P, Pariante CM (2020): Cortisol, inflammatory biomarkers and neurotrophins in children and adolescents with attention deficit hyperactivity disorder (ADHD) in Taiwan. *Brain, Behavior, and Immunity* 88: 105–113.

64. Fond G, Lançon C, Auquier P, Boyer L (2018): C-Reactive Protein as a Peripheral Biomarker in Schizophrenia. An Updated Systematic Review. *Front Psychiatry* 9: 392.
65. Hartwig FP, Borges MC, Horta BL, Bowden J, Davey Smith G (2017): Inflammatory Biomarkers and Risk of Schizophrenia: A 2-Sample Mendelian Randomization Study. *JAMA Psychiatry* 74: 1226.
66. Prins BP, Abbasi A, Wong A, Vaez A, Nolte I, Franceschini N, *et al.* (2016): Investigating the Causal Relationship of C-Reactive Protein with 32 Complex Somatic and Psychiatric Outcomes: A Large-Scale Cross-Consortium Mendelian Randomization Study. *PLoS Med* 13: e1001976.
67. Ligthart S (2019): Commentary: CRP and schizophrenia: cause, consequence or confounding? *Int J Epidemiol* 48: 1514–1515.
68. Reay WR, Cairns MJ (2020): The role of the retinoids in schizophrenia: genomic and clinical perspectives. *Mol Psychiatry* 25: 706–718.
69. Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, *et al.* (2016): The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* 167: 1415-1429.e19.
70. Vuckovic D, Bao EL, Akbari P, Lareau CA, Mousas A, Jiang T, *et al.* (2020): The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell* 182: 1214-1231.e11.
71. Zametkin AJ, Nordahl TE, Gross M, King AC, Semple WE, Rumsey J, *et al.* (1990): Cerebral Glucose Metabolism in Adults with Hyperactivity of Childhood Onset. *N Engl J Med* 323: 1361–1366.
72. Lindblad F, Eickhoff M, Forslund AH, Isaksson J, Gustafsson J (2015): Fasting blood glucose and HbA1c in children with ADHD. *Psychiatry Res* 226: 515–516.

73. Beydoun MA, Canas J-A, Dore GA, Beydoun HA, Rostant OS, Fanelli-Kuczmarski MT, *et al.* (2016): Serum Uric Acid and Its Association with Longitudinal Cognitive Change Among Urban Adults. *J Alzheimers Dis* 52: 1415–1430.
74. Singh JA, Cleveland JD (2018): Comparative effectiveness of allopurinol versus febuxostat for preventing incident dementia in older adults: a propensity-matched analysis. *Arthritis Res Ther* 20: 167.
75. Sproston NR, Ashworth JJ (2018): Role of C-Reactive Protein at Sites of Inflammation and Infection. *Front Immunol* 9: 754.
76. Slaats J, ten Oever J, van de Veerdonk FL, Netea MG (2016): IL-1 β /IL-6/CRP and IL-18/ferritin: Distinct Inflammatory Programs in Infections ((J. B. Bliska, editor)). *PLoS Pathog* 12: e1005973.
77. Del Giudice M, Gangestad SW (2018): Rethinking IL-6 and CRP: Why they are more than inflammatory biomarkers, and why it matters. *Brain, Behavior, and Immunity* 70: 61–75.
78. Jacomb I, Stanton C, Vasudevan R, Powell H, O'Donnell M, Lenroot R, *et al.* (2018): C-Reactive Protein: Higher During Acute Psychotic Episodes and Related to Cortical Thickness in Schizophrenia and Healthy Controls. *Front Immunol* 9: 2230.
79. Li Y-C, Yang S-S, Gao W-J (2016): Disruption of Akt signaling decreases dopamine sensitivity in modulation of inhibitory synaptic transmission in rat prefrontal cortex. *Neuropharmacology* 108: 403–414.
80. Solmi M, Veronese N, Favaro A, Santonastaso P, Manzato E, Sergi G, Correll CU (2015): Inflammatory cytokines and anorexia nervosa: A meta-analysis of cross-sectional and longitudinal studies. *Psychoneuroendocrinology* 51: 237–252.

81. Patsalos O, Dalton B, Himmerich H (2020): Effects of IL-6 Signaling Pathway Inhibition on Weight and BMI: A Systematic Review and Meta-Analysis. *Int J Mol Sci* 21. <https://doi.org/10.3390/ijms21176290>
82. Tiosano S, Yavne Y, Watad A, Langevitz P, Lidar M, Feld J, *et al.* (2020): The impact of tocilizumab on anxiety and depression in patients with rheumatoid arthritis. *Eur J Clin Invest* 50. <https://doi.org/10.1111/eci.13268>
83. Knight JM, Costanzo ES, Singh S, Yin Z, Szabo A, Pawar DS, *et al.* (2021): The IL-6 antagonist tocilizumab is associated with worse depression and related symptoms in the medically ill. *Transl Psychiatry* 11: 58.
84. Abramovitch A, Anholt GE, Cooperman A, van Balkom AJLM, Giltay EJ, Penninx BW, van Oppen P (2019): Body mass index in obsessive-compulsive disorder. *J Affect Disord* 245: 145–151.
85. Koellinger PD, de Vlaming R (2019): Mendelian randomization: the challenge of unobserved environmental confounds. *International Journal of Epidemiology* 48: 665–671.
86. Gkatzionis A, Burgess S (2019): Contextualizing selection bias in Mendelian randomization: how bad is it likely to be? *International Journal of Epidemiology* 48: 691–701.
87. Davies NM, Holmes MV, Davey Smith G (2018): Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* k601.
88. Labrecque JA, Swanson SA (2019): Interpretation and Potential Biases of Mendelian Randomization Estimates With Time-Varying Exposures. *American Journal of Epidemiology* 188: 231–238.

Supplementary Figure 1

Supplementary Figure 1. Analysis of an LCV model violation in the HbA1c to PTSD model. As outlined in the supplementary text, we demonstrated that the two mixed fourth moments (cokurtosis) estimates for HbA1c to PTSD were opposite in sign, with an outlier jackknifed estimate of $E[z_1, z_2^3]$ which suggests discordant SNP effects relative to the genome-wide signal. Conversely, the CRP to AN and CRP to MDD had consistent signs of $E[z_1^3, z_2]$ and $E[z_1, z_2^3]$ relative to each other and the genetic correlation. In both CRP models, $E[z_1, z_2^3]$, which implies $GCP > 0$.

Supplementary Figure 2

Supplementary Figure 2. The effect of different prior probabilities for the hypothesis that the SZ and CRP share a same causal variant. Each of these plots represent the colocalisation results for the five regions outlined in supplementary table 24, in order from region 1 to 5, left to right, top to bottom. We plot the effects of varying the default prior probability for the shared causal variant above or below its default of 1×10^{-5} .

Supplementary Figure 3

Supplementary Figure 3. Local genetic covariance and correlation between schizophrenia and CRP as estimated by ρ -HESS. The first two panels are estimates of local genetic correlation and covariance for each approximately independent LD block throughout the genome. Coloured bars represent significant estimates. The remaining two panels are local estimates SNP heritability in those same regions for CRP and SZ, respectively.