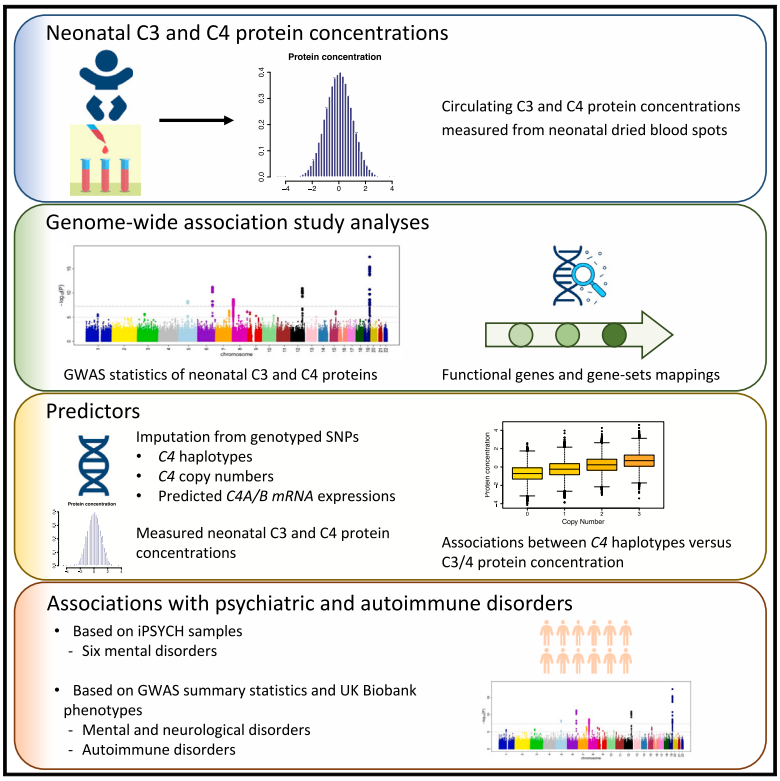
Cell Genomics

Article

The correlates of neonatal complement component 3 and 4 protein concentrations with a focus on psychiatric and autoimmune disorders

Graphical abstract



Highlights

- The genetic correlates of two key complement components (C3, C4) are described
- Neonatal C4 concentration was not associated with six mental disorders
- Higher concentration of C3 was associated with reduced schizophrenia in women
- Mendelian randomization linked C4 concentration and five autoimmune disorders

Authors

Nis Borbye-Lorenzen, Zhihong Zhu, Esben Agerbo, ..., Naomi R. Wray, Bjarni J. Vilhjálmsson, John J. McGrath

Correspondence

z.zhu.ncrr@au.dk (Z.Z.), j.mcgrath@uq.edu.au (J.J.M.)

In brief

The innate immune system has been linked to schizophrenia and autoimmune disorders. We examined neonatal circulating C3 and C4 protein concentrations in 68,768 neonates and the risk of six mental disorders. We found no associations between C4 concentration and mental disorders, but C3 concentration was associated with a reduced risk of schizophrenia in females. We described the genetic correlates of neonatal circulating C3 and C4. Mendelian randomization linked C4 and an altered risk of five types of autoimmune disorder.

Borbye-Lorenzen et al., 2023, Cell Genomics 3, 100457
December 13, 2023
https://doi.org/10.1016/j.xgen.2023.100457

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Article

The correlates of neonatal complement component 3 and 4 protein concentrations with a focus on psychiatric and autoimmune disorders

Nis Borbye-Lorenzen,[1,27] Zhihong Zhu,[2,27,*] Esben Agerbo,[2,3,4] Clara Albiñana,[2,3] Michael E. Benros,[5,6] Beilei Bian,[7] Anders D. Børglum,[3,8,9] Cynthia M. Bulik,[10,11,12] Jean-Christophe Philippe Goldtsche Debost,[2,13] Jakob Grove,[3,9,14,15] David M. Hougaard,[3,16] Allan F. McRae,[7] Ole Mors,[3,17] Preben Bo Mortensen,[2,3,4] Katherine L. Musliner,[18] Merete Nordentoft,[3,19,20] Liselotte V. Petersen,[2] Florian Privé,[2] Julia Sidorenko,[7] Kristin Skogstrand,[1] Thomas Werge,[3,20,21,22] Naomi R. Wray,[7,23,24,25] Bjarni J. Vilhjálmsson,[2,3,15] and John J. McGrath[2,23,26,28,*]

[1]Center for Neonatal Screening, Department of Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark
[2]National Center for Register-Based Research, Aarhus University, 8210 Aarhus V, Denmark
[3]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, 8210 Aarhus V, Denmark
[4]Center for Integrated Register-based Research, Aarhus University, CIRRAU, 8210 Aarhus V, Denmark
[5]Copenhagen Research Center for Mental Health, Mental Health Center Copenhagen, Copenhagen University Hospital, Hellerup, Denmark
[6]Department of Immunology and Microbiology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[7]Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, Australia
[8]Department of Biomedicine and the iSEQ Center, Aarhus University, Aarhus, Denmark
[9]Center for Genomics and Personalized Medicine, CGPM, Aarhus, Denmark
[10]Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[11]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
[12]Department of Nutrition, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[13]Department of Psychosis, Aarhus University Hospital Skejby, Aarhus Nord, Denmark
[14]Department of Biomedicine (Human Genetics), Aarhus University, Aarhus, Denmark
[15]Bioinformatics Research Center, Aarhus University, 8000 Aarhus C, Denmark
[16]Department for Congenital Disorders, Statens Serum Institut, 2300 Copenhagen S, Denmark
[17]Psychosis Research Unit, Aarhus University Hospital – Psychiatry, Aarhus, Denmark
[18]Department of Affective Disorders, Aarhus University and Aarhus University Hospital –Psychiatry, Aarhus, Denmark
[19]Mental Health Services in the Capital Region of Denmark, Mental Health Center Copenhagen, University of Copenhagen, 2100 Copenhagen, Denmark
[20]Department of Clinical Medicine, University of Copenhagen, 2200 Copenhagen N, Denmark
[21]Department of Clinical Medicine, Institute of Biological Psychiatry, Mental Health Services, Copenhagen University Hospital, University of Copenhagen, 2200 Copenhagen N, Denmark
[22]Lundbeck Center for Geogenetics, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark
[23]Queensland Brain Institute, The University of Queensland, Brisbane, QLD 4072, Australia
[24]Department of Psychiatry, University of Oxford, Oxford OX3 7JX, UK
[25]Big Data Institute, University of Oxford, Oxford OX3 7LF, UK
[26]Queensland Centre for Mental Health Research, The Park Centre for Mental Health, Brisbane, QLD 4076, Australia
[27]These authors contributed equally
[28]Lead contact
*Correspondence: z.zhu.ncrr@au.dk (Z.Z.), j.mcgrath@uq.edu.au (J.J.M.)
https://doi.org/10.1016/j.xgen.2023.100457

SUMMARY

Complement components have been linked to schizophrenia and autoimmune disorders. We examined the association between neonatal circulating C3 and C4 protein concentrations in 68,768 neonates and the risk of six mental disorders. We completed genome-wide association studies (GWASs) for C3 and C4 and applied the summary statistics in Mendelian randomization and phenome-wide association studies related to mental and autoimmune disorders. The GWASs for C3 and C4 protein concentrations identified 15 and 36 independent loci, respectively. We found no associations between neonatal C3 and C4 concentrations and mental disorders in the total sample (both sexes combined); however, post-hoc analyses found that a higher C3 concentration was associated with a reduced risk of schizophrenia in females. Mendelian randomization based on C4 summary statistics found an altered risk of five types of autoimmune disorders. Our study adds to our understanding of the associations between C3 and C4 concentrations and subsequent mental and autoimmune disorders.

INTRODUCTION

The complement systems are an integral part of the innate immune response.[1–3] These phylogenetically ancient systems involve complex and interlinked amplification cascades, which can be triggered to protect the body from pathogens. Elements of the system are also involved in a range of additional physiological functions. For example, a growing body of evidence links elements of the complement systems (e.g., complement component 4; C4) to brain development, which could subsequently have implications for the risk of mental disorders.[4–7]

The coding gene (*C4* gene) of C4 is located within the major histocompatibility complex (MHC). It has two homologous isoforms (*C4A* and *C4B*), each of which can vary according to an insertion of a human endogenous retrovirus (*HERV*) transposon, and which can vary between one and three genocopies per haplotype.[8] Sekar et al.[9] reported a suite of coordinated studies that implicate an increased *C4A* copy number as a causal risk factor for schizophrenia. These studies include postmortem brain analyses (*C4A* mRNA expression was found to be higher in postmortem brain samples in schizophrenia versus control), genetic studies (fine-mapping, conditional-association, and allelic-series analyses based on Psychiatric Genomics Consortium Schizophrenia samples implicated *C4A* as a risk factor for schizophrenia), and experimental animal studies (mechanisms related to C4 and C3 were implicated in synaptic pruning). A subsequent study by Kamitaki et al.[10] extended these *C4A* findings and reported sex-related biases for both schizophrenia (associated with higher *C4A* copy number) and autoimmune disorders (in particular, systemic lupus erythematous [SLE] associated with lower C4A copy number). These findings are of interest with regard to neurodevelopmental disorders such as schizophrenia (SCZ), given evidence that *C4* and related members of the complement systems (e.g., C1q, C3) are involved in synaptic pruning during brain development.[11–15]

While the (premortem) measurement of brain C4 protein concentration is clearly not feasible in epidemiological studies, access to stored blood samples in biobanks provides a window into the association between circulating C4 concentration and the risk of a range of subsequent adverse health outcomes. We are aware of only one study that has measured neonatal C4 protein concentration in an SCZ case-control study (75 cases and 644 controls).[16] This study found evidence that an increased concentration of one of the two measured peptides within the protein encoded by *C4A* was associated with an increased risk of subsequent SCZ. There is a need for studies that examine both the underlying C4-related genetic variants associated with SCZ and measure the end product of these variants, such as circulating neonatal C4 protein concentration. Furthermore, in light of the shared genetic architecture among different types of mental disorders,[17] it is feasible that C4-related measures may also be associated with the risk of a wider range of mental disorders in addition to SCZ. We are not aware of studies that have previously examined this research question.

Apart from C4, there is evidence that complement component C3 (encoded by the C3 gene) is also involved in synaptic pruning,[11–15] and both C4 and C3 protein concentrations are often used in the monitoring of autoimmune disorders.[18] It is thus of interest to measure additional members of this cascade (e.g., C3 is immediately downstream of C4). A study based on induced pluripotent stem cells found an association between *C4A* copy number and neuronal C3 complement deposition.[15] Measuring C3 and C4 protein concentration in a large genotyped sample allows for genome-wide association study (GWAS) analyses, extending the previously published GWASs of serum complement components C3 and C4 (studies based on 3,495 Han Chinese men[19]). This information is required for post-GWAS analyses based on summary statistics (e.g., Mendelian randomization and phenome-wide association studies). Access to this information can also be used to explore other C4-related disease outcomes, in particular the link between increased *C4A* copy number with a decreased risk of autoimmune disorders.[10,20] Finally, we imputed *C4A* copy number, *C4* haplotypes, and brain *C4A* mRNA expression to examine the association between these variables and the (1) observed neonatal C4 protein concentration and (2) risk of six types of mental disorders. A summary of the overall methods is shown in Figure 1.

RESULTS

The iPSYCH2012 study, a population-based case-cohort study, was designed to investigate the genetic and environmental factors of six mental disorders (Table S1): SCZ, bipolar disorder (BIP), depression (DEP), autism spectrum disorder (ASD), attention deficit/hyperactivity disorder (ADHD), and anorexia nervosa (AN). The study included 80,873 individuals of multiple ancestries; 75,764 European individuals were retained by principal components projection (Figure S1). The following analyses in our study were based on the European ancestry individuals. We imputed C4 haplotypes using the reference data[9,10] (Table S2). Eight common *C4* haplotypes were imputed with allele frequency (AF) $\geq$0.01 (Table S3). Their frequencies were respectively *BS* (12%), *AL* (4%), *AL-BS* (23%), *AL-BL* (43%), *AL-BS-BS* (2%), *AL-AL* (11%), *AL-AL-BS* (3%), and *AL-AL-BL* (2%), consistent with the published studies.[10,21] We counted the copy numbers of the three types of *C4* alleles (*C4A*, *C4B*, and *HERV*, Figure S2). The copy numbers (i.e., count) for the different types of *C4* alleles were correlated. *C4A* count was negatively correlated with *C4B* count ($r = -0.52$, $p < 1.0 \times 10^{-100}$). *HERV* count was positively correlated with *C4A* count ($r = 0.73$, $p < 1.0 \times 10^{-100}$) but negatively correlated with *C4B* count ($r = -0.17$, $p < 1.0 \times 10^{-100}$).

There were 68,768 participants of European ancestry with measures of C3 and C4 protein concentrations. The distributions of the observed neonatal C3 and C4 protein concentrations were right skewed, with mean, median, SD, and interquartile range being respectively 7.1, 6.7, 3.6, and 5.1–9.2 mg/L for C3 protein concentration and 6.9, 6.5, 3.3, and 4.9–9.0 mg/L for C4 protein concentration (Table S4). Significant differences were observed in the protein concentrations between males and females (C3 protein concentration: difference = −0.32, SE = 0.03, $p = 6.5 \times 10^{-32}$; C4 protein concentration: difference = −0.30, SE = 0.03, $p = 2.7 \times 10^{-33}$). While the variance captured by sex was small (R^2 = 0.19% for C3 protein concentration and 0.20% for C4 protein concentration), we fitted sex as a covariate

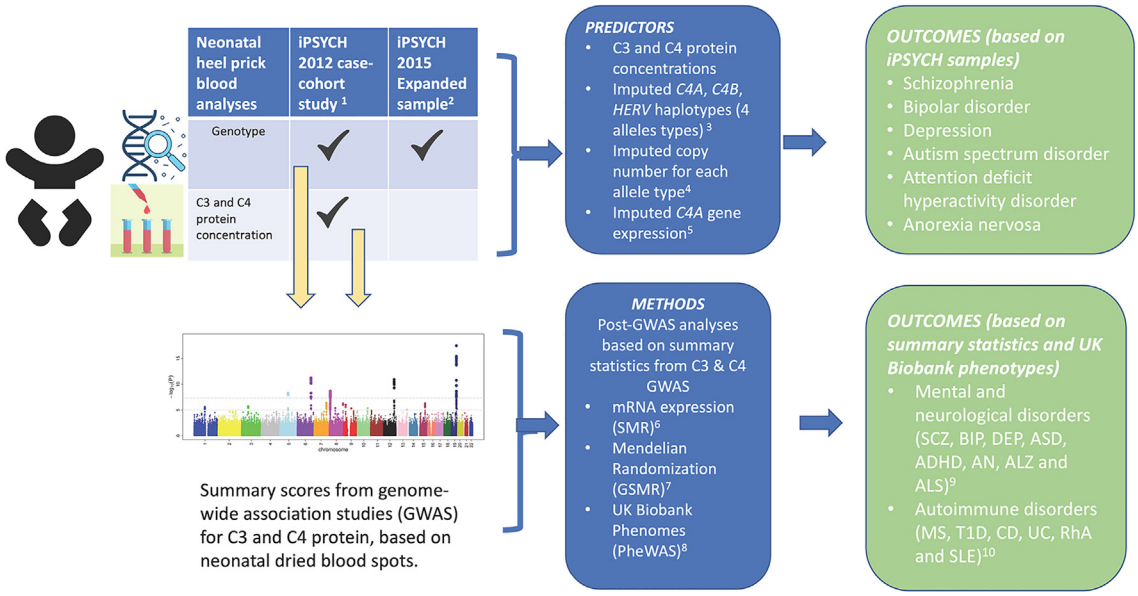


Figure 1. Concise summary of methods

(1) iPSYCH 2012 case-cohort study: N (protein concentration) = 68,768; N (genotype) = 75,764. (2) iPSYCH 2015 expanded sample provided additional genotyped samples: N ~ 56,000 samples. (3) *C4A* and *C4B* isotypes with or without HERV insertions were imputed (long and short, respectively) results in four alleles (*AL*, *AS*, *BL*, *BS*). See Tables S2 and S3. (4) Based on the imputed *C4A*, *C4B*, and *HERV* allele count, copy numbers were estimated. See Figure 3. (5) Based on Sekar et al.,[9] brain *C4A* expression was estimated based on imputed *C4* haplotypes. (6) We examined the association between variants associated with C4 and C3 protein concentration and mRNA expression in GTEx datasets, using summary-data-based Mendelian randomization (SMR). See Tables S12 and S16. (7) We examined the bidirectional association between variants associated with C4 and C3 protein concentration and summary statistics for a range of outcomes (see below) using generalized summary-data-based Mendelian randomization (GSMR). See Tables S22–S26. (8) We examined the association between variants associated with C4 and C3 protein concentration and a range of phenotypes identified in the UK Biobank. See Tables S27 and S28, and Figures S39 and S40. (9) The eight mental and neurological disorders were schizophrenia (SCZ), depression (DEP), bipolar disorder (BIP), autism spectrum disorder (ASD), attention deficit/hyperactivity disorder (ADHD), anorexia nervosa (AN), Alzheimer's disease (ALZ), and amyotrophic lateral sclerosis (ALS). (10) The six autoimmune disorders were multiple sclerosis (MS), type 1 diabetes (T1D), Crohn's disease (CD), ulcerative colitis (UC), rheumatoid arthritis (RhA), and systemic lupus erythematosus (SLE).

in the following analyses. To account for the influence of duration of storage (Figure S3) and between-protein assay plate variation, we regressed the concentrations of the plates using a linear mixed model (LMM) approach. The residuals were standardized (mean 0, variance 1) using rank-based inverse normal transformation. After standardization, the concentrations of C3 and C4 were positively correlated: $r_P = 0.65$ ($p < 1 \times 10^{-100}$, Figure S4).

With respect to the 5,109 individuals of non-European ancestry, we were able to infer 101 individuals of South Asian ancestry and 159 individuals of African ancestry. The remaining individuals of non-European ancestry could not be confidently allocated to any specific ancestry group, which is consistent with an earlier, separate analysis based on this same iPSYCH sample.[22] We imputed the *C4* haplotypes and counted the three *C4* alleles (*C4A*, *C4B*, and *HERV*) in the two additional ancestry groups using the same method as used for European individuals. The allele frequencies in African ancestry individuals were consistent with the study by Kamitaki et al.[10] (Tables S3C and S3D). The C3 and C4 neonatal protein concentrations were measured in 94 South Asian ancestry and 150 African ancestry individuals. The distributions of the two protein concentrations in the two additional ancestry groups were comparable to those found in individuals of European ancestry (Table S4B).

Heritability of C3 and C4 protein concentrations

The h^2 of C4 by Zaitlen's method[23] was 40% (SE = 0.03, p = 2.7×10^{-44}, Table S5) while the h^2_{SNP} was 26% (SE = 0.006, p < 1.0×10^{-100}). For C3, h^2 was 21% (SE = 0.03, p = 1.1×10^{-11}) and the h^2_{SNP} was 4% (SE = 0.005, p = 3.2×10^{-14}). The high genetic variance of C4 concentration was confirmed by BayesR[24,25]—the h^2_{SNP} was 24% for C4 (SE = 0.004, p = 1.0×10^{-100}) and 6% for C3 (SE = 0.005, p = 4.4×10^{-39}). These results indicate that both C3 and C4 concentrations were heritable traits. Moreover, we found that SNPs on the chromosome where the coding gene is located (*cis*-chr SNPs) explained a higher genetic variance than those on the remaining chromosomes (*trans*-chr SNPs) (Data S1 and Table S5). Especially for C4, using GREML,[9,16] $h^2_{cis\text{-}chr}$ = 14% (SE = 0.005, p < 1.0×10^{-100}) and $h^2_{trans\text{-}chr}$ = 4% (SE = 0.006, p = 9.4×10^{-13}). This indicates that SNPs positioned in the *C4* gene accounted for a substantial proportion of the genetic variance of C4 concentration.

The genetic correlation (r_g) between the two concentrations based on BOLT-REML[26] was 0.38 (Table S6, SE = 0.03, p = 1.9×10^{-35}), smaller than the phenotypic correlation (0.65). Given the large genetic effects at the coding genes for C3 and C4, we then estimated r_g using SNPs other than chromosomes 6 or 19 (related to the location of *C4* and *C3* genes, respectively)

to further investigate whether the correlation was driven by *cis*-chr SNPs, $r_g = 0.82$ (SE = 0.05, p = 4.8×10^{-65}). The high correlation was confirmed by Haseman-Elston regression[27] ($r_g = 0.78$, SE = 0.19, p = 4.1×10^{-5}) using *trans*-chr SNPs. These results indicated that C3 and C4 were genetically correlated, and this genetic correlation was not only driven by SNPs in or near their respective encoding genes.

We did not find that the SNP-based h^2 of C3/4 protein concentration differed between males and females (Data S2 and Table S5). The between-sex genetic correlation of C3/4 protein concentration (Table S6B) was, for C3 protein concentration, 0.74 (SE = 0.21, p[H0 $r_g = 1$] = 0.13), and for C4 protein concentration 0.97 (SE = 0.03, p[H0 $r_g = 1$] = 0.16). The findings from these analyses do not support the hypothesis that the genetic variation of C3/4 protein concentration differs between males and females.

GWASs of C3 and C4 protein concentrations with a focus in *cis*- and *trans*-protein quantitative trait loci

We used fastGWA[28] to conduct the GWAS analysis based on 68,768 participants of European ancestry and using 5,327,833 common SNPs, 5,201,724 in autosomes and 126,109 on the X chromosome (Figure 2). We conducted a Genome-wide Complex Trait Conditional and Joint Analysis (GCTA-COJO[29]) to help identify putative independent SNPs. For C4 protein concentration, 34 autosomal SNPs were identified as genome-wide significant (Table S7) and all were autosomal. For C3 protein concentration, 14 significant SNPs were identified, and again all were autosomal (Table S8).

Of the 34 SNPs significantly associated with C4 protein concentration, 30 (88.2%, 30/34) were found on chromosome 6. Of these, 29 were in the MHC region and 27 (79.4%, 27/34) SNPs were positioned within 2 Mb of the *C4* gene (chr6, 31.9 Mb). These 27 SNPs explained 16.7% of phenotypic variance in C4 concentration, which is consistent with the estimated $h^2_{\text{cis-chr}}$. SNP rs113720465 (32,005,355 bp, ~1 kb away from *C4B-AS1* [32,000–32,004 kb]) having the largest effect size (the A allele was associated with an increase of 0.76 SD units of C4 protein concentration); however, SNP rs3117579 had the smallest p value (within an exon of *GPANK1*). Given this large effect size, it is possible that SNPs in linkage disequilibrium (LD) at $R^2 < 0.01$ (the COJO threshold of independence) could also be reported as genome-wide significant through correlation. Thus, we conducted a GWAS fitting the COJO SNPs in and near the MHC region as fixed effects (Figure 2). We identified eight significant loci by COJO, six of which were significant from a GWAS of unadjusted C4 protein concentration. The two additional loci were on chromosomes 9 (rs6477754) and X (rs12012736). Interestingly, nearly all the eight COJO SNPs were annotated to the genes biologically related to complement-related pathways (Figures S5–S11). For example, *C4BPA* (rs12057769) encodes a binding protein of C4. The *IL6* gene (rs2066992) encodes a cytokine stimulated in response to infections and injuries. *C1S* (7.1 Mb on chr12) and *C1R* (7.2 Mb on chr12), the nearest genes of rs11064501, are the protein-coding genes of two C1 subcomponents.

With respect to C3 protein concentration, seven COJO SNPs were positioned within 2 Mb of the *C3* gene (chr19, 6.7 Mb)—these loci explained 3% of phenotypic variance in C3 concentration. After fitting these seven COJO SNPs as covariates, eight significant COJO SNPs were identified (Figures S12–S19). We found a SNP within the *ABO* gene, which has recently been identified as a "master regulator" of plasma protein concentration.[30,31] The gene annotations of the remaining SNPs encode proteins which involve immune- and/or C3-related pathways: (1) *FCGR2B* (rs844), which encodes an inhibitory receptor for the Fc region of immunoglobulin gamma (IgG); (2) *CFH* (rs558103 and rs11580821), which encodes complement factor H, a key factor that inhibits the alternative pathway and the amplification loop downstream of C3; (3) *STK19* gene (rs114492815), which is close to the *C4A* gene; and (4) *FAM117A* (rs12949906), which has enhanced gene expressions in dendritic cells (i.e., antigen-presenting cells) involved in the immune system.[32]

We then explored potential differences in effect sizes of SNPs on C3 and C4 protein concentrations between males and females. All the SNPs from the unadjusted GWASs were used in the analysis. No significant between-sex differences were found for these SNPs ($p < 5.0 \times 10^{-8}$).

For both GWASs of C3 and C4 protein concentration, we found no evidence of potential ascertainment bias related to the enrichment of cases with mental disorders in the iPSYCH2012 case-cohort study (Figures S20–S22). Therefore, the following post-GWAS analyses were based on the results from the full iPSYCH2012 sample.

C4 haplotypes are associated with C4 protein concentration

Both h^2_{SNP} and GWAS results indicated strong effects of the SNPs in the MHC region for both C4 and C3 protein concentrations. Due to the complex LD structure in this region, we used the imputed C4 haplotypes to investigate phenotypic associations of these genetic variants. We first examined the associations between the imputed *C4* haplotypes with the observed C4 protein concentration using an LMM approach. As expected, more copies of *C4* allele (either *C4A*, *C4B*, with or without *HERV*) were strongly associated with higher C4 protein concentration (Figure 3). The *C4A* copy number (b_{C4A}) had a greater effect than *C4B* (b_{C4B}) and *HERV* (b_{HERV}): $b_{\text{C4A}} = 0.3$ (Table S9, SE = 0.01, $p < 1.0 \times 10^{-100}$), $b_{\text{C4B}} = 0.2$ (SE = 0.01, $p < 1.0 \times 10^{-100}$), and $b_{\text{HERV}} = 0.2$ (SE = 0.004, $p < 1.0 \times 10^{-100}$). The *C4* copy numbers were correlated. Therefore, we fitted all three gene copy numbers in a regression model to estimate the joint effects. The *C4A* copy number had an effect nearly identical to that of *C4B* copy number, $b_{\text{C4A}} = 0.6$ (SE = 0.01, $p < 1.0 \times 10^{-100}$), $b_{\text{C4B}} = 0.6$ (SE = 0.01, $p < 1.0 \times 10^{-100}$). The beta estimates associated with the *HERV* copy number were less than the comparable estimates for *C4A* and *C4B* and were negatively associated with C4 protein concentration, $b_{\text{HERV}} = -0.08$ (SE = 0.005, p = 5.0×10^{-51}). This may reflect the strong correlation with *C4A* ($r = 0.73$) and negative correlation with *C4B* ($r = -0.17$). The result suggested one more copy of *C4A* or *C4B* is likely to have 1.6 mg/L (~0.6 × SD unit) higher C4 protein concentration given the same amount of *HERV*. We calculated the captured variance ($R^2 = s^2 b^2$) that was comparable between the *C4* copy numbers. In the formula, s^2 was variance of C4 copy number, analogous to variance of allele count. Of interest, the s^2 of *C4A* count was greater than *C4B* count (Table S3, $s^2 = 0.55$ for *C4A* and 0.31 for *C4B*).

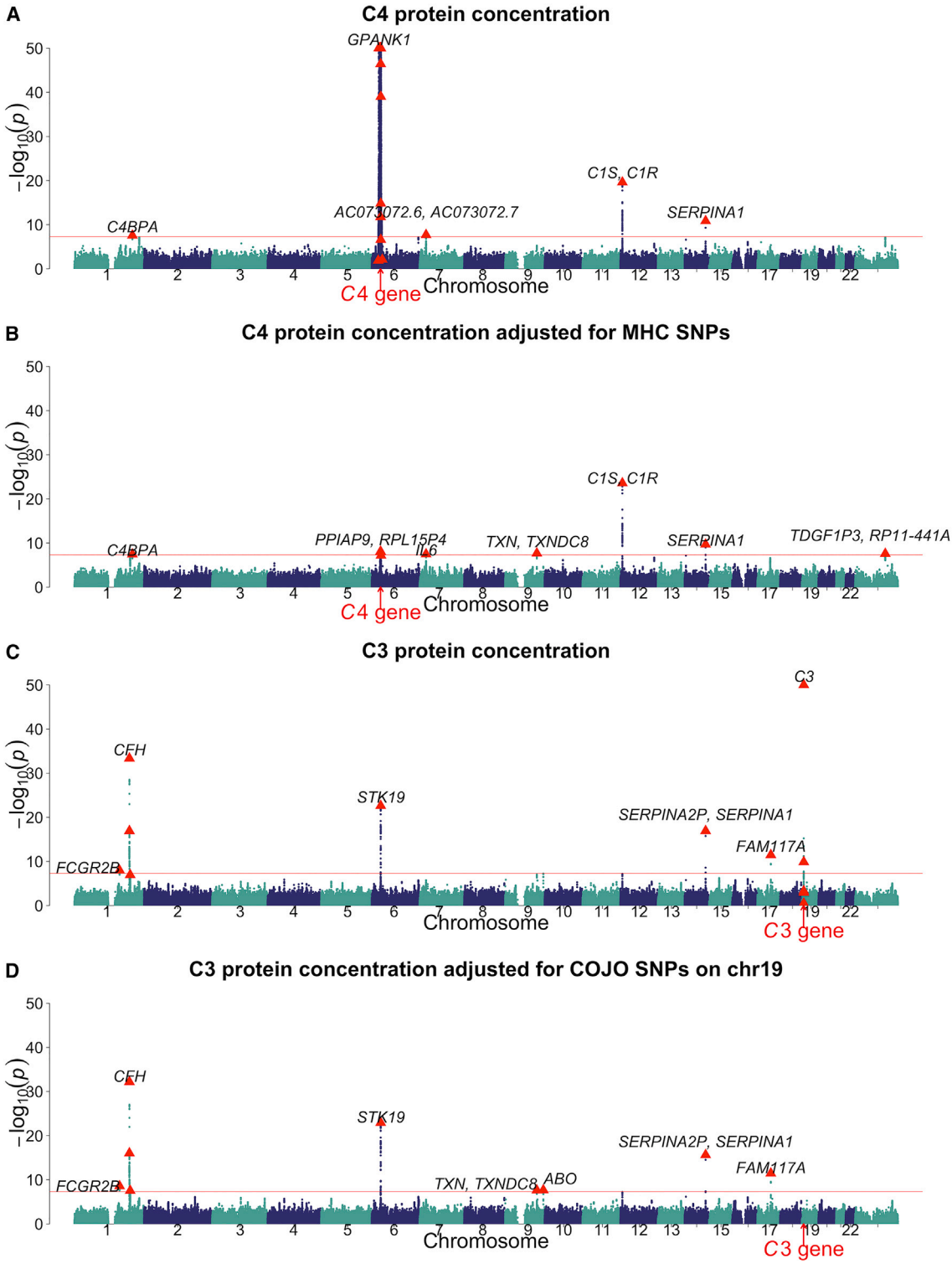


Figure 2. GWASs of neonatal C4 and C3 protein concentrations

(A) Unadjusted C4 protein concentration, (B) C4 protein concentration adjusted for the *cis*-pQTLs from COJO (fitted as covariates in the regression model), (C) unadjusted C3 protein concentration, and (D) C3 protein concentration adjusted for the *cis*-pQTLs from COJO. The COJO SNPs fitted as covariates in GWAS of adjusted protein concentration (B and D) were identified from GCTA-COJO of unadjusted protein concentration. The COJO SNPs are highlighted with red triangles. The location of the *C3* (on chromosome 19) and *C4* (on chromosome 6) genes are highlighted in the relevant panels. The top-associated SNPs were annotated with their overlapped or nearest genes. The GWAS threshold was 5.0×10^{-8}.

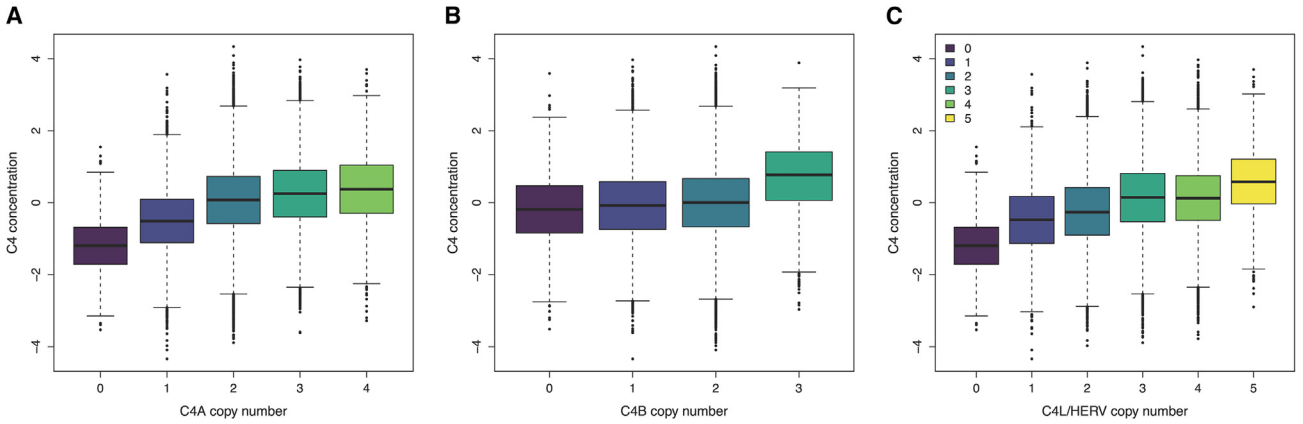


Figure 3. Plot of *C4* copy number versus C4 protein concentration

The center line shows the median value, and the lower and upper edges of the box shows the 25th and 75th quantiles respectively. The upper and lower whiskers extend 1.5 times the interquartile range from the top and bottom of the box. Individual outliers are shown as dots. There were three C4 alleles: (A) *C4A*, (B) *C4B*, and (C) *C4L/HERV*. The colors represent *C4* allele counts.

Therefore, *C4A* count had a larger contribution to C4 protein concentration than the *C4B* count (R^2 = 23% for *C4A* and 11% for *C4B*). In total, both counts captured 17.3% of variance in C4 protein concentration, accounting for the negative correlation between the two allele counts ($r = -0.52$). The captured genetic variance was in line with $h^2_{\text{cis-chr}}$, and the genetic variance at the MHC SNPs. In summary, the imputed counts of both *C4A* and *C4B* were associated with the observed C4 protein concentration, and the *C4A* count had a greater contribution than the *C4B* count.

Based on the effects of *C4* allele counts, we then examined the association between the commonly observed *C4* haplotypes and C4 protein concentration. For this analysis, we used the *BS* haplotype as the reference category because of (1) the positive association between *C4* allele count and C4 protein concentration and (2) the greater contribution of *C4A* count to C4 protein concentration. All the remaining seven common haplotypes were associated with increased C4 protein concentration. Due to their higher frequencies, *AL-BS* and *AL-BL* haplotypes captured greater variance of C4 protein concentration (Table S10, R^2 = 7.6% for *AL-BS* and 7.1% for *AL-BL*).

We then examined C3 protein concentration. In keeping with expectations, we did not identify any significant associations between either *C4* copy number or *C4* haplotype versus C3 protein concentration. However, we found the *AL-BS* haplotype was nominally significantly associated with C3 protein concentration ($b_{\text{AL-BS}} = 0.23$, SE = 0.03, $p = 5.4 \times 10^{-3}$). From the GWAS of C3 concentration, there was a COJO SNP positioned within the MHC region (rs114492815). While this SNP was in very weak association with each of the *C4* allele counts ($R^2 < 0.005$ for *C4A* count and *C4B* count, 0.01 for *C4L/HERV* count), there was a moderate association with *AL-BS* (R^2 = 0.11). Therefore, we ran the analysis again, fitting rs114492815 as an additional covariate. After this adjustment, none of the haplotypes were associated with the C3 concentration (Table S10). In general, C3 concentration was independent of *C4* alleles.

With respect to the associations in the individuals of non-European ancestry, we found a nominally significant association between a higher number of *C4B* allele copies and increased C4 protein concentration in African ancestry individuals (Table S9B, b = 0.43, SE = 0.14, $p = 3 \times 10^{-3}$). The imputation accuracy of *C4* haplotypes in individuals of African ancestry is lower than in European ancestry individuals.[10]

Functional mapping of GWAS findings

Having found the significant SNPs from the GWASs, we explored the genes associated with both concentrations using Multi-marker Analysis of GenoMic Annotation[33] nested in Functional Mapping and Annotation of Genome-wide Association Studies[34] (FUMA/MAGMA) and Summary-data-based Mendelian Randomization (SMR). The majority of the identified genes associated with C4 concentration were on chromosome 6, FUMA/MAGMA (257/263, Table S11), and SMR (55/56, Table S12) (Data S3). Interestingly, SMR found strong genetic correlates between higher brain *C4A* gene expressions in GTEx and neonatal C4 protein concentration in eight brain tissues (the mean of b_{XY} = 0.73). Subsequent analyses confirmed the associations between *C4A* expression in 15 brain-related tissues and neonatal circulating C4 protein concentration (Data S4; Tables S13 and S14; Figure S23). Overall, the findings indicate strong associations between genes in the MHC region and C4 protein concentration, and the *C4A* gene was likely to have a causal effect on C4 protein concentration in brain tissues. For C3 protein concentration, we identified 19 genes by FUMA/MAGMA (Table S15) and the *DXO* gene (chr6: 31.9 Mb) by SMR (Table S16). These significant genes associated with C3 and C4 protein concentrations were enriched with the Kyoto Encyclopedia of Genes and Genomes gene sets of SLE and complement systems (Figure S24).

Associations between C3 and C4 protein concentrations with mental disorders within the iPSYCH case-cohort study

In the models that accounted for the strong correlation between C3 and C4 concentration, we found no significant association between C3 concentration and any of the six mental disorders based on the entire sample (i.e., males and females

combined); however, post-hoc analyses found that higher C3 concentration was associated with a reduced risk of SCZ in females only (hazard ratio [HR] = 0.74, 95% confidence interval [CI] = 0.63–0.87, p = 2.36 × 10^{-4}) (Table S17). With respect to C4 concentration, we did not detect any association between any of the six mental disorders in analyses based on the entire sample or in post-hoc analyses stratified by sex (Table S17).

Associations between C3- and C4-related genotypes with mental disorders within the iPSYCH case-cohort study

We did not identify significant associations between *C4A*, *C4B*, or *HERV* copy numbers (Table S18) or C4-related haplotypes (Table S19) and any of the six mental disorders. Based on the formula between imputed *C4* haplotypes and observed *C4* gene expression (i.e., RNA concentration) in postmortem brain tissue,[9] we found no significant associations between these estimates and any of the six mental disorders (Table S20). However, we note that while the association between *C4A* copy number and SCZ was non-significant (HR = 1.27, 95% CI = 0.95–1.69, p = 0.11), the effect size was comparable to that reported by the larger sample in Sekar et al.[9] As a post-hoc analysis, we had access to the expanded iPSYCH2015 sample, which allowed us to re-examine this association with a larger sample size (original iPSYCH2012 sample cases = 2,515, controls = 51,751; expanded iPSYCH2015 sample cases = 4,398, controls = 77,368; these individuals were genetically unrelated). Additional methods and results based on the expanded sample can be found in Tables S18–S20). In the expanded sample, a nominally significant association was found (HR = 1.19, 95% CI = 1.06–1.35, p = 4.63 × 10^{-3}; the Bonferroni-corrected threshold is 2.8 × 10^{-3}, 0.05/[3 × 6]). None of the other mental disorders were associated with any of the C4-related genetic scores. Based on the sample sizes of cases and controls in iPSYCH2012, for each of the six mental disorders, we calculated the smallest detectable HRs of predicted brain *C4A/B* expression and C4 concentration (Table S21). For example, with respect to SCZ, we had sufficient power to detect an HR of brain *C4A* expression $\geq$1.60 and HR of C4 protein concentration $\geq$1.29. Larger sample sizes would be required to confidently detect small to medium-sized effects for several of the mental disorders included in this study.

We explored the associations between C4-related genotypes and six mental disorders in males and females. Due to the insufficient power in the iPSYCH2012 sample, we conducted the associations between copy numbers of three *C4* alleles (*C4A*, *C4B*, and *HERV*) versus six mental disorders in the iPSYCH2015 extension study. We found a nominally significant association between *C4A* copy number and SCZ in males (Table S18E, HR = 1.33, 95% CI = 1.11–1.59, p = 1.7 × 10^{-3}). The association was not significant in females (HR = 1.09, 95% CI = 0.93–1.29, p = 0.28).

While the iPSYCH case-cohort sample lacked power to undertake stand-alone GWASs for the six mental disorders included in the sample (see Figures S25–S36 for the GWAS analyses based on iPSYCH2012 and iPSYCH2015), we explored the findings from the most recent Psychiatric Genetics Consortium (PGC) for SCZ, BIP, and DEP. We inspected locus plots for the extended MHC region (Chr6, 23–38 Mb) for (1) the results of our C4 protein concentration GWAS (only available on the iPSYCH2012 sample), (2) PGC SCZ, (3) PGC BIP, and (4) PGC DEP. The findings of these three mental disorders were from the most recent GWAS[35] (Figure 4).

GSMR relationships with candidate neuropsychiatric and autoimmune disorders

We conducted Mendelian randomization analyses to examine relationships between the two protein concentrations (C3 and C4) and neuropsychiatric and autoimmune disorders (Figure 5 and Table S22). In the unadjusted forward analyses (i.e., all loci including the MHC region, with and without HEIDI filtering), higher C4 protein concentration was found to be associated with three mental disorders (Figure S37 and Table S23; SCZ, DEP, and BIP). The odds ratios (ORs) for these three findings were small (1.05 or less). We found that the majority of SNP instruments used in these generalized summary-data-based Mendelian randomization (GSMR) analyses were in and near the MHC region (e.g., for SCZ and BIP 126 out of 130 SNPs, and for DEP 103 out of 107 SNPs). Because of the strong LD in the MHC region, HEIDI-outlier methods designed to identify potential pleiotropic/confounding SNPs as outliers are unreliable (see Data S5). These three disorders also had significant findings in the unadjusted reverse analyses when using all variants (Table S25), which suggests the presence of pleiotropy from circulating C4 protein concentration to the three mental disorders. We advise caution in the interpretation of these GSMR findings.

We found protective effects of C4 concentration for several autoimmune disorders (Figure 5). In the unadjusted analyses, higher C4 concentration was associated with lower risks of multiple sclerosis (MS), type 1 diabetes (T1D), rheumatoid arthritis, and SLE. The effects were very large, especially for T1D (OR = 0.54, 95% CI = 0.50–0.58, N_{SNP} = 47) and SLE (OR = 0.37, 95% CI = 0.34–0.42, N_{SNP} = 103) (Figure S38 and Table S23). We identified that higher C4 concentration increased the risk of Crohn's disease (CD) (OR = 1.26, 95% CI = 1.19–1.34, N_{SNP} = 86). The strong association between neonatal C4 protein concentration and these autoimmune disorders was not indentified in reverse analyses (i.e., the association is unidirectional) (Table S24). When we examined the relationships adjusted for the MHC region SNPs, the significant association with SLE persisted. The effect size was comparable to that found using unadjusted C4 GWAS (with adjustment, OR = 0.24, 95% CI = 0.12–0.47, N_{SNP} = 7; without adjustment, OR = 0.37, 95% CI = 0.34–0.42, N_{SNP} = 103). Overall, these findings support the hypothesis that higher C4 protein concentration is causally related to a reduced risk of SLE—it is predicted that an increase of 2.46 mg/L (1 SD unit) of C4 concentration would be associated with a 76% reduced risk (1–0.24) of SLE.

We then explored the relationships between C3 concentration and neuropsychiatric and autoimmune disorders by bidirectional GSMR (Tables S25 and S26). Mindful that analyses based on fewer instruments may be underpowered to detect small effects, no significant associations were identified with pleiotropic SNPs removed. Our findings provide no support for the hypothesis that

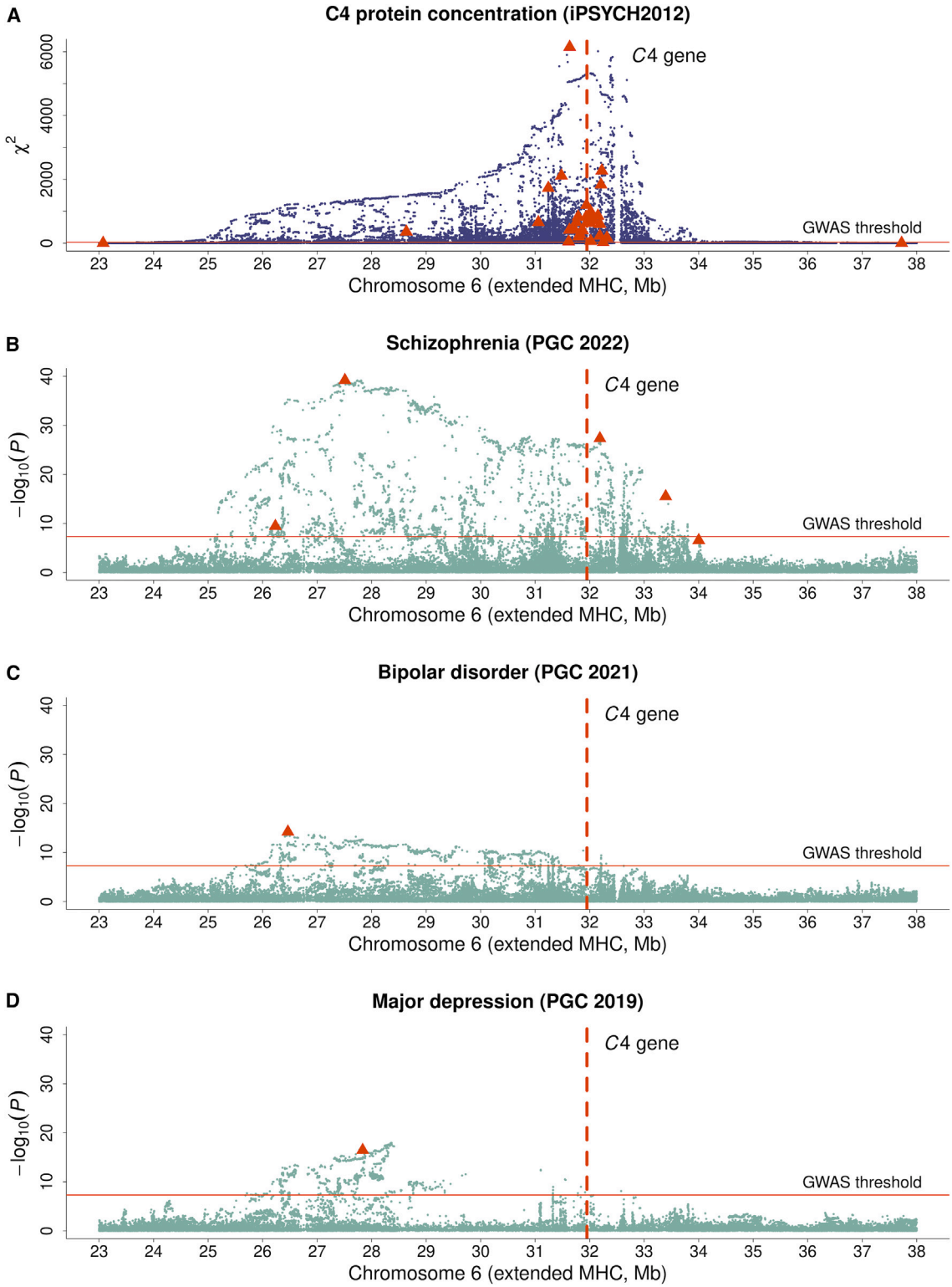


Figure 4. Locus plots for GWASs of C4 protein concentration and three mental disorders

The four panels show the differences in locations and distributions of the top-associated GWAS SNPs in the MHC region between C4 protein concentration (A) and three mental disorders, SCZ (B), BIP (C), and DEP (D). The y axis in the plots are squared *Z* scores from GWASs of (A) C4 protein concentration in iPSYCH2012, and $-\log_{10}$(p values) from GWASs of (B) SCZ,[35] (C) BIP,[36] and (D) DEP.[37] The GWAS summary statistics of these three mental disorders were from the

(legend continued on next page)

C3 protein concentration is related to the risk of the neuropsychiatric or autoimmune disorders examined in this study.

GSMR relationships between C3 and C4

Finally, we used Mendelian randomization analyses to explore the inter-relationships between C3 and C4 concentrations. Using all significant SNPs, we found bidirectional effects (C4→C3, b = 0.03, 95% CI = 0.02–0.05, p = 1.4 × 10^{-4}, N_{SNP} = 125; C3→C4, b = 0.03, 95% CI = 0.09–0.20, p = 1.4 × 10^{-6}, N_{SNP} = 10). The HEIDI-outlier method identified four potentially pleiotropic variants near the *C4* gene. When these were excluded, the C3→C4 findings were no longer significant (b = −0.03, 95% CI = −0.09–0.04, p = 0.43, N_{SNP} = 6); however the C4→C3 findings remained significant (b = 0.03, 95% CI = 0.02–0.05, p = 8.6 × 10^{-5}, N_{SNP} = 119). These unidirectional findings are consistent with the concentration of C3 being influenced by the "upstream" concentration of C4 (but not vice versa).

C3 and C4 phenome-wide association studies in the UK Biobank

We conducted phenome-wide association study (PheWAS) analysis in the UK Biobank, based on the polygenic scores (PGSs), which were predicted from SNPs across the whole genome. With respect to C4 protein concentration, we found 35 significant associations (Table S27 and Figure S39). Many of these were related to autoimmunity (Data S6). One of the top associations was SLE (ICD10 = M32, OR = 0.74, 95% CI = 0.69–0.80). There were no significant findings between C4 and any neuropsychiatric disorders (ICD10 F codes). No significant differences were found between the pattern of PheWAS findings between males and females. Overall, these findings lend weight to the hypotheses that neonatal C4 protein concentration is associated with an altered risk of autoimmune disorders, in particular SLE. There were no significant associations between C3 and any of the 1,148 phenotypes, which was in line with our GSMR findings (Table S28 and Figure S40).

DISCUSSION

Our findings provide new insights into the relationship between complement and the risks of mental and autoimmune disorders. First, we found associations between *C4*-related copy number and brain *C4A* expression versus measured circulating C4 protein concentration, which is consistent with the evidence from transgenic mouse experiments[38] and human observational studies.[19,39] The biologically plausible loci/genes identified in the C3 and C4 GWASs lend weight to the validity of our protein assays. Second, we found no association between neonatal circulating C4 protein concentration and risk of any of the mental disorders included in the case-cohort sample. In an expanded sample, we found support for a link between an increased *C4A* copy number and an increased risk of SCZ (N_{cases} = 4,398, $N_{controls}$ = 77,368, HR = 1.19, 95% CI = 1.06–1.35, p = 4.63 × 10^{-3}), which is broadly consistent with the findings from Sekar et al.[9] (N_{cases} = 28,799, $N_{controls}$ = 35,986). However, we found no link between the major *C4*-related haplotypes or imputed brain *C4A* RNA expression with risk of SCZ. Furthermore, there were no associations between these *C4*-related variables and any of the other five iPSYCH target psychiatric disorders. We also note that the PheWAS study found no significant associations between the summary statistics of C4 protein concentration and the UK Biobank-measured brain volumes (N = 28,613).

The relationship between C3-related measures and both psychiatric and autoimmune disorders was uniformly null, apart from a single post-hoc finding that higher circulating C3 protein concentration was associated with a decreased risk of schizophrenia in females only. C3 has been implicated in synaptic pruning,[11–15] and evidence suggests that estrogen may influence C3-related activation of microglia and subsequent phagocytosis of synapsis.[40] We hope that this finding can guide future hypothesis-driven research.

The genetic architecture of C3 and C4

The neonatal protein concentrations of both C3 and C4 were highly heritable. Both pedigree-based and SNP-based h^2 estimates (SE) were appreciable for C4: 0.40 (0.03) and 0.26 (0.006), respectively. The same estimates for C3 were smaller: 0.21 (0.03) and 0.04 (0.005), respectively. As expected, *cis*-protein quantitative trait loci (pQTLs) contributed to more than half of the genetic variance of their related proteins. Our sample sizes for C3 and C4 concentration GWASs were nearly 20 times larger than the only published GWAS for these proteins.[19] In the C4 GWAS, 30 quasi-independent hits were on chromosome 6, within the MHC region. Six additional loci were found on chromosomes 1, 7, 9, 12, 14, and X, and, reassuringly, several of these loci are linked to the complement system. We identified a locus on chromosome 1 within *C4BPA*, which encodes C4 binding protein (closely involved in C4 protein regulation). The locus on chromosome 12 (rs11064501) is adjacent to two genes that encode proteins involved in complement cascade initiation (C1s, C1R). Interestingly, a locus (rs12012736) was identified on the X chromosome. This locus may be one of the factors that contributed to the small sex differences found for the C3 and C4 protein concentrations and to the known sex differences in the risk of autoimmune disorders.[41]

With respect to C3 protein concentration, apart from loci within the *C3* gene (seven quasi-independent loci within this gene on chromosome 19), we found a locus within *FCGR2B* (Fc gamma receptor IIb), which encodes a receptor for the Fc region of IgG complexes. The IgG complex forms part of the machinery required for the phagocytosis of immune complexes. One locus in the MHC complex was identified, which is adjacent to the *C4A* gene. In keeping with the prior GWAS,[19] we identified two loci within *CFH*, the gene that encodes complement factor H. This protein

latest meta-analysis. The p values from GWAS of C4 protein at the *cis*-SNPs were extremely small. Squared Z scores (i.e., $Z = b/SE$) were used to show the significance at SNPs. The GWAS threshold was 5 × 10^{-8}, and the equivalent squared Z-score threshold was 29.8. The quasi-independent SNPs are highlighted with red triangles. The quasi-independent SNPs from GWAS of C4 were conducted from COJO (Table S7). The quasi-independent SNPs for the three types of mental disorders were based on those reported in the related publications.[35–37] The red dashed line represents the position of *C4* gene.

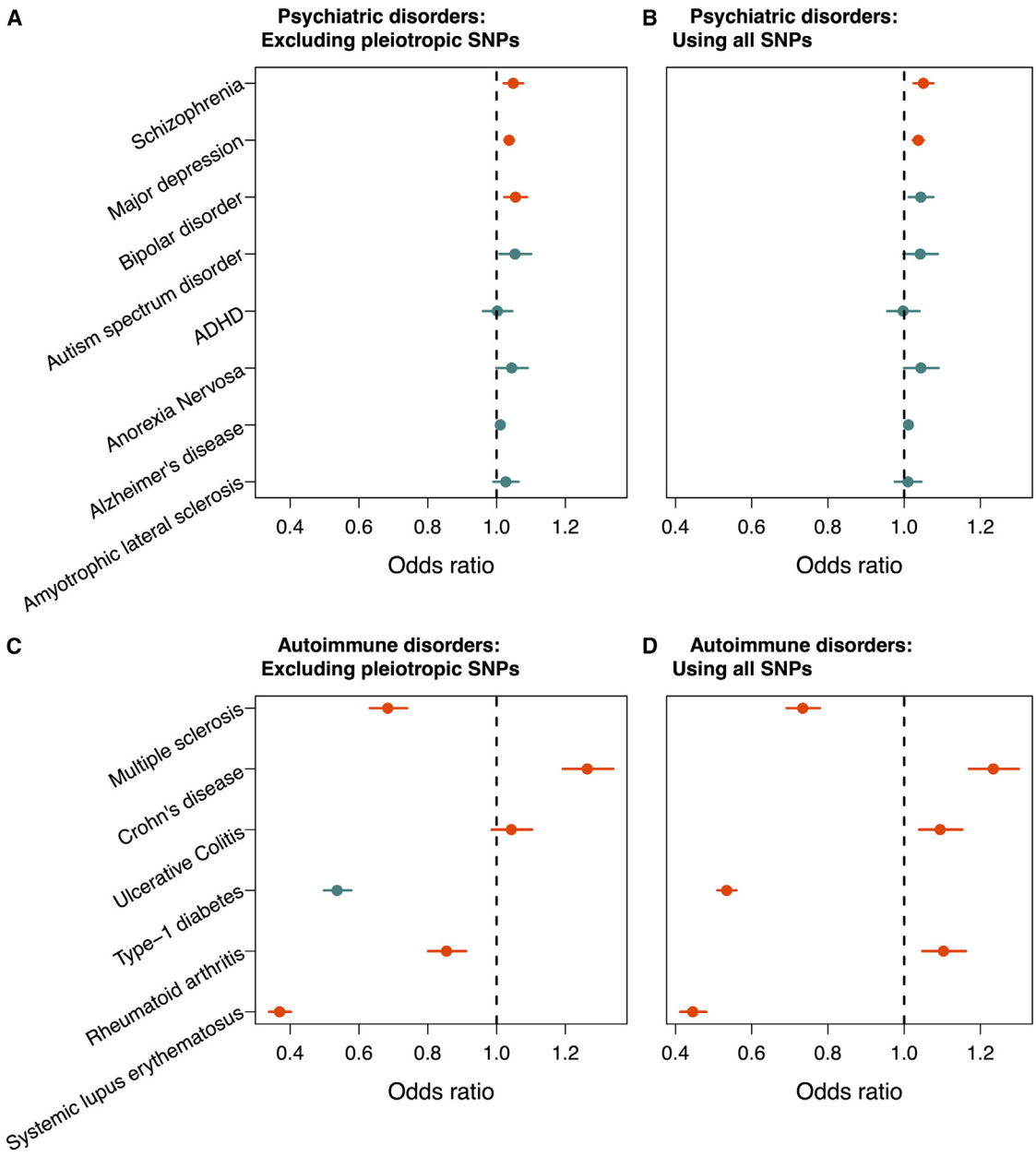


Figure 5. Forward GSMR results of C4 protein concentration

The GSMR analyses examined the relationships between circulating C4 protein concentration versus psychiatric disorders (A and B) and autoimmune disorders (C and D), using the unadjusted C4 GWAS. The pleiotropic SNPs identified by HEIDI-outlier were excluded in (A) and (C). All genome-wide significant SNPs were used in (B) and (D). The dot symbols show the estimates, and 95% confidence intervals are provided. The red dots represent the significant results with p value of GSMR $<1.8 \times 10^{-3}$, the Bonferroni-corrected threshold.

is involved in complement regulation and has been linked to several disease phenotypes (most notably with age-related macular degeneration).[42] *CHF* specifically regulates C3, which slows the downstream complement activation. We also found a locus within *ABO*, which was identified as having associations with over 50 other protein concentrations[30,31]; thus, variants in this gene could directly or indirectly influence generic protein metabolic pathways (e.g., upstream metabolic steps and downstream protein degradation and excretion). In summary, our study has highlighted how genetic variants within several components of the complement cascades (i.e., at the systems level) could influence the concentration of key circulation proteins such as C3 and C4. We have summarized these findings in Figure 6.

Findings linking C4 with autoimmune disorders

We found convergent evidence linking higher C4 protein concentration and an altered risk of autoimmune disorders. Based on Mendelian randomization analyses, there was robust evidence

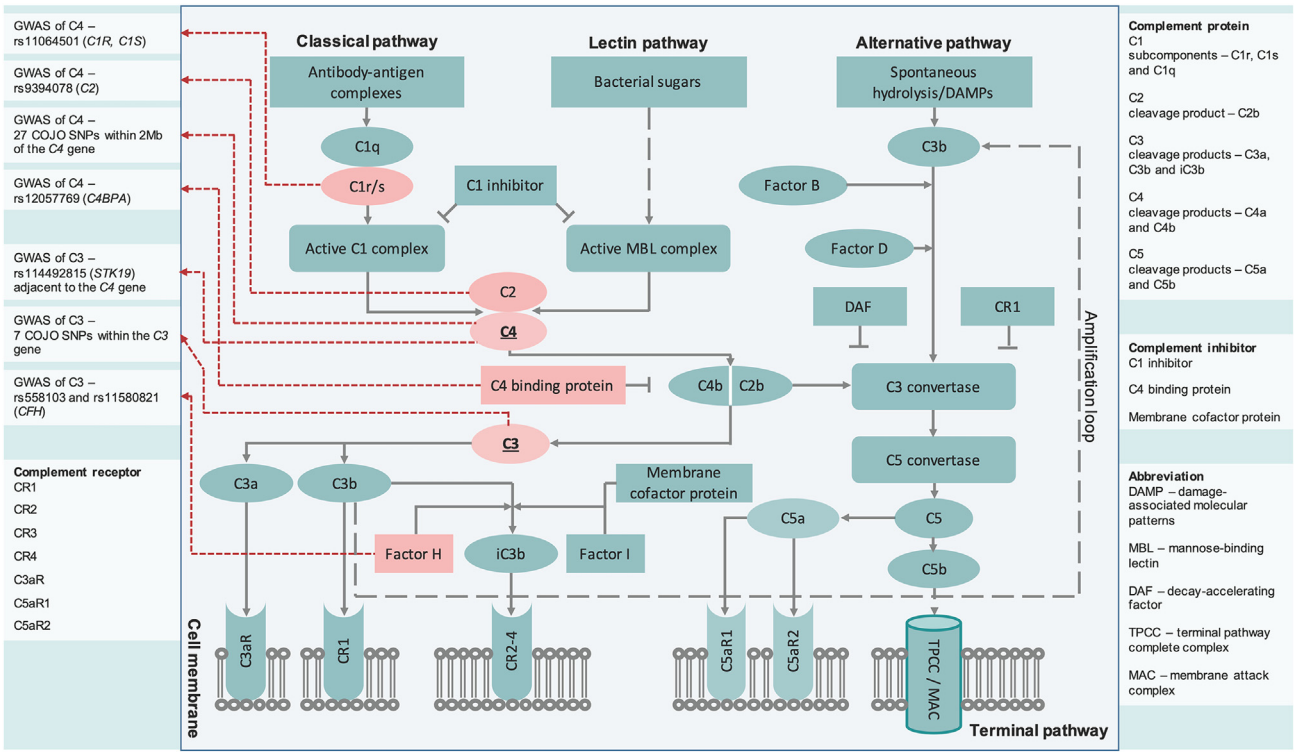


Figure 6. Summary of the results from GWASs of neonatal C3 and C4 protein concentrations displayed within the complement cascade
For significant loci identified from COJO, proteins encoded by annotated genes are highlighted in red.

with respect to a lower risk of SLE. Reassuringly, the UK Biobank-based PheWASs found that variants associated with increased neonatal C4 protein concentration were associated with (1) reduced risks of a wide range of disorders (including celiac disease, thyrotoxicosis, hypothyroidism, T1D, sarcoidosis, SLE, nephrotic syndrome, and MS; Sjögren's syndrome was nominally significant) and (2) increased risks of several disorders (including psoriasis, ankylosing spondylitis, and iridocyclitis; CD was nominally significant). Our findings are consistent with a meta-analysis based on 16 case-control studies, where low *C4* gene copy number (<4) was associated with an increased risk of any type of autoimmune disorder, including SLE.[20] A study based on large-scale genetic and transcriptomic datasets by Kim et al.[43] suggested that *C4A*-related gene expression was not associated with risk of SCZ-related synaptic gene expression but was associated with disorders including inflammatory bowel disease, RA, and SLE. Our findings also support stronger associations of *C4A* with immune disorders compared to SCZ. Variants in *C4A* and *C4B*, which were thought to increase the risk for SCZ, are protective for two autoimmune disorders (SLE and Sjögren's syndrome).[10] The mechanisms of action underpinning the links between C4 and both increased and decreased risk of different autoimmune disorders remain poorly understood.[44–46]

Strengths and limitations of the study

Our study has several strengths. Our sample was nearly 20 times larger than other published GWASs of C3 and C4.[19] With respect to the hypothesis linking complement to brain development, our complement assays were collected from neonatal samples (versus adult samples). Because the onset of mental disorders such as SCZ is often in the second and third decade of life, our samples are unlikely to be impacted by reverse causation (e.g., smoking may be linked to complement gene expression in the brain[43]), and medication effects may impact on postmortem gene expression studies.[47] Sager et al.[13] examined *C4* mRNA expression and C4 protein expression in human brain tissue from neonatal to young adult age points. These authors note that *C4* mRNA and C4 protein brain expression are both more prominent in early life compared to adolescence/young adulthood. A recent study from Hernandez et al.[48] reported good agreement between imputed *C4A* and *C4B* mRNA expression as published by Sekar et al.[9] (based on adult brain tissue) versus C4A expression in EUR PsychENCODE in samples aged between 5 and 15 years. It would be of interest to examine whether our findings based on adult brain mRNA expression were consistent with fetal brain tissue. There is evidence from animal experiments (mouse pups at postnatal days 5 and 10)[38,49] suggesting that C4-related processes impact on synapse elimination during this early postnatal phase.

With respect to limitations, because our samples were based on neonatal C3 and C4 protein concentrations, it remains to be seen whether the genetic correlates we identified for these proteins remain stable across the lifespan. Furthermore, there is evidence from adult samples (N = 47) that the serum

concentration of C4 protein is not correlated with that measured in the cerebrospinal fluid.[50] However, we examined the correlation between the effect sizes for SNPs associated with (1) circulating neonatal C4 protein concentration versus (2) the expression of *C4A* mRNA reported in brain tissue (using GTEx data[51]). These findings lend weight to the hypothesis that there are shared genetic influences on circulating neonatal C4 protein concentration and brain *C4A* mRNA expression across the lifespan.

Also, we used an antibody that has been demonstrated to measure total C4 (i.e., both C4A and C4B), so we are unable to isolate the concentrations of the two isoforms. The C3 and C4 concentrations in our study were derived from circulating plasma proteins, whereas the concentration of these proteins may vary between organs/tissues and in response to local tissue activation pathways. Apart from the genetic factors influencing neonatal circulating C3 and C4 protein concentration, it is feasible that exposures such as prenatal infection and maternal immune activation,[52,53] and obstetric complications (e.g., hypoxia),[54] may influence neonatal C3 and C4 protein concentration. We plan to explore these issues in future studies.

The mental disorders examined in this study were based on registers, which only cover inpatient, outpatient, and accident-emergency sites. These registers do not include people who do not seek help for their condition or who are treated only by their general practitioners. While studies have reported good validity for these register-based diagnoses (compared to research criteria),[55–60] it is known that Danish registers are biased, with milder disorders (e.g., depression) being under-represented.[61,62]

Our sample may have been underpowered to confidently detect small to medium-sized relationships between neonatal circulating C4 protein concentration and the risk of the six mental disorders. We estimate that our sample had sufficient power (assuming 90% of power and significance level of 0.05) to confidently detect an increased neonatal C4 protein concentration for higher risks of our target mental disorders, which ranged between 11% for ASD to 66% for BIP (Table S21).

Conclusions

Our study provides new insights into the genetic and phenotypic correlates of C3 and C4 protein concentration and helps unravel the contribution of different C4-related copy numbers and haplotypes to C4 protein concentration. We found no evidence that either C3 or C4 neonatal protein concentration was associated with any of the six mental disorders examined. However, we found an association between increased *C4A* copy number and an increased risk of SCZ, and evidence from Mendelian randomization suggests that some genetic variants may be pleiotropic for C4 and three mental disorders (i.e., SCZ, bipolar disorder, and depression). We found convergent evidence linking C4 protein concentration and an altered risk of autoimmune disorders. Apart from a post-hoc finding between higher neonatal circulating C3 concentration and a reduced risk of schizophrenia in females, we found no other evidence linking neonatal C3 protein concentration and risk of either mental or autoimmune disorders. We hope that our findings can guide future research related to the association between two complement components (C3 and C4) and health outcomes.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - The iPSYCH2012 study
  - Ethical framework
  - C3 and C4 protein concentrations
  - Imputation of genotypes
  - Imputation of *C4* haplotypes
  - Quality control of the C3 and C4 protein concentrations
  - Heritability and SNP-based heritability of the C3 and C4 protein concentrations
  - GWAS of C3 and C4 protein concentrations
  - Associations between C4 haplotypes and protein concentrations
  - FUMA/MAGMA and SMR
  - Associations between C4 haplotypes and mental disorders observed within the iPSYCH case-cohort study
  - Associations between protein concentrations and mental disorders observed within the iPSYCH2012 case-cohort study
  - Mendelian Randomization analysis based on summary statistics
  - PheWAS based on UK Biobank phenotypes

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2023.100457.

ACKNOWLEDGMENTS

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. This study used the GTEx data under the database of Genotypes and Phenotypes (dbGaP) study, dbGaP: phs000424. This research has been conducted using the data resource under dbGaP: phs001992, and using the UK Biobank Resource under Application Number 12505. The authors thank GenomeDK and Aarhus University for providing computational resources and support that contributed to these research results. We thank Dr. Joana A. Revez for her insightful comments on BayesR and R tools. This study was supported by the Danish National Research Foundation, via a Niels Bohr Professorship to J.J.M. B.J.V. was also supported by a Lundbeck Foundation Fellowship (R335-2019-2339). This research was conducted using the Danish National Biobank resource, supported by the Novo Nordisk Foundation. The iPSYCH team was supported by grants from the Lundbeck Foundation (R102-A9118, R155-2014-1724, and R248-2017-2003), NIMH (1R01MH124851-01 to A.D.B.), and the Universities and University Hospitals of Aarhus and Copenhagen. High-performance computer capacity for handling and statistical analysis of iPSYCH data on the GenomeDK HPC facility was provided by the Center for Genomics and Personalized Medicine and the Centre for Integrative Sequencing, iSEQ, Aarhus University, Denmark (grant to A.D.B.). The Anorexia Nervosa Genetics Initiative (ANGI) was an initiative of the Klarman Family Foundation. Genotyping of the AN samples was funded by the Klarman Family Foundation. M.E.B. was

supported by the Independent Research Fund Denmark (grant number 7025-00078B) and by an unrestricted grant from The Lundbeck Foundation (grant number R268-2016-3925); J.-C.P.G.D. was supported by a grant from the Danish Council for Independent Research (grant number 0134-00227B); A.F.M. was supported by an ARC Future Fellowship (FT200100837); K.L.M. was supported by grants from The Lundbeck Foundation and the Brain & Behavior Research Foundation; N.R.W. was supported by NHMRC 1173790 and 1113400; L.V.P. was supported by NIMH (R01MH120170) and The Lundbeck Foundation (grant no. R276-2018-4581); and C.M.B. was supported by NIMH (R56MH129437, R01MH120170, R01MH124871, R01MH119084, R01MH118278, R01 MH124871), Brain and Behavior Research Foundation Distinguished Investigator grant, Swedish Research Council (Vetenskapsrådet, award number 538-2013-8864), and the Lundbeck Foundation (grant number R276-2018-4581).

AUTHOR CONTRIBUTIONS

Funding acquisition, C.M.B., L.V.P., and J.J.M.; conceptualization, N.B.-L., Z.Z., K.S., and J.J.M.; resources, D.M.H.; supervision, D.M.H., P.B.M., N.R.W., B.J.V., and J.J.M.; methodology, N.B.-L., Z.Z., A.D.B., D.M.H., O.M., P.B.M., M.N., K.S., N.R.W., and T.W.; validation and investigation, N.B.-L.; data curation, software, and visualization, Z.Z.; formal analysis, Z.Z., C.A., B.B., F.P., and J.S.; writing – original draft, N.B.-L., Z.Z., and J.J.M.; writing – review & editing, N.B.-L., Z.Z., E.A., C.A., M.E.B., B.B., A.D.B., C.M.B., J.-C.P.G.D., J.G., D.M.H., A.F.M., O.M., P.B.M., K.L.M., M.N., L.V.P., F.P., J.S., K.S., N.R.W., T.W., B.J.V., and J.J.M.

DECLARATION OF INTERESTS

C.M.B. reports Pearson (author, royalty recipient) and Equip Health Inc. (Stakeholder Advisory Board). B.J.V. reports Allelica (Scientific Advisory Board).

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: June 6, 2023
Revised: September 3, 2023
Accepted: November 8, 2023
Published: December 13, 2023

REFERENCES

1. Minton, K. (2014). Innate immunity: The inside story on complement activation. Nat. Rev. Immunol. *14*, 61.
2. Merle, N.S., Church, S.E., Fremeaux-Bacchi, V., and Roumenina, L.T. (2015). Complement System Part I - Molecular Mechanisms of Activation and Regulation. Front. Immunol. *6*, 262.
3. Reis, E.S., Mastellos, D.C., Hajishengallis, G., and Lambris, J.D. (2019). New insights into the immune functions of complement. Nat. Rev. Immunol. *19*, 503–516.
4. Mayilyan, K.R., Weinberger, D.R., and Sim, R.B. (2008). The complement system in schizophrenia. Drug News Perspect. *21*, 200–210.
5. Magdalon, J., Mansur, F., Teles E Silva, A.L., de Goes, V.A., Reiner, O., and Sertié, A.L. (2020). Complement System in Brain Architecture and Neurodevelopmental Disorders. Front. Neurosci. *14*, 23.
6. Stephan, A.H., Barres, B.A., and Stevens, B. (2012). The complement system: an unexpected role in synaptic pruning during development and disease. Annu. Rev. Neurosci. *35*, 369–389.
7. Presumey, J., Bialas, A.R., and Carroll, M.C. (2017). Complement System in Neural Synapse Elimination in Development and Disease. Adv. Immunol. *135*, 53–79.
8. Blanchong, C.A., Chung, E.K., Rupert, K.L., Yang, Y., Yang, Z., Zhou, B., Moulds, J.M., and Yu, C.Y. (2001). Genetic, structural and functional diversities of human complement components C4A and C4B and their mouse homologues, Slp and C4. Int. Immunopharmacol. *1*, 365–392.
9. Sekar, A., Bialas, A.R., de Rivera, H., Davis, A., Hammond, T.R., Kamitaki, N., Tooley, K., Presumey, J., Baum, M., Van Doren, V., et al. (2016). Schizophrenia risk from complex variation of complement component 4. Nature *530*, 177–183.
10. Kamitaki, N., Sekar, A., Handsaker, R.E., de Rivera, H., Tooley, K., Morris, D.L., Taylor, K.E., Whelan, C.W., Tombleson, P., Loohuis, L.M.O., et al. (2020). Complement genes contribute sex-biased vulnerability in diverse disorders. Nature *582*, 577–581.
11. Stevens, B., and Johnson, M.B. (2021). The complement cascade repurposed in the brain. Nat. Rev. Immunol. *21*, 624–625.
12. Stevens, B., Allen, N.J., Vazquez, L.E., Howell, G.R., Christopherson, K.S., Nouri, N., Micheva, K.D., Mehalow, A.K., Huberman, A.D., Stafford, B., et al. (2007). The classical complement cascade mediates CNS synapse elimination. Cell *131*, 1164–1178.
13. Sager, R.E.H., Walker, A.K., Middleton, F., Robinson, K., Webster, M.J., and Weickert, C.S. (2021). Trajectory of change in brain complement factors from neonatal to young adult humans. J. Neurochem. *157*, 479–493.
14. Westacott, L.J., and Wilkinson, L.S. (2022). Complement Dependent Synaptic Reorganisation During Critical Periods of Brain Development and Risk for Psychiatric Disorder. Front. Neurosci. *16*, 840266.
15. Sellgren, C.M., Gracias, J., Watmuff, B., Biag, J.D., Thanos, J.M., Whittredge, P.B., Fu, T., Worringer, K., Brown, H.E., Wang, J., et al. (2019). Increased synapse elimination by microglia in schizophrenia patient-derived models of synaptic pruning. Nat. Neurosci. *22*, 374–385.
16. Cooper, J.D., Ozcan, S., Gardner, R.M., Rustogi, N., Wicks, S., van Rees, G.F., Leweke, F.M., Dalman, C., Karlsson, H., and Bahn, S. (2017). Schizophrenia-risk and urban birth are associated with proteomic changes in neonatal dried blood spots. Transl. Psychiatry *7*, 1290.
17. Cross-Disorder Group of the Psychiatric Genomics Consortium; Lee, S.H., Ripke, S., Neale, B.M., Faraone, S.V., Purcell, S.M., Perlis, R.H., Mowry, B.J., Thapar, A., Goddard, M.E., et al. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat. Genet. *45*, 984–994.
18. Hebert, L.A., Cosio, F.G., and Neff, J.C. (1991). Diagnostic significance of hypocomplementemia. Kidney Int. *39*, 811–821.
19. Yang, X., Sun, J., Gao, Y., Tan, A., Zhang, H., Hu, Y., Feng, J., Qin, X., Tao, S., Chen, Z., et al. (2012). Genome-wide association study for serum complement C3 and C4 levels in healthy Chinese subjects. PLoS Genet. *8*, e1002916.
20. Li, N., Zhang, J., Liao, D., Yang, L., Wang, Y., and Hou, S. (2017). Association between C4, C4A, and C4B copy number variations and susceptibility to autoimmune diseases: a meta-analysis. Sci. Rep. *7*, 42628.
21. Bian, B., Couvy-Duchesne, B., Wray, N.R., and McRae, A.F. (2022). The role of critical immune genes in brain disorders: insights from neuroimaging immunogenetics. Brain Commun. *4*, fcac078.
22. Albiñana, C., Zhu, Z., Borbye-Lorenzen, N., Boelt, S.G., Cohen, A.S., Skogstrand, K., Wray, N.R., Revez, J.A., Privé, F., Petersen, L.V., et al. (2023). Genetic correlates of vitamin D-binding protein and 25-hydroxyvitamin D in neonatal dried blood spots. Nat. Commun. *14*, 852.
23. Zaitlen, N., Kraft, P., Patterson, N., Pasaniuc, B., Bhatia, G., Pollack, S., and Price, A.L. (2013). Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. PLoS Genet. *9*, e1003520.
24. Moser, G., Lee, S.H., Hayes, B.J., Goddard, M.E., Wray, N.R., and Visscher, P.M. (2015). Simultaneous discovery, estimation and prediction analysis of complex traits using a bayesian mixture model. PLoS Genet. *11*, e1004969.
25. Zeng, J., de Vlaming, R., Wu, Y., Robinson, M.R., Lloyd-Jones, L.R., Yengo, L., Yap, C.X., Xue, A., Sidorenko, J., McRae, A.F., et al. (2018).

Signatures of negative selection in the genetic architecture of human complex traits. Nat. Genet. *50*, 746–753.

26. Loh, P.R., Bhatia, G., Gusev, A., Finucane, H.K., Bulik-Sullivan, B.K., Pollack, S.J., Schizophrenia Working Group of Psychiatric Genomics Consortium; de Candia, T.R., Lee, S.H., Wray, N.R., et al. (2015). Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. Nat. Genet. *47*, 1385–1392.
27. Elston, R.C., Buxbaum, S., Jacobs, K.B., and Olson, J.M. (2000). Haseman and Elston revisited. Genet. Epidemiol. *19*, 1–17.
28. Jiang, L., Zheng, Z., Qi, T., Kemper, K.E., Wray, N.R., Visscher, P.M., and Yang, J. (2019). A resource-efficient tool for mixed model association analysis of large-scale data. Nat. Genet. *51*, 1749–1755.
29. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Genetic Investigation of ANthropometric Traits GIANT Consortium; DIAbetes Genetics Replication And Meta-analysis DIAGRAM Consortium; Madden, P.A.F., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat. Genet. *44*, 369–375. S1-S3.
30. Ferkingstad, E., Sulem, P., Atlason, B.A., Sveinbjornsson, G., Magnusson, M.I., Styrmisdottir, E.L., Gunnarsdottir, K., Helgason, A., Oddsson, A., Halldorsson, B.V., et al. (2021). Large-scale integration of the plasma proteome with genetics and disease. Nat. Genet. *53*, 1712–1721.
31. Pietzner, M., Wheeler, E., Carrasco-Zanini, J., Cortes, A., Koprulu, M., Wörheide, M.A., Oerton, E., Cook, J., Stewart, I.D., Kerrison, N.D., et al. (2021). Mapping the proteo-genomic convergence of human diseases. Science *374*, eabj1541.
32. Karlsson, M., Zhang, C., Méar, L., Zhong, W., Digre, A., Katona, B., Sjöstedt, E., Butler, L., Odeberg, J., Dusart, P., et al. (2021). A single-cell type transcriptomics map of human tissues. Sci. Adv. *7*, eabh2169.
33. de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. *11*, e1004219.
34. Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. Nat. Commun. *8*, 1826.
35. Trubetskoy, V., Pardiñas, A.F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T.B., Bryois, J., Chen, C.Y., Dennison, C.A., Hall, L.S., et al. (2022). Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature *604*, 502–508.
36. Mullins, N., Forstner, A.J., O'Connell, K.S., Coombes, B., Coleman, J.R.I., Qiao, Z., Als, T.D., Bigdeli, T.B., Børte, S., Bryois, J., et al. (2021). Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. Nat. Genet. *53*, 817–829.
37. Howard, D.M., Adams, M.J., Clarke, T.K., Hafferty, J.D., Gibson, J., Shirali, M., Coleman, J.R.I., Hagenaars, S.P., Ward, J., Wigmore, E.M., et al. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat. Neurosci. *22*, 343–352.
38. Yilmaz, M., Yalcin, E., Presumey, J., Aw, E., Ma, M., Whelan, C.W., Stevens, B., McCarroll, S.A., and Carroll, M.C. (2021). Overexpression of schizophrenia susceptibility factor human complement C4A promotes excessive synaptic loss and behavioral changes in mice. Nat. Neurosci. *24*, 214–224.
39. Yang, Y., Chung, E.K., Zhou, B., Blanchong, C.A., Yu, C.Y., Füst, G., Kovács, M., Vatay, A., Szalai, C., Karádi, I., and Varga, L. (2003). Diversity in intrinsic strengths of the human complement system: serum C4 protein concentrations correlate with C4 gene size and polygenic variations, hemolytic activities, and body mass index. J. Immunol. *171*, 2734–2745.
40. Yang, H., Oh, C.K., Amal, H., Wishnok, J.S., Lewis, S., Schahrer, E., Trudler, D., Nakamura, T., Tannenbaum, S.R., and Lipton, S.A. (2022). Mechanistic insight into female predominance in Alzheimer's disease based on aberrant protein S-nitrosylation of C3. Sci. Adv. *8*, eade0764.
41. Schurz, H., Salie, M., Tromp, G., Hoal, E.G., Kinnear, C.J., and Möller, M. (2019). The X chromosome and sex-specific effects in infectious disease susceptibility. Hum. Genomics *13*, 2.
42. Poppelaars, F., Goicoechea de Jorge, E., Jongerius, I., Baeumner, A.J., Steiner, M.S., Józsi, M., Toonen, E.J.M., and Pauly, D.; SciFiMed consortium (2021). A Family Affair: Addressing the Challenges of Factor H and the Related Proteins. Front. Immunol. *12*, 660194.
43. Kim, M., Haney, J.R., Zhang, P., Hernandez, L.M., Wang, L.K., Perez-Cano, L., Loohuis, L.M.O., de la Torre-Ubieta, L., and Gandal, M.J. (2021). Brain gene co-expression networks link complement signaling with convergent synaptic pathology in schizophrenia. Nat. Neurosci. *24*, 799–809.
44. Jain, U., Otley, A.R., Van Limbergen, J., and Stadnyk, A.W. (2014). The complement system in inflammatory bowel disease. Inflamm. Bowel Dis. *20*, 1628–1637.
45. Cleynen, I., Konings, P., Robberecht, C., Laukens, D., Amininejad, L., Théâtre, E., Machiels, K., Arijs, I., Rutgeerts, P., Louis, E., et al. (2016). Genome-Wide Copy Number Variation Scan Identifies Complement Component C4 as Novel Susceptibility Gene for Crohn's Disease. Inflamm. Bowel Dis. *22*, 505–515.
46. Coss, S.L., Zhou, D., Chua, G.T., Aziz, R.A., Hoffman, R.P., Wu, Y.L., Ardoin, S.P., Atkinson, J.P., and Yu, C.Y. (2023). The complement system and human autoimmune diseases. J. Autoimmun. *137*, 102979.
47. Hoffman, G.E., Jaffe, A.E., Gandal, M.J., Collado-Torres, L., Sieberts, S.K., Devlin, B., Geschwind, D.H., Weinberger, D.R., and Roussos, P. (2023). Comment on: What genes are differentially expressed in individuals with schizophrenia? A systematic review. Mol. Psychiatry *28*, 523–525.
48. Hernandez, L.M., Kim, M., Zhang, P., Bethlehem, R.A.I., Hoftman, G., Loughnan, R., Smith, D., Bookheimer, S.Y., Fan, C.C., Bearden, C.E., et al. (2023). Multi-ancestry phenome-wide association of complement component 4 variation with psychiatric and brain phenotypes in youth. Genome Biol. *24*, 42.
49. Schafer, D.P., Lehrman, E.K., Kautzman, A.G., Koyama, R., Mardinly, A.R., Yamasaki, R., Ransohoff, R.M., Greenberg, M.E., Barres, B.A., and Stevens, B. (2012). Microglia sculpt postnatal neural circuits in an activity and complement-dependent manner. Neuron *74*, 691–705.
50. Gallego, J.A., Blanco, E.A., Morell, C., Lencz, T., and Malhotra, A.K. (2021). Complement component C4 levels in the cerebrospinal fluid and plasma of patients with schizophrenia. Neuropsychopharmacology *46*, 1140–1144.
51. GTEx Consortium; Laboratory, Data Analysis &Coordinating Center LDACC—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Enhancing GTEx eGTEx groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site—NDRI (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213. https://www.nature.com/articles/nature24277#supplementary-information.
52. Duchatel, R.J., Meehan, C.L., Harms, L.R., Michie, P.T., Bigland, M.J., Smith, D.W., Jobling, P., Hodgson, D.M., and Tooney, P.A. (2018). Increased complement component 4 (C4) gene expression in the cingulate cortex of rats exposed to late gestation immune activation. Schizophr. Res. *199*, 442–444.
53. Choudhury, Z., and Lennox, B. (2021). Maternal Immune Activation and Schizophrenia-Evidence for an Immune Priming Disorder. Front. Psychiatry *12*, 585742.
54. Zhang, H., Zhang, Y., Yang, F., Li, L., Liu, S., Xu, Z., Wang, J., and Sun, S. (2011). Complement component C4A and apolipoprotein A-I in plasmas as biomarkers of the severe, early-onset preeclampsia. Mol. Biosyst. *7*, 2470–2479.
55. Kessing, L. (1998). Validity of diagnoses and other clinical register data in patients with affective disorder. Eur. Psychiatry. *13*, 392–398.

56. Phung, T.K.T., Andersen, B.B., Høgh, P., Kessing, L.V., Mortensen, P.B., and Waldemar, G. (2007). Validity of dementia diagnoses in the Danish hospital registers. Dement. Geriatr. Cogn. Disord *24*, 220–228.

57. Lauritsen, M.B., Jørgensen, M., Madsen, K.M., Lemcke, S., Toft, S., Grove, J., Schendel, D.E., and Thorsen, P. (2010). Validity of childhood autism in the Danish Psychiatric Central Register: findings from a cohort sample born 1990-1999. J. Autism Dev. Disord. *40*, 139–148.

58. Bock, C., Bukh, J.D., Vinberg, M., Gether, U., and Kessing, L.V. (2009). Validity of the diagnosis of a single depressive episode in a case register. Clin. Pract. Epidemiol. Ment. Health *5*, 4.

59. Mohr-Jensen, C., Vinkel Koch, S., Briciet Lauritsen, M., and Steinhausen, H.C. (2016). The validity and reliability of the diagnosis of hyperkinetic disorders in the Danish Psychiatric Central Research Registry. Eur. Psychiatry. *35*, 16–24.

60. Jakobsen, K.D., Frederiksen, J.N., Hansen, T., Jansson, L.B., Parnas, J., and Werge, T. (2005). Reliability of clinical ICD-10 schizophrenia diagnoses. Nord. J. Psychiatry *59*, 209–212.

61. Musliner, K.L., Liu, X., Gasse, C., Christensen, K.S., Wimberley, T., and Munk-Olsen, T. (2019). Incidence of medically treated depression in Denmark among individuals 15-44 years old: a comprehensive overview based on population registers. Acta Psychiatr. Scand. *139*, 548–557.

62. Weye, N., McGrath, J.J., Lasgaard, M., Momen, N.C., Knudsen, A.K., Musliner, K., and Plana-Ripoll, O. (2023). Agreement between survey- and register-based measures of depression in Denmark. Acta Psychiatr. Scand. *147*, 581–592.

63. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. *42*, 565–569.

64. Grove, J., Ripke, S., Als, T.D., Mattheisen, M., Walters, R.K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O.A., Anney, R., et al. (2019). Identification of common genetic risk variants for autism spectrum disorder. Nat. Genet. *51*, 431–444.

65. Demontis, D., Walters, R.K., Martin, J., Mattheisen, M., Als, T.D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., et al. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat. Genet. *51*, 63–75.

66. Watson, H.J., Yilmaz, Z., Thornton, L.M., Hübel, C., Coleman, J.R.I., Gaspar, H.A., Bryois, J., Hinney, A., Leppä, V.M., Mattheisen, M., et al. (2019). Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. Nat. Genet. *51*, 1207–1214.

67. Marioni, R.E., Harris, S.E., Zhang, Q., McRae, A.F., Hagenaars, S.P., Hill, W.D., Davies, G., Ritchie, C.W., Gale, C.R., Starr, J.M., et al. (2018). GWAS on family history of Alzheimer's disease. Transl. Psychiatry *8*, 99.

68. van Rheenen, W., van der Spek, R.A.A., Bakker, M.K., van Vugt, J.J.F.A., Hop, P.J., Zwamborn, R.A.J., de Klein, N., Westra, H.J., Bakker, O.B., Deelen, P., et al. (2021). Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. Nat. Genet. *53*, 1636–1648.

69. International Multiple Sclerosis Genetics Consortium (2019). Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. Science *365*, eaav7188.

70. de Lange, K.M., Moutsianas, L., Lee, J.C., Lamb, C.A., Luo, Y., Kennedy, N.A., Jostins, L., Rice, D.L., Gutierrez-Achury, J., Ji, S.G., et al. (2017). Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nat. Genet. *49*, 256–261.

71. Chiou, J., Geusz, R.J., Okino, M.L., Han, J.Y., Miller, M., Melton, R., Beebe, E., Benaglio, P., Huang, S., Korgaonkar, K., et al. (2021). Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. Nature *594*, 398–402.

72. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., et al. (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature *506*, 376–381.

73. Julià, A., López-Longo, F.J., Pérez Venegas, J.J., Bonàs-Guarch, S., Olivé, À., Andreu, J.L., Aguirre-Zamorano, M.Á., Vela, P., Nolla, J.M., de la Fuente, J.L.M., et al. (2018). Genome-wide association study meta-analysis identifies five new loci for systemic lupus erythematosus. Arthritis Res. Ther. *20*, 100.

74. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature *562*, 203–209.

75. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. *88*, 76–82.

76. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., and Yang, J. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. *48*, 481–487.

77. Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419.

78. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006.

79. R Core Team (2021). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

80. Zhu, Z., Zheng, Z., Zhang, F., Wu, Y., Trzaskowski, M., Maier, R., Robinson, M.R., McGrath, J.J., Visscher, P.M., Wray, N.R., and Yang, J. (2018). Causal associations between risk factors and common diseases inferred from GWAS summary data. Nat. Commun. *9*, 224.

81. Pedersen, C.B., Bybjerg-Grauholm, J., Pedersen, M.G., Grove, J., Agerbo, E., Bækvad-Hansen, M., Poulsen, J.B., Hansen, C.S., McGrath, J.J., Als, T.D., et al. (2018). The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. Mol. Psychiatry *23*, 6–14.

82. Thornton, L.M., Munn-Chernoff, M.A., Baker, J.H., Juréus, A., Parker, R., Henders, A.K., Larsen, J.T., Petersen, L., Watson, H.J., Yilmaz, Z., et al. (2018). The Anorexia Nervosa Genetics Initiative (ANGI): Overview and methods. Contemp. Clin. Trials *74*, 61–69.

83. Munk-Jørgensen, P., and Mortensen, P.B. (1997). The Danish Psychiatric Central Register. Dan. Med. Bull. *44*, 82–84.

84. Mors, O., Perto, G.P., and Mortensen, P.B. (2011). The Danish Psychiatric Central Research Register. Scand. J. Public Health *39*, 54–57.

85. Borgan, O., Langholz, B., Samuelsen, S.O., Goldstein, L., and Pogoda, J. (2000). Exposure stratified case-cohort designs. Lifetime Data Anal. *6*, 39–58.

86. Nørgaard-Pedersen, B., and Hougaard, D.M. (2007). Storage policies and use of the Danish Newborn Screening Biobank. J. Inherit. Metab. Dis. *30*, 530–536.

87. Hollegaard, M.V., Sørensen, K.M., Petersen, H.K., Arnardottir, M.B., Nørgaard-Pedersen, B., Thorsen, P., and Hougaard, D.M. (2007). Whole genome amplification and genetic analysis after extraction of proteins from dried blood spots. Clin. Chem. *53*, 1161–1162.

88. Bybjerg-Grauholm, J., Bøcker Pedersen, C., Bækvad-Hansen, M., Giørtz Pedersen, M., Adamsen, D., Søholm Hansen, C., Agerbo, E., Grove, J., Als, T.D., Schork, A.J., et al. (2020). The iPSYCH2015 Case-Cohort sample: updated directions for unravelling genetic and environmental architectures of severe mental disorders. Preprint at medRxiv.

89. Thygesen, L.C., Daasnes, C., Thaulow, I., and Brønnum-Hansen, H. (2011). Introduction to Danish (nationwide) registers on health and social

issues: structure, access, legislation, and archiving. Scand. J. Public Health *39*, 12–16.

90. Mortensen, P.B. (2019). Response to "Ethical concerns regarding Danish genetic research". Mol. Psychiatry *24*, 1574–1575.
91. Albiñana, C., Zhu, Z., Borbye-Lorenzen, N., Boelt, S.G., Cohen, A.S., Skogstrand, K., Wray, N.R., Revez, J.A., Privé, F., Petersen, L.V., et al. (2023). Genetic correlates of vitamin D-binding protein and 25-hydroxyvitamin D in neonatal dried blood spots. Nat. Commun. *14*, 852.
92. Gunderson, K.L., Steemers, F.J., Ren, H., Ng, P., Zhou, L., Tsan, C., Chang, W., Bullis, D., Musmacker, J., King, C., et al. (2006). Whole-genome genotyping. Methods Enzymol. *410*, 359–376.
93. Schork, A.J., Won, H., Appadurai, V., Nudel, R., Gandal, M., Delaneau, O., Revsbech Christiansen, M., Hougaard, D.M., Bækved-Hansen, M., Bybjerg-Grauholm, J., et al. (2019). A genome-wide association study of shared risk across psychiatric disorders implicates gene regulation during fetal neurodevelopment. Nat. Neurosci. *22*, 353–361.
94. Lam, M., Awasthi, S., Watson, H.J., Goldstein, J., Panagiotaropoulou, G., Trubetskoy, V., Karlsson, R., Frei, O., Fan, C.C., De Witte, W., et al. (2020). RICOPILI: Rapid Imputation for COnsortias PIpeLIne. Bioinformatics *36*, 930–933.
95. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al. (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nat. Genet. *48*, 1279–1283.
96. Browning, B.L., Zhou, Y., and Browning, S.R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. Am. J. Hum. Genet. *103*, 338–348.
97. Wouters, D., van Schouwenburg, P., van der Horst, A., de Boer, M., Schooneman, D., Kuijpers, T.W., Aarden, L.A., and Hamann, D. (2009). High-throughput analysis of the C4 polymorphism by a combination of MLPA and isotype-specific ELISA's. Mol. Immunol. *46*, 592–600.
98. Jiang, J., and Nguyen, T. (2021). Linear Mixed Models: Part I. In Linear and Generalized Linear Mixed Models and Their Applications (Springer), pp. 1–61.
99. Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Softw. *67*, 1–48.
100. Beasley, T.M., Erickson, S., and Allison, D.B. (2009). Rank-based inverse normal transformations are increasingly used, but are they merited? Behav. Genet. *39*, 580–595.
101. Visscher, P.M. (2009). Whole genome approaches to quantitative genetics. Genetica *136*, 351–358.
102. Yang, C., Farias, F.H.G., Ibanez, L., Suhy, A., Sadler, B., Fernandez, M.V., Wang, F., Bradley, J.L., Eiffert, B., Bahena, J.A., et al. (2021). Genomic atlas of the proteome from brain, CSF and plasma prioritizes proteins implicated in neurological disorders. Nat. Neurosci. *24*, 1302–1312.
103. Gudjonsson, A., Gudmundsdottir, V., Axelsson, G.T., Gudmundsson, E.F., Jonsson, B.G., Launer, L.J., Lamb, J.R., Jennings, L.L., Aspelund, T., Emilsson, V., and Gudnason, V. (2022). A genome-wide association study of serum proteins reveals shared loci with common diseases. Nat. Commun. *13*, 480.
104. Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M., and Price, A.L. (2014). Advantages and pitfalls in the application of mixed-model association methods. Nat. Genet. *46*, 100–106.
105. Galinsky, K.J., Bhatia, G., Loh, P.R., Georgiev, S., Mukherjee, S., Patterson, N.J., and Price, A.L. (2016). Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. Am. J. Hum. Genet. *98*, 456–472.
106. Sidorenko, J., Kassam, I., Kemper, K.E., Zeng, J., Lloyd-Jones, L.R., Montgomery, G.W., Gibson, G., Metspalu, A., Esko, T., Yang, J., et al. (2019). The effect of X-linked dosage compensation on complex trait variation. Nat. Commun. *10*, 3009.
107. Woo, J.J., Pouget, J.G., Zai, C.C., and Kennedy, J.L. (2020). The complement system in schizophrenia: where are we now and what's next? Mol. Psychiatry *25*, 114–130.
108. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. *12*, e1001779.
109. UK10K Consortium; Walter, K., Min, J.L., Huang, J., Crooks, L., Memari, Y., McCarthy, S., Perry, J.R.B., Xu, C., Futema, M., et al. (2015). The UK10K project identifies rare variants in health and disease. Nature *526*, 82–90.
110. Revez, J.A., Lin, T., Qiao, Z., Xue, A., Holtz, Y., Zhu, Z., Zeng, J., Wang, H., Sidorenko, J., Kemper, K.E., et al. (2020). Genome-wide association study identifies 143 loci associated with 25 hydroxyvitamin D concentration. Nat. Commun. *11*, 1647.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Antibodies of C3 and C4	Statens Serum Institut	HYB030-07 (C3), HYB030-06 (C3), HYB162-04 (C4)
Antibodies of C4	Thermo Fisher	RRID AB_923305
Deposited data		
Neonatal C3 protein concentration	This study	https://www.ebi.ac.uk/gwas/ (GWAS Catalog: GCST90281041)
Neonatal C4 protein concentration	This study	https://www.ebi.ac.uk/gwas/ (GWAS Catalog: GCST90281042)
Schizophrenia	Trubetskoy et al.[35]	https://www.med.unc.edu/pgc/download-results/
Major depression	Howard et al.[37]	https://www.med.unc.edu/pgc/download-results/
Bipolar disorder	Mullins et al.[36]	https://www.med.unc.edu/pgc/download-results/
Autism spectrum disorder	Grove et al.[64]	https://www.med.unc.edu/pgc/download-results/
Attention deficit hyperactivity disorder	Demontis et al.[65]	https://www.med.unc.edu/pgc/download-results/
Anorexia nervosa	Watson et al.[66]	https://www.med.unc.edu/pgc/download-results/
Alzheimer's disease	Marioni et al.[67]	https://www.ebi.ac.uk/gwas/studies/GCST005921
Amyotrophic lateral sclerosis	Van Rheenen et al.[68]	https://www.ebi.ac.uk/gwas/studies/GCST90027164
Multiple sclerosis	International Multiple Sclerosis Genetics[69]	https://imsgc.net/
Crohn's disease	de Lange et al.[70]	https://www.ebi.ac.uk/gwas/studies/GCST004132
Ulcerative colitis	de Lange et al.[70]	https://www.ebi.ac.uk/gwas/studies/GCST004133
Type 1 diabetes	Chiou et al.[71]	https://www.ebi.ac.uk/gwas/studies/GCST90014023
Rheumatoid arthritis	Okada et al.[72]	http://plaza.umin.ac.jp/~yokada/datasource/software.htm
Systemic lupus erythematosus	Julia et al.[73]	http://www.urr.cat/
Individual-level data from UK Biobank*	UK Biobank[74]	https://biobank.ndph.ox.ac.uk/showcase/
GTEx version 8	GTEx Consortium[51]	https://gtexportal.org/home/datasets/; https://www.ncbi.nlm.nih.gov/gap/ (dbGaP: phs000424)
The *C4* haplotype imputation protocol[&]	Sekar et al.[9] Kamitaki et al.[10]	https://github.com/freeseek/imputec4/; https://www.ncbi.nlm.nih.gov/gap/ (dbGaP: phs001992)
Software and algorithms		
PLINK2	PLINK Working Group	https://www.cog-genomics.org/plink/2.0/
GCTA	Yang et al.[75]	https://yanglab.westlake.edu.cn/software/gcta/#Overview/
GCTB (BayesR)	Zeng et al.[25] Moser et al.[24]	https://cnsgenomics.com/software/gctb/#Overview/
BOLT-REML	Loh et al.[26]	https://alkesgroup.broadinstitute.org/BOLT-LMM/BOLT-LMM_manual.html
FUMA/MAGMA v1.50	Watanabe et al.[34]	https://fuma.ctglab.nl/
SMR	Zhu et al.[76]	https://yanglab.westlake.edu.cn/software/smr/
Human Protein Atlas	Uhlen et al.[77]	https://www.proteinatlas.org/
UCSC Genome Browser	Kent et al.[78]	https://genome.ucsc.edu/
R 4.0.5	R core team[79]	https://www.R-project.org/
GSMR 1.09	Zhu et al.[80]	https://yanglab.westlake.edu.cn/software/gsmr/

* The UK Biobank data is an individual-level data. The remaining datasets are GWAS summary statistics that are publicly available.
[&] *C4* haplotype imputation accessed via dbGAP: phs001992.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the corresponding author Prof. John McGrath (j.mcgrath@uq.edu.au). The information and requests for resources can be also fulfilled by the first author Dr Zhihong Zhu, email: z.zhu.ncrr@au.dk.

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The summary statistics from the GWAS for C3 and C4 will be made available via the GWAS Catalog https://www.ebi.ac.uk/gwas/ (Study accession numbers, GWAS of C3 protein concentration, GWAS Catalog: GCST90281041, GWAS of C4 protein concentration, GWAS Catalog: GCST90281042). All other GWAS summary data are publicly available and listed in the key resources table.
- Owing to the sensitive nature of these data, individual level data can be accessed only through secure servers where download of individual level information is prohibited. Each scientific project must be approved before initiation, and approval is granted to a specific Danish research institution. International researchers may gain data access through collaboration with a Danish research institution. More information about getting access to the iPSYCH data can be obtained at https://ipsych.dk/en.
- This study did not generate any unique datasets or code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

The iPSYCH2012 study

Key elements of this study were based on the Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH) sample,[81] a population-based case-cohort designed to study the genetic and environmental factors of SCZ, BIP, DEP, ASD and ADHD. The original iPSYCH sample (known as iPSYCH2012) included information on case status complete through 31 December 2012. We also included 4,791 AN cases from the Anorexia Nervosa Genetics Initiative (ANGI-DK),[82] which had the same design as iPSYCH2012. Henceforth, we refer to iPSYCH2012 as the combined dataset with the ANGI samples. The iPSYCH2012 sample is nested within the entire Danish population born between 1981 and 2005 (N=1,472,762). Diagnoses were identified in the Danish Central Psychiatric Research Register,[83,84] which includes all inpatient contacts in Danish psychiatric hospitals since 1969 and all outpatient and emergency contacts since 1995. The ICD-10 codes used to classify the psychiatric disorder cases can be found in Table S1. The phenotype information for the iPSYCH2012 participants was updated for the target mental disorders until December 2016. The case-cohort sample includes a population-based random sub-cohort[85] (n = 30,000) with an inclusion probability of 2.04% of the study base (30,000 / 1,472,762). This sub-cohort also includes some participants with the target mental disorders of interest. The genotypes and C3 and C4 protein concentrations were measured in neonatal dried bloodspots (DBSs) taken as part of routine screening at birth from all babies born in Denmark since 1981 and stored in the Danish Neonatal Screening Biobank.[86] Dried blood spot samples have been collected from practically all neonates born in Denmark since 1st May 1981 and stored at −20°C. Samples are collected 4–7 days after birth. After the dried blood spots were retrieved from the biobank, samples were extracted in a PBS buffer and stored for further use at -80°C. Subsequently, DNA was extracted according to previously published methods.[87] After storage the protein extracts were assayed for C3 and C4 concentrations. Thus, all genotypes and C3/C4 protein concentration data originated from a single DBS extraction. Additional details related to blood spot extraction and storage are provided in Method S1. As a post-hoc analysis, we had access to additional genotyped samples from the iPSYCH2015 extension study[88] (n ∼56,000 samples,[88] Method S2). We were not able to assess C3 or C4 protein concentration in these samples.

Ethical framework

Material from the Danish Neonatal Screening Biobank has been used primarily for screening for congenital disorders, but are also stored for follow-up diagnostics, screening, quality control and research. According to Danish legislation, material from The Danish Neonatal Screening Biobank can be used for research after approval from the Biobank, and the relevant Scientific Ethical Committee.[89,90] There is also a mechanism in place ensuring that one can opt out of having the stored material used for research. The Danish Data Protection Agency and the Danish Health Data Authority approved this study. According to Danish law, informed consent is not required for register-based studies. All data accessed were deidentified.

C3 and C4 protein concentrations

These methods have been described in a related study.[91] Two 3.2 mm discs of DBS were punched into 96 well polymerase chain reaction plates (72.1981.202, Sarstedt). The extracts were analyzed with a multiplex immunoassay (also measuring vitamin D binding protein[91]) using U-plex plates (Meso-Scale Diagnostics (MSD), Maryland, US) employing antibodies specific for complement C3 (HYB030-07 and HYB030-06) and complement C4 (MA1-72520 (ThermoFisher Scientific) and HYB162-04). The antibodies were purchased from SSI Antibodies (Copenhagen, Denmark) except if otherwise stated. Extracts were analyzed diluted 1:70 in diluent 101 (#R51AD, MSD). Capture antibodies (used at 10 μg/mL as input concentration) were biotinylated in-house using EZ-Link Sulfo-NHS-LC-Biotin (#21327, Thermo Fisher Scientific) and detection antibodies were SULFO-tagged (#R91AO, MSD), both at a challenge ratio of 20:1. As calibrators, we used complement components purified from human: C3: #PSP-109 (Nordic Biosite, Copenhagen, DK), C4: abx060108 (Abbexa, Cambridge, UK). Calibrators were diluted in diluent 101, detection antibodies (used at 1 μg/mL) were diluted in diluent 3 (#R50AP, MSD). Controls were made in-house from part of the calibrator solution in one batch, aliquoted

in portions for each plate, and stored at -20°C until use. The samples were prepared on the plates as recommended by the manufacturer and were read on the QuickPlex SQ 120 (MSD) 4 min after adding 2x Read buffer T (#R92TC, MSD). Analyte concentrations were calculated from the calibrator curves on each plate using 4PL logistic regression using the MSD Workbench software.

Intra-assay variations were calculated from 38 measurements analyzed on the same plate of a pool of extract made from 304 samples. Inter-assay variations were calculated from controls analyzed in duplicate on each plate during the sample analysis, 1022 plates in total. Lower limits of detections were calculated as 2.5 standard deviations from 40 replicate measurements of the zero calibrator. The higher detection limit was defined as the highest calibrator concentration. The lower and upper detection limits for: (a) C3 were 95.4 μg/L and 79.8 mg/L respectively, and (b) C4 were 55.2μg/L and 79.8 mg/L respectively. The intra- and inter-assay coefficient of variation (CV) for (a) C3 were 5.2% and 18.1% respectively; and for (b) C4 were 3.9% and 8.5% respectively. To validate the stability of the samples during storage, we randomly selected 15-16 samples from five years (1984, 1992, 2000, 2008, and 2016; a total of 76 samples). After extracting the samples and adding them to an MSD plate, the rest of the extracts were frozen for 2 months, thawed and measured as described above to imitate the freeze-thaw cycle of the samples in the study. The oldest samples (from 1984) recorded lower concentrations, most probably due to a change in the type of filter paper after 1989. In light of this artifact, we adjusted all values by plate (the sequence of testing followed the date of birth of the sample). This is described below. Additional details related to pre-analytic variation are provided in Method S1.

Imputation of genotypes

DNA genotyping was conducted at the Broad Institute (Boston, MA, USA) using the Infinium PsychChip v1.0 array (Illumina, San Diego, CA, USA).[92] We restricted the genotyped SNPs to 252,339 high-quality and common SNPs based on build hg19 (the same human genome reference build was used throughout this study). Details of the filtering can be found elsewhere.[93] Briefly, we excluded SNPs with minor allele frequency (MAF) < 0.01, Hardy Weinberg Equilibrium (HWE) p-value < 1.0×10^{-6} or non-SNP alleles (i.e., insertions and deletions, INDELs). 245,328 autosomal and 7,011 X-chromosome (chrX) SNPs were retained and used to impute SNPs using the Ricopili pipeline[94] with the Haplotype Reference Consortium (HRC)[95] as the imputation reference panel (accession number: EGAD00001002729). 6,743,499 autosomal SNPs, 227,371 chrX SNPs for males and 184,517 chrX SNPs for females were retained with missing rate < 0.02 and genotype call probability > 0.8. We further excluded the imputed SNPs with imputation info score < 0.8, MAF < 0.01 or HWE p-value < 1.0×10^{-6}. 5,201,724 SNPs were retained in autosomes and 126,109 SNPs were retained on chrX. We then used the common SNPs to infer the genetic ancestries of 80,873 participants in the iPSYCH2012 study, 75,764 individuals of European ancestry and 5,109 individuals of non-European ancestry. With respect to the individuals of non-European ancestry, we identified 159 individuals of African ancestry and 101 individuals of South Asian ancestry. Details are provided in Method S3.

Imputation of *C4* haplotypes

C4 haplotypes were imputed from reference data[9,10] (the database of Genotypes and Phenotypes [dbGaP]: phs001992) using the genotyped SNPs in the iPSYCH2012 sample. The human *C4* haplotypes have various copy numbers, including two isotypic polymorphisms, *C4A* (*A*) and *C4B* (*B*). Each isotype has two length-polymorphisms due to a human endogenous retroviral (*HERV*) insertion, long form (*L*, with *HERV* insertion) and short form (*S*, without *HERV* insertion). The isotypic and length polymorphisms lead to four alleles in a *C4* copy, *AL*, *AS*, *BL* and *BS*. Using the genotyped SNPs, the *C4* haplotype reference was used to impute the C4 alleles and the number of *C4* copies (with a maximum copy number of 4). The *C4* haplotype imputation panel comprised whole genome sequencing data from 1,265 individuals of multiple ancestries, which enabled us to identify *C4* alleles with high accuracy. We used Beagle software[96] for the imputation with the *C4* haplotype reference. The imputation results provided the counts of alleles, but were unable to confidently distinguish all combinations of variants, for example, between the haplotypes *AS-BL* and *AL-BS*. We counted the two *C4* alleles (*C4A* and *C4B*) with combination of *HERV* using a subset of the imputed result, where combinations can be confidently distinguished (details are provided in Method S4). Both counts of *C4* allele combinations and reported studies[97] indicated that the *C4A* gene is more likely to carry *HERV* insertion than the *C4B* gene. Therefore, the *C4* haplotype is assumed to be *AL-BS* rather than *AS-BL*, consistent with methods described by Sekar et al.[9] The imputed counts were converted to the *C4* haplotypes. Eight common *C4* haplotypes (allele frequencies $\geq$ 0.01) were imputed in the iPSYCH2012 study (Table S3). The allele frequencies of the 8 haplotypes were consistent with other studies.[10,21] We counted the copy numbers of the *C4* alleles (Figure S2) for each participant. 28 individuals (0.04%) carried 4 copies of *C4B* and 35 individuals (0.05%) carried 6 copies of *HERV* insertion. Therefore, we excluded these individuals with very rare copy numbers. The *C4A* copy number is strongly correlated with *C4B* and *HERV* copy numbers (Pearson correlation between *C4A* and *C4B* = -0.52; between *C4A* and *HERV* = 0.73). We imputed the *C4* haplotypes and the *C4* copy numbers in the iPSYCH2015 extension study using the same method. The copy numbers of *C4* alleles were counted from the imputed haplotypes. Details are provided in the Method S5. Since the C4 haplotype imputation reference data included individuals of multiple ancestries, we applied the same method that was used in the European cohort to impute *C4* haplotypes in the 159 individuals of African ancestry and the 101 individuals of South Asian ancestry.

Quality control of the C3 and C4 protein concentrations

The C3 and C4 protein concentrations were measured in 78,268 iPSYCH2012 participants of multiple ancestries. We focused on 68,768 individuals of European ancestry with measures of C3 and C4 protein concentrations. The protein assay plates captured a

substantial amount of variance (C3 = 49.4%, C4 = 45.3%). Therefore, we used a linear mixed model (LMM)[98] approach to adjust protein concentrations, $\mathbf{y} = \mathbf{Z}_{plate}\mathbf{u}_{plate} + \mathbf{e}$, where $\mathbf{y}$ represents the C3/4 protein concentration; $\mathbf{Z}_{plate}$ represents protein assay plate, a random variable; $\mathbf{u}_{plate}$ represents the random effect of protein assay plate; and $\mathbf{e}$ represents residual. The mixed model regression was conducted by the R package of lme4.[99] The rank-based inverse normal transformation (RINT)[100] was applied to the residuals to have mean 0 and variance 1. The standard deviations (SDs) adjusted for variance captured by protein assay plate were used for the interpretation of results of C3 and C4 protein concentrations, for C3 protein concentration, 1 SD unit = 2.56 mg/L (3.60 mg/L × $\sqrt{(1 - 0.49)}$), and for C4 protein concentration, 1 SD unit = 2.46 mg/L (3.33 mg/L × $\sqrt{(1 - 0.45)}$). We then performed quality control analysis in 150 individuals of African ancestry and 94 individuals of South Asian ancestry. These individuals of non-European ancestries had the measures of both C3 and C4 protein concentrations. Due to the small sample sizes, nearly all these neonatal blood samples were separately measured on different protein assay plates. We were unable to use the LMM approach to adjust protein concentration for protein assay plate. Therefore, we used a linear regression model where the protein assay plate was a fixed variable. After adjustment, we applied RINT to standardize the residuals with mean 0 and variance 1.

Heritability and SNP-based heritability of the C3 and C4 protein concentrations

The iPSYCH2012 cohort had 75,764 participants of European ancestry. 19,113 participants who shared a genetic relatedness (entry of genetic relationship matrix (GRM), r_{GRM}) $\geq$ 0.05 with at least one other individual were considered as relatives; 3,253 first degree ($r_{GRM} \geq 0.4$), 2,077 second degree relatives ($0.2 \leq r_{GRM} < 0.4$) and 13,783 third degree relatives ($0.05 \leq r_{GRM} < 0.2$). The cut-offs of relatives reflect their the respective expectations (1/2, 1/4, and 1/8 for first, second, and third degree relatives,[101] respectively), because the pair-wise GRM-based coefficients between individuals were estimated from observed genetic variants, which resulted in minor variations in the kinship coefficients. We jointly estimated both the heritability (h^2) and the SNP-based h^2 (h^2_{SNP}) of the C3 and C4 protein concentrations by using the method proposed by Zaitlen et al.[23] The method requires two GRMs, 1) the full GRM and 2) the GRM with all entries below a threshold t (t = 0.05 in the study) set to 0. The first variance component provides the estimate of h^2_{SNP} and the sum of two variance components provides the estimate of h^2. This method assumes a normal distribution of SNP effect sizes. The GWAS studies of protein concentrations[31,102,103] observed that *cis*-pQTLs (significant SNPs in or near the coding genes) often capture more phenotypic variance than the remaining SNPs (including *trans*-pQTLs). Therefore, we used two approaches to further explore the h^2_{SNP} using genetically unrelated participants (no GRM entries > 0.05) (Data S1); 1) estimating it using all common SNPs by BayesR,[24] and 2) partitioning h^2_{SNP} into (a) $h^2_{cis\text{-}chr}$, explained by SNPs on the chromosome where the coding gene (*cis*-chr SNPs) was positioned, and (b) $h^2_{trans\text{-}chr}$, explained by the remaining SNPs (*trans*-chr SNPs). This analysis was conducted by GREML.[63] The genetic relationship matrix used in the Zaitlen and GREML analyses were estimated from 5,201,724 common SNPs. Only the subset of HapMap phase 3 (HM3) SNPs were included in the BayesR analyses because of the computation complexity (853,129 HM3 SNP in total. The Zaitlen method and GREML were implemented in Genome-wide Complex Trait Analysis (GCTA).[75] BayesR was implemented in Genome-wide Complex Trait Bayesian analysis[25] (GCTB). The URLs for these programs are provided below.

We estimated the genetic correlation between C3 and C4 concentrations by BOLT-REML.[26] To further examine if the genetic correlation was primarily driven by the two protein-coding genes (i.e. *C3* and *C4*), we conducted the BOLT-REML and Haseman-Elston regression[27] (implemented in GCTA) analyses using the *trans*-chr SNPs.

We conducted post-hoc sex-specific analyses of SNP-based h^2 and between-sex genetic correlation in the 28,750 males and 22,436 females of European ancestry in iPSYCH2012 (Data S2). These individuals who had the measures of C3 and C4 protein concentrations were genetically unrelated.

GWAS of C3 and C4 protein concentrations

We performed the GWAS analysis of the C3 and C4 protein concentrations by fastGWA.[28] The fastGWA is a LMM method which can include all individuals of European ancestry regardless of relatedness. 5,201,724 imputed SNPs were analyzed in the GWAS. Since all variants are included as random variables, fastGWA loses power for identification of candidate markers especially when those particular *cis*-pQTLs capture a large proportion of the total variance[104]— in the study, we defined *cis*-pQTLs as significant SNPs within ±10Mb of the respective coding genes. Therefore, we excluded the SNPs in and near the coding gene for the required GRM in the GWAS, C3: chr19, 4.67Mb – 8.74Mb, C4: chr6, 24.8Mb – 33.9Mb. For each GWAS, we fitted birthyear, sex, wave (i.e., genotyping batch) and the first 20 PCs as covariates in the model. The PCs were estimated by FastPCA,[105] excluding the same SNPs as we did for the required GRM. We conducted the GWASs using all SNPs on autosomal and sex chromosomes. SNPs on the X chromosome for males (coded as 0/2) were tested as diploid, assuming X chromosome of males has half dosage compensation.[106] We used GCTA-COJO[75] to identify the SNPs which were independently associated with the two concentrations. We randomly sampled 10,000 participants from the population-based sub-cohort of iPSYCH2012 as the LD reference cohort. The GWAS significance threshold was 5.0×10^{-8}. To increase power for the identification of *trans*-pQTLs (which we defined as significant SNPs other than chromosome 6 because of long-range LD in MHC region), we used fastGWA to perform the GWAS analysis of the C3 and C4 protein concentration adjusted for the respective *cis*-pQTLs. In the model, the GRMs were the same as above. The covariates included those we used in the unadjusted GWAS and the independent *cis*-pQTLs from the COJO analyses. The GWAS significance threshold remained 5.0×10^{-8}. We used the unadjusted GWAS of C3 and C4 protein concentrations in the discovery of post-GWAS analyses. The adjusted GWASs were only used in GSMR as planned sensitivity analyses.

We performed post-hoc analyses, with sex-specific GWAS analyses of C3 and C4 protein concentrations, to explore potential differences in effect sizes of SNPs on protein concentration between males and females. The GWAS analyses were conducted by fastGWA. 32,361 males were included in the male-specific GWAS analysis and 36,407 females were included in the female-specific GWAS analysis. All the individuals were of European ancestry. The required male and female GRMs were subsets from the full matrix. We fitted birthyear and the first 20 PCs as covariates in the model. All the SNPs from the unadjusted GWAS were included in the sex-specific GWAS analyses. We then examined differences in effect sizes between males and females. The test statistic was estimated from $T_{difference} = (b_{male} - b_{female})^2 / [var(b_{male}) + var(b_{female})]$. All the statistics used in the formula (b_{male}, $var(b_{male})$, b_{female} and $var(b_{female})$) were from the sex-specific GWAS analyses. We used the genome-wide significant threshold (5.0×10^{-8}) to identify SNPs that have different effect sizes between the two sexes.

To explore if the enrichment of mental disorder cases in the iPSYCH2012 case-cohort could induce bias within the GWASs, we conducted simulations with ascertained individuals and performed GWASs in the population-based sub-cohort (Method S6).

Associations between C4 haplotypes and protein concentrations

We examined the associations between the imputed *C4* haplotypes and the two observed C3 and C4 protein concentrations. We first examined the associations of *C4* copy numbers using a LMM approach, in matrix form, $\mathbf{y}_{protein} = \mathbf{x}_{copy}b_{copy} + \mathbf{X}_c\mathbf{b}_c + \mathbf{Z}_{-MHC}\mathbf{u}_{-MHC} + \mathbf{e}$, where $\mathbf{y}_{protein}$ was C3/4 protein concentration; $\mathbf{x}_{copy}$ was copy number of *C4* allele, either *C4A*, *C4B* or *HERV*; b_{copy} represents effect of copy number; $\mathbf{X}_c$ was covariate with $\mathbf{b}_c$ being its effect; Both b_{copy} and $\mathbf{b}_c$ were fixed effects. The covariates in the model were the same as those fitted in the GWAS of C4 protein concentration. We fitted the SNPs outside the MHC region ($\mathbf{Z}_{-MHC}$) in the model with $\mathbf{U}_{-MHC}$ being their random effects. Fitting SNPs in the MHC region is likely to underestimate the effects of *C4* allele count due to multi-collinearity. Therefore, these SNPs were excluded from the model. In practice, the effect of copy number from the linear mixed model could be estimated by generalized least squares (GLS) method, $\mathbf{b} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}_{protein}$, where $\mathbf{X} = \{\mathbf{x}_{copy}, \mathbf{X}_c\}$ and $\mathbf{V}$ was phenotypic covariance matrix of C3/4 protein concentration. It was implemented by GCTA-GREML. All the individuals of European ancestry were included in the analysis. The three *C4* allele counts were correlated. Therefore, we estimated the joint effects using the same LMM approach as above, in matrix form, $\mathbf{y}_{protein} = \mathbf{x}_{C4A}b_{C4A} + \mathbf{x}_{C4B}b_{C4B} + \mathbf{x}_{HERV}b_{HERV} + \mathbf{X}_c\mathbf{b}_c + \mathbf{Z}_{-MHC}\mathbf{u}_{-MHC} + \mathbf{e}$. In the model, $\mathbf{x}_{C4A}$, $\mathbf{x}_{C4B}$ and $\mathbf{x}_{HERV}$ represent respectively counts of *C4A*, *C4B* and *HERV*. The remaining variables were defined as above. Secondly, we further examined the associations of the imputed *C4* haplotypes using the LMM approach. Previous studies have reported a strong effect for the *C4A* gene,[9] while the effect of *C4B* remains unclear.[107] Therefore, we used '*BS*' as the reference haplotype to estimate joint effects of the remaining haplotypes. The regression model can be expressed as, in matrix form, $\mathbf{y}_{protein} = \mathbf{X}_{allele}\mathbf{b}_{allele} + \mathbf{X}_c\mathbf{b}_c + \mathbf{Z}_{-MHC}\mathbf{u}_{-MHC} + \mathbf{e}$, where $\mathbf{X}_{allele}$ represents *C4* haplotype. Seven *C4* haplotypes were included in the model, except for BS. The remaining parameters were defined as above. The estimated effect can be interpreted as the effect of the *C4* haplotype compared to BS. All the European participants in the iPSYCH2012 study were included in the analysis. The significance threshold for these analyses was the same as the main GWAS significance threshold (i.e., 5.0×10^{-8}).

To explore the associations in non-European ancestry cohorts, we examined the joint effect sizes of three copy numbers (i.e., *C4A*, *C4B* and *HERV*) on C4 protein concentration. The effect sizes were estimated using an LMM approach which was the same as we used in the individuals of European ancestry. All the three copy numbers were jointly fitted in the model. In practice, the effect sizes were estimated by GCTA-GREML and the GRMs were estimated in the respective ancestry cohorts.

FUMA/MAGMA and SMR

We conducted gene-based analysis by Multi-marker Analysis of GenoMic Annotation[33] nested in Functional Mapping and Annotation of Genome-Wide Association Studies[34] (FUMA/MAGMA, v1.5.0). The gene mapping was conducted based on positions of SNPs and genes using default parameters. There were 18,305 genes available for the gene-based analysis, thus the Bonferroni corrected threshold was 1.4×10^{-6} (=0.05/(18,305 × 2)). We conducted Summary-data-based Mendelian Randomization (SMR)[76] to identify genes for C3 and C4 concentrations based on associations between estimated gene expressions and protein concentrations. The SMR analysis requires summary-level statistics, including eQTL data (i.e., summary statistics from association analysis of gene expressions) and GWAS of neonatal C3/4 protein concentration. In our study, the eQTL data was Genotype-Tissue Expression version 8 (GTEx v8).[51] The LD reference sample with 10,000 participants was the same as for the GCTA-COJO analysis. 22,338 gene-tagged probes within 49 tissues (200,144 probes in total) which had significant SNPs were included in the SMR analysis. The Bonferroni significance threshold was 1.2×10^{-6} (= 0.05 / (200,144 × 2)). The threshold of HEIDI to filter association caused by LD between underlying causal variants was 0.01.

To further examine the association between brain *C4A/B* gene expression and circulating C4 protein concentration, we additionally conducted three analyses, 1) the correlation between effect sizes of SNPs on brain *C4A/B* gene expression and effect sizes on C4 protein concentration, 2) the correlation between predicted C4 protein concentration in GTEx and brain *C4A/B* gene expression, 3) the correlation between predicted *C4A/B* gene expression in iPSYCH2012 and circulating C4 protein concentration. In the first analysis, we used the effect sizes of SNPs on brain *C4A/B* gene expression in 15 brain-related tissues from GTEx v8. All provided GTEx *cis*-eQTLs (dbGaP: phs000424) were used in this analysis. The effect sizes of SNPs on the neonatal circulating C4 protein concentration were provided from the GWAS of C4 in the study. In the second analysis, we used BayesR[24,25] to predict the polygenic score (PGS) of C4 protein concentration in the GTEx data (dbGaP: phs000424). We examined the correlations with the *C4A/B* gene expression in 15 brain-related tissues. Details of the sample sizes were provided in Table S14. In the third analysis, we predicted the brain

C4A/B mRNA expression in the iPSYCH2012 study using the formula provided from the Sekar et al. study.[9] We included 50,881 unrelated individuals in examining the Pearson correlation coefficient between predicted *C4A/B* expression and standardized C4 protein concentration.

Associations between C4 haplotypes and mental disorders observed within the iPSYCH case-cohort study

Based on the associations with protein concentrations, we conducted the associations between *C4* haplotypes and 6 iPSYCH disorders (SCZ, BIP, DEP, ASD, ADHD and AN). We used three approaches to examine the relationships, 1) associations with *C4* allele counts, 2) associations with imputed *C4* haplotypes, 3) associations with predicted *C4* gene expression in the brain. Because the iPSYCH case-cohort study has person-level data on the age-at-first contact with psychiatric services, we were able to assess the risk of mental disorders within the time-to-event framework, using Cox proportional hazards regression (Cox PH) to analyze the hazards of *C4* allele counts and haplotypes with respect to the mental disorder of interest. For *C4* allele count, we examined the joint effects due to their correlations, $h(t) = h_0(t)\exp(\mathbf{x}_{C4A}b_{C4A} + \mathbf{x}_{C4B}b_{C4B} + \mathbf{x}_{HERV}b_{HERV} + \mathbf{X}_c\mathbf{b}_c)$. In the model, $h_0(t)$ represents the baseline hazard while $h(t)$ represents the hazard at time t between baseline and December 2016. The remaining variables were defined as above. For *C4* haplotypes, we examined the joint effects using the Cox PH model, $h(t) = h_0(t)\exp(\mathbf{X}_{allele}\mathbf{b}_{allele} + \mathbf{X}_c\mathbf{b}_c)$. All the variables were defined as above. In addition to *C4* haplotypes, we used the predicted *C4A* and *C4B* gene expressions as outlined in the postmortem brain study of Sekar et al.[9] The association was conducted with a Cox PH model, $h(t) = h_0(t)\exp(\mathbf{x}_{C4A\_predicted}b_{C4A\_predicted} + \mathbf{x}_{C4B\_predicted}\,b_{C4B\_predicted} + \mathbf{X}_c\mathbf{b}_c)$, where $\mathbf{x}_{C4A\_predicted}$ and $\mathbf{x}_{C4B\_predicted}$ represent the predicted *C4A* and *C4B* gene expressions, respectively. We first conducted the three analyses in the iPSYCH2012 study, and subsequently conducted additional post-hoc analyses based on the expanded iPSYCH2015 study. We included only unrelated individuals of European ancestry in all the analyses. In the time-to-event analysis, the cases were the diagnosed participants by December 2016, and the non-cases are defined as the entire cohort excluding those individuals with the disorder of interest. Therefore, we defined six psychiatric-disorder samples for the time-to-event analyses. The sample sizes for cases and non-cases are shown in Table S1.

To optimize the power of the sex-specific associations between C4-related genotypes and mental disorders in males and females, we conducted the time-to-event analysis in the iPSYCH2015 extension study. We used a joint model in the sex-specific analysis, which fitted all the three copy numbers, *C4A*, *C4B* and *HERV*. This model was the same as we used in the iPSYCH2012 cohort. The covariates fitted in the model were subsets from those we used in the primary analysis. The male and female individuals of European ancestry included in the analysis were genetically unrelated.

Associations between protein concentrations and mental disorders observed within the iPSYCH2012 case-cohort study

Based on the associations between *C4* haplotypes and 1) the two protein concentrations (C3 and C4) and 2) six mental disorders, we explored the associations between C3 and C4 protein concentrations and mental disorders observed in the iPSYCH2012 case-cohort study, using Cox PH models. Due to the high correlation, we fitted both concentrations jointly, $h(t) = h_0(t)\exp(\mathbf{x}_{C3\_protein}b_{C3\_protein} + \mathbf{x}_{C4\_protein}b_{C4\_protein} + \mathbf{X}_c\mathbf{b}_c)$, where $\mathbf{x}_{C3\_protein}$ and $\mathbf{x}_{C4\_protein}$ represent C3 and C4 concentration, respectively. The effect sizes of two protein concentrations, $b_{C3\_protein}$ and $b_{C4\_protein}$, were fixed effects. The remaining variables were defined as above. In the three analyses, we included only unrelated individuals of European ancestry in the iPSYCH2012 study. In the sex-specific association post-hoc analysis between C3/4 protein concentration and mental disorders, we applied the same method as we used in the full iPSYCH2012 study. All the variables in the sex-specific model were subsets from the primary analysis.

Mendelian Randomization analysis based on summary statistics

We explored the relationships between protein concentrations and mental and autoimmune disorders using the generalized summary-data-based Mendelian Randomization (GSMR) method.[80] Because of the possible link between C3 and C4 versus brain function,[7] we also included two neurodegenerative disorders—Alzheimer's disease, and amyotrophic lateral sclerosis in these analyses. Thus, there were 8 broadly-defined neuropsychiatric disorders (i.e., SCZ,[35] DEP,[37] BIP,[36] ASD,[64] ADHD,[65] AN,[66] Alzheimer's disease,[67] and amyotrophic lateral sclerosis[68]), and 6 autoimmune disorders (i.e., MS,[69] T1D,[71] CD[73], UC[73], RhA,[72] and SLE[73]). The GWAS summary statistics for these disorders were publicly available (additional details provided in Table S22). Unfortunately, detailed GWAS summary statistics for Sjögren's syndrome were not available. The GSMR method was implemented in GCTA.

We used the GCTA-GSMR default settings to select SNPs from the unadjusted GWAS (i.e., independent SNPs from LD clumping, P-value $< 5.0\times10^{-8}$, LD $r^2 < 0.05$, LD clumping widow size = 10Mb). The GSMR method includes options to exclude potentially pleotropic SNPs (via the HEIDI-outlier method). The HEIDI-outlier identifies SNPs whose individual b_{zy}/b_{zx} deviates significantly from other SNPs (assumed to be valid instruments). In theory, these pleiotropic loci may be associated with both the exposure (e.g., C4 protein concentration) and the outcome (e.g., mental disorders) via two independent pathways (i.e., 'horizontal pleiotropy'). In the study, we set HEIDI-outlier threshold at 0.01 to filter horizontal pleiotropy.

We then conducted reverse GSMR using the default settings. The comparison of the results from forward GSMR (e.g. C3/4 protein concentration → mental and autoimmune disorders) and reverse GSMR (e.g., mental and autoimmune disorders → C3/4 protein concentration) can help identify the presence of causality or the presence of pleiotropy. With respect to the forward GSMR (with disease outcomes), we reported the log odds ratio (logOR) and 95% confidence intervals (CI). With respect to the reverse GSMR (with continuous measures of protein concentrations), we report beta and SE. However, we only used the reverse GSMR results

to examine the presence of reverse causation and pleiotropy. Therefore, we reported the GSMR estimates directly. As planned sensitivity analyses, we repeated the GSMR analysis using the adjusted C4 GWAS summary statistics.

The LD reference sample required in GSMR included 10,000 participants, the same as we used in the GCTA-COJO analysis. The GSMR Bonferroni corrected significance threshold was 1.9×10^{-3} (= 0.05 / (2 × 13)).

PheWAS based on UK Biobank phenotypes

Based on the GSMR analysis results, we conducted phenome-wide association studies (PheWASs) to explore the relationships with disorder outcomes in the UK Biobank (UKB) cohort,[108] a large population cohort with 487,409 participants of multiple ancestries. The PheWAS analyses were regressions of measured phenotypes on the PGS of C3 or C4 protein concentration. The PGS of protein concentration was predicted by BayesR using SNPs across the whole genome, including those in MHC. This method conducts Bayesian posterior inference on effects of SNPs, where effects of null SNPs are shrunk toward zero. The genotypes were imputed to the HRC[95] and UK10K[109] reference panels by the UKB group. The quality controls were described in detail elsewhere,[110] including genetic ancestry determination, quality controls of imputed SNPs, and estimation of principal components. In the study, we included 1,130,559 HM3 SNPs on autosomal chromosomes, with MAF $\geq$ 0.01, HWE P-value $\geq$ 1.0×10^{-6}, because only effects of HM3 SNPs were predicted by BayesR. The genetic relationship matrix was estimated by GCTA. 347,769 unrelated participants of European ancestry were retained with genetic relationship < 0.05. In the PheWAS analysis, we included 1,148 UKB phenotypes, 1) 1,027 disorders which were classified by ICD-10 codes, 2) 51 anthropometric measurements and brain imaging traits, and 3) 70 infectious disease antigens. The quantitative traits were standardized by RINT to have mean 0 and variance 1. We then used the model to test the associations, for quantitative traits, $\mathbf{y} = \mathbf{x}_{protein\_prs}b_{protein\_prs} + \mathbf{X}_c\mathbf{b}_c + \mathbf{e}$, where y represents quantitative trait in UKB; $\mathbf{x}_{protein\_prs}$ represents polygenic scores for neonatal C3/4 protein concentration predicted by BayesR; x_c represent the covariate variables including birth year, sex and 20 PCs. For dichotomous traits, $\text{logit}(\mathbf{y}) = \mathbf{x}_{protein\_prs}b_{protein\_prs} + \mathbf{X}_c\mathbf{b}_c + \mathbf{e}$, where $\mathbf{y}$ represents the dichotomous trait and definitions of the remaining variables were the same as above. In addition, we conducted the PheWAS analyses for males and females separately using the same approach. Polygenic scores were predicted using GWASs in both sexes. The Bonferroni corrected significance threshold was 7.3×10^{-6} (= 0.05 / (1148 × 3 × 2)).