Title: Genetic correlations among brain-behavioral and immune-related phenotypes based on genome-wide association data.


1 Psychiatric Genetic Epidemiology & Neurobiology Laboratory (PsychGENe Lab); Departments of Psychiatry and Behavioral Sciences & Neuroscience and Physiology; SUNY Upstate Medical University; Syracuse, NY, U.S.A.

2 Institute for Stroke and Dementia Research, Klinikum der Universität München, Ludwig-Maximilians-University (LMU), Munich, Germany.

3 Departments of Neurology and Public Health Sciences, University of Virginia School of Medicine, Charlottesville, VA, U.S.A.

4 Department of Neurology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, U.S.A.

5 Department of Infectious Diseases and Microbiology, Graduate School of Public Health, University of Pittsburgh, PA, U.S.A.

6 Department of Medicine, University of Cambridge, Cambridge, U.K.

7 Genetics Institute, University College London, London, WC1E 6BT, U.K.

8 Institute of Clinical Molecular Biology, Kiel University and University Hospital Schleswig-Holstein, Kiel, Germany.

9 Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, U.S.A.

10 BHF Centre for Cardiovascular Science, Queen’s Medical Research Institute, University of Edinburgh, Edinburgh, EH16 4TJ, U.K.

11 School of Social and Community Medicine, MRC Integrated Epidemiology Unit, University of Bristol, Bristol, BS8 2BN, UK.

12 Center for Applied Genomics, Division of Human Genetics, The Children’s Hospital of Philadelphia, Philadelphia, PA, U.S.A.

13 Division of Endocrinology and Diabetes, The Children’s Hospital of Philadelphia, Philadelphia, PA, U.S.A.

14 Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, U.S.A.
Genetic correlations among behavioral and immune phenotypes

15 Institute of Diabetes, Obesity and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, U.S.A.

16 Endocrine Genetics Laboratory, Department of Pediatrics and the Child Health Program of the Research Institute, McGill University Health Centre, Montreal, Quebec, Canada.

17 Quantinuum Research LLC, San Diego, CA, U.S.A.

18 Division of Pulmonary Medicine, The Children’s Hospital of Philadelphia, Philadelphia, PA, U.S.A.

19 K.G. Jebsen Centre for Research on Neuropsychiatric Disorders, University of Bergen, Bergen, Norway

**Target Journal:** Brain, Behavior, and Immunity (Impact 5.89)
Abstract

Evidence of altered immune function and inflammatory signaling has been reported in samples of individuals affected by major psychiatric and neurodevelopmental disorders; it remains unclear how these altered immunological states arise, though a genetic basis has been postulated. The present study sought to use existing summary-level data generated from previous genome-wide association studies (GWAS) in order to explore whether common variant genetic risk factors might be shared between a set of psychiatric/neurodevelopmental phenotypes and a set of medical phenotypes enriched with immune and inflammatory processes. Based on the available GWAS summary data, we calculated the estimated heritability for each phenotype and we examined the genetic correlations between pair-wise combinations of phenotypes, using the LD Score Regression method. We observed positive genetic correlations between bipolar disorder and both celiac disease ($rg = 0.31 \pm 0.09$) and ulcerative colitis ($rg = 0.25 \pm 0.06$), which survived correction for multiple testing. We also observed several robust genetic correlations amongst the set of medical phenotypes enriched for immune and inflammatory processes. We review the relevant clinical literature and suggest that similarities in common variant genetic diatheses may contribute to increased comorbidity between bipolar and autoimmune/inflammatory conditions involving the gastrointestinal tract. We also discuss the limitations of the present approach and important caveats for interpreting the findings.
Introduction

The biological bases of major psychiatric and neurodevelopmental disorders have been studied for decades, yet they remain largely unresolved for the majority affected by apparently idiopathic disease. Evidence from both clinical and biomedical literature has demonstrated that samples of individuals with these conditions show differences in circulating immunologic markers, functional capacities of isolated immune cells, and atypical prevalence of clinical immune-related phenotypes, compared to samples of individuals not affected by psychiatric or neurodevelopmental disorder. It remains unclear what roles (if any) altered immunologic functions may play within the major psychiatric phenotypes, though plausible mechanisms linking altered immune functions with neurobiological changes have been identified. While some studies have already suggested potential genetic bases for the immune-dysregulation observed in a subset of psychiatric patients, the extent to which co-occurrence or segregation of clinical phenotypes may be influenced by similarities in genome-wide genetic risk signals warrants further examination. The goal of the present study was to explore the genetic correlations among these phenotypes using summary data from independent genome-wide association studies (GWAS) or GWAS meta-analyses, which shed light on the disease-related risk associated with single nucleotide polymorphisms (SNPs) that are relatively common in population samples. We hypothesized that some phenotype pairs with evidence for clinical comorbidity might also share similarities in their genome-wide association profile (i.e. significant positive correlations). On the other hand, negative genetic correlations may highlight phenotype pairs with unexpectedly low rates of clinical comorbidity, or may highlight phenotypes at the opposite ends of a polygenic diathesis.

Methods

Literature Search

We searched the published literature (Pubmed, SCOPUS), data repositories (database of Genotypes and Phenotypes; dbGaP) and the downloads page of the Psychiatric Genomics Consortium...
Genetic correlations among behavioral and immune phenotypes - 5

website (https://www.med.unc.edu/pgc/downloads) to identify phenotypes with potentially usable GWAS and GWAS meta-analysis summary statistics. In order to facilitate cross-study comparison, we utilized studies that reported samples of European ancestry, broadly defined to include Central, Southern and Eastern Europe, Scandinavia, and Western Russia. One exception to this was the selection criteria was PGC-II schizophrenia summary data, which reflects a meta-analysis of 46 European cohorts and 3 cohorts of East Asian ancestry. When multiple studies were identified for a given phenotype, we first pursued the study with the largest effective sample size. For studies identified in the published literature, we contacted corresponding authors to request summary statistics. The full list of phenotypes identified in the search and retained for analyses is shown in Supplementary Table 1, along with identifications of the study cohorts and consortia efforts that generated these data, full citations of the respective publications, and indications of sample size, information regarding genomic inflation, and estimated heritability (described below).

**GWAS Phenotypes Obtained and Retained For Genetic Correlation**

For our brain- and behavior-related phenotypes, we obtained summary data reflecting studies of alcohol-related problems in youth, Alzheimer’s disease, angry temperament, anxiety-spectrum disorders, attention-deficit/hyperactivity disorder (ADHD), autism (PGC download page), borderline personality disorder, bipolar disorder, epilepsy (partial seizures), head circumference, intracranial volume, reading disability and language impairment, major depression, obsessive-compulsive disorder (OCD; personal communication from Carol Mathews), personality (5 domains), cigarette smoking (ever-smoked status and cigarettes per day), schizophrenia, substance dependence (including alcohol, opiates, and cocaine), and Tourette Syndrome (personal communication from Carol Mathews). We also included studies of birth weight and length, because of their relevance to intrauterine development and potential relevance to early brain development and neurodevelopmental hypotheses of disorders like schizophrenia and autism. These brain/behavior/development phenotypes were treated as a set in subsequent analyses.
Genetic correlations among behavioral and immune phenotypes - 6

We obtained GWAS data for a variety of medical conditions, many of which are known or suspected to involve alterations to immune cells and/or inflammatory signaling, as well as several continuously measured immune/inflammatory biomarkers. These included atopic dermatitis,38 autoimmune hepatitis,39 blood pressure (systolic, diastolic, mean arterial and pulse pressure),40,41 celiac disease,42 coronary artery disease,43 Crohn’s disease,44 diabetes (type 145 and 246), eosinophilic esophagitis,47 morning plasma cortisol,48 granulomatosis with polyangiitis,49 human immunodeficiency virus (HIV) control50 and HIV–associated sequelae (including dementia, neurocognitive impairment, processing speed, and executive function),51 IgA nephropathy,52 myeloid neoplasia,53 otitis media (chronic),54 psoriatic arthritis,55 rheumatoid arthritis,56 sarcoidosis,57 systemic sclerosis,58 stroke (large vessel, small vessel, cardio-embolic, and all subtypes combined),59,60 systemic lupus erythematosus,61 susceptibility to pulmonary tuberculosis,62 and ulcerative colitis.63 These phenotypes were treated as a set in subsequent analyses.

**Data Pre-Processing and Analysis**

All data pre-processing and analyses were performed using tools supplied within the Linkage Disequilibrium (LD) Score Regression software suite (LDSC; https://github.com/bulik/ldsc).64,65 Briefly, this set of tools can be used with existing GWAS summary data in order to (1) distinguish polygenicity from confounding caused by uncontrolled population stratification or cryptic relatedness among samples,64 (2) estimate the heritability of a given phenotype,65 and (3) estimate the genetic correlation between two phenotypes based on two separate or related sets of summary statistics.65 This method relies on adjustment for the linkage between SNPs (i.e. covariance caused by genomic proximity); for our analyses, we used the set of LD scores provided by the software’s creators, based on the 1000 Genomes Project’s European sample (file = eur_w_ld_chr, URL = https://data.broadinstitute.org/alkesgroup/LDSCORE). Because minor allele frequencies and imputation quality scores were not available for all the obtained sets of GWAS results, we filtered the GWAS results to retain only single nucleotide polymorphisms (SNPs) that were included within the HapMap3 panel and
Genetic correlations among behavioral and immune phenotypes had a minor allele frequency \( \geq 5\% \) within the 1000 Genomes Project Phase 3 European samples. Additional SNP quality controls (QC) routines followed those implemented by the GWAS authors and the defaults employed with the LDSC `munge_sumstats.py` function; this function checks alleles to ensure that the supplied alleles match those in the HapMap3 reference panel. For each data set, we estimated the phenotype’s heritability using the `ldsc.py` function. The results of this analysis, along with features of each GWAS data set (sample size, number of QC+ SNPs, genomic inflation factor, etc.), are shown in Supplementary Table 1. Phenotypes with estimated heritability greater than the error of the estimate were retained for correlational analysis (indicated in Supplementary Table 1); this resulted in the inclusion of seven phenotypes with an effective sample size less than 5000 individuals (indicated in Supplementary Table 1), which was previously recommended as a minimum sample size in order to reduce to error. Pairwise genetic correlations were assessed between all retained phenotypes based on the intersection of QC-positive SNPs and heatmaps were constructed to depict these relationships. For correlation coefficients returned within the bounds of the LDSC software, \( p \)-values were corrected using the Benjamini-Hochberg (BH) method for the total number of unique tests depicted in each correlation matrix. In subsequent figures, the correlations reaching a trend-level significance (0.05 < \( p \) < 0.10) are depicted as colored panels, while relationships surpassing uncorrected \( p < 0.05 \) are additionally denoted with *, and relationships surpassing BH-\( p < 0.05 \) (for the total number of tests depicted in the figure) are denoted with **. The ordering of each phenotype within the rows and columns of each heatmap were based on hierarchical clustering of the correlation coefficients. The extended major histocompatibility (MHC) region contains high amounts of long-range LD, making it challenging to accurately map association signals in this region. For this reason, and following the work of others, we repeated our analyses after the exclusion of this region (chromosome 6, base positions \( 25 \times 10^6 \) to \( 35 \times 10^6 \)). Estimated heritability and features of each GWAS data set after MHC removal are shown in Supplementary Table 2. In the following sections, we primarily report genetic relationships observed under MHC inclusion, and we indicate relationships that changed (i.e. gain or loss of nominal significance) when MHC was excluded.
Results

Correlations between Brain/Behavior/Development and Immune/Inflammatory Phenotypes

We first examined relationships between the set of brain-behavior-development phenotypes and the set of immune and inflammatory phenotypes. The GWAS signals for personality domains and angry temperament did not show significant correlations with immune and inflammatory phenotypes, so they were omitted from the following set of analyses. Pair-wise genetic correlations between brain/behavioral and immune/inflammatory phenotypes are shown in Figure 1 and detailed results are provided in Supplementary Table 3. These same genetic correlations, repeated in the absence of the extended MHC region, are described in Supplementary Table 4. Unless specifically indicated, all correlations described in the main text were maintained at the same significance threshold in the absence of the MHC region.

We attempt to focus our reporting on genetic relationships that have not previously been reported using this method. Notably, four correlations survived BH multiple-test correction and these relationships maintained a corrected level of significance after removal of the MHC region. Among these, were significant positive relationships between bipolar disorder and celiac disease ($r_g = 0.31 + 0.09$) and bipolar disorder and ulcerative colitis ($r_g = 0.25 + 0.06$). We also detected a significant negative correlation between birth weight and coronary artery disease ($r_g = -0.37 + 0.09$), as well as birth weight and type 2 diabetes ($r_g = -0.37 + 0.09$). In the following paragraphs, we highlight relationships that surpassed a nominal significance threshold (uncorrected $p < 0.05$).

In addition to celiac disease and ulcerative colitis, bipolar disorder also showed a nominally significant correlation with Crohn’s disease ($r_g = 0.14 + 0.06$). In the absence of the MHC region, a previously undetected negative relationship between bipolar and rheumatoid arthritis rose to nominal significance ($r_g = -0.15 + 0.07$). Obsessive-compulsive disorder was also positively related to celiac disease ($r_g = 0.39 + 0.18$) and ulcerative colitis ($r_g = 0.29 + 0.09$), but negatively related to type 1 diabetes ($r_g = -0.28 + 0.10$). Schizophrenia showed positive relationships with ulcerative colitis ($r_g =
Genetic correlations among behavioral and immune phenotypes

0.12 ± 0.05), Crohn’s disease (rg = 0.09 ± 0.04), all-cause ischemic stroke (rg = 0.22 ± 0.08), and myeloid leukemias (rg = 0.35 ± 0.13). Schizophrenia showed a positive relationship with susceptibility to tuberculosis (rg = 0.15 ± 0.07) and with HIV viral control (rg = 0.17 ± 0.08), but the latter relationship was diminished to a statistical trend (p < 0.09) upon MHC removal. Curiously, autism showed a negative correlation (rg = -0.43 ± 0.21) with myeloid neoplasms, but this relationship was also diminished to a trend (p < 0.06) upon MHC removal. The GWAS signal of epilepsy with partial seizures was positively related to all-cause ischemic stroke (rg > 0.95 ± 0.35) and with type 2 diabetes (rg = 0.62 ± 0.28). The combined signal reflecting reading disability and language impairment was negatively related to type 2 diabetes (rg = -0.46 ± 0.18). A nominally significant positive relationship between ADHD and systolic blood pressure was detected in the absence of the MHC region (rg = 0.09 ± 0.05), but was diminished to a statistical trend (p < 0.06) upon MHC inclusion.

Interestingly, there was a positive relationship between smoking (cigarettes per day; ever-smoked-status) with all-cause ischemic stroke (rg = 0.46 ± 0.20; rg = 0.53 ± 0.16, respectively), susceptibility to tuberculosis infection (rg = 0.50 ± 0.18, rg = 0.31 ± 0.13), and coronary artery disease (rg = 0.34 ± 0.12; rg = 0.24 ± 0.07). Additionally, cigarettes per day was positively related to type 2 diabetes (rg = 0.27 ± 0.11), but negatively related to morning plasma cortisol levels (rg = -0.80 ± 0.39). Furthermore, ever-smoked-status was positively related to rheumatoid arthritis (rg = 0.14 ± 0.05) and large vessel stroke (rg = 0.71 ± 0.34), though the latter relationship was diminished to a statistical trend (p < 0.07) upon removal of the MHC.

In addition to the relationships involving birth weight, several interesting relationships involving other developmental phenotypes were observed. Birth length was negatively related to type 2 diabetes (rg = -0.23 ± 0.09). Head circumference was negatively related to atopic dermatitis (rg = -0.31 ± 0.14), ulcerative colitis (rg = -0.24 ± 0.10), and pulse pressure (rg = -0.07 ± 0.03), though the latter relationship was diminished to a statistical trend upon MHC removal. Additionally, upon MHC removal, head circumference showed a significant negative relationship (rg = -0.39 ± 0.19) with eosinophilic
esophagitis. Intracranial volume was positively related to Crohn’s disease \( (r_g = 0.29 \pm 0.11) \) and negatively related to coronary artery disease \( (r_g = -0.32 \pm 0.13) \).

We observed a number of highly significant genetic correlations among the brain/behavior/development phenotypes; these are depicted in Figure 2, with full summary statistics in Supplementary Tables 5 (with MHC) and 6 (without MHC). Additionally, we observed several highly significant and previously unquantified genetic correlations among the immune and inflammatory phenotypes; these are depicted in Figure 3, with full summary statistics in Supplementary Table 7 (with MHC) and 8 (without MHC). Among those surviving family-wise BH correction were positive relationships between celiac and Crohn’s disease \( (r_g = 0.31 \pm 0.09) \), celiac and type 1 diabetes \( (r_g = 0.32 \pm 0.09) \), Crohn’s and ulcerative colitis \( (r_g = 0.59 \pm 0.06) \), and coronary artery disease and type 2 diabetes \( (r_g = 0.43 \pm 0.07) \). Notably, these relationships maintained a corrected level of significance after removal of the MHC region. Additionally, several nominally significant relationships were observed. Celiac disease showed modest positive relationships with ulcerative colitis, rheumatoid arthritis, and systemic sclerosis \( (r_g = 0.30 \pm 0.11, 0.24 \pm 0.08, \text{and} 0.39 \pm 0.19, \text{respectively}) \). Ulcerative colitis also showed positive relationships with rheumatoid arthritis \( (r_g = 0.13 \pm 0.06) \). Rheumatoid arthritis showed a positive relationship with systemic sclerosis \( (r_g = 0.38 \pm 0.14) \). Type 1 diabetes was positively related to autoimmune hepatitis \( (r_g = 0.55 \pm 0.19) \) and rheumatoid arthritis \( (r_g = 0.30 \pm 0.15) \). IgA nephropathy was negatively correlated with rheumatoid arthritis \( (r_g = -0.47 \pm 0.17) \), Crohn’s disease \( (r_g = -0.47 \pm 0.20) \), and atopic dermatitis \( (r_g = -0.66 \pm 0.30) \). Atopic dermatitis was positively related to sarcoidosis \( (r_g = 0.45 \pm 0.21) \). Additionally, several aspects of blood pressure showed a weak positive relationship with myeloid neoplasia \( (r_g = 0.15 \pm 0.08) \). Susceptibility to tuberculosis was negatively related \( (r_g = -0.58 \pm 0.24) \) to the ability to control HIV infection. Upon removal of the MHC region, several previously non-significant correlations rose to nominal significance, including a negative relationship between type 1 and type 2 diabetes \( (r_g = -0.25 \pm 0.09) \) and positive relationships between rheumatoid arthritis and
autoimmune hepatitis ($rg = 0.58 \pm 0.21$), Crohn’s and susceptibility to tuberculosis ($rg = 0.23 \pm 0.10$), and myeloid neoplasia and all-cause ischemic stroke ($rg = 0.73 \pm 0.34$).

**Discussion**

Quantitative SNP-based genetic relationships between disorders have been explored previously using other statistical approaches (restricted maximum likelihood [REML] co-heritability, polygenic risk scores, genetic analysis incorporating pleiotropy and annotations, and other permutation-based methods). It is apparent that different approaches to examining genetic relationships across disorders are prone to arriving at different conclusions. For example, a recent study by Pouget and colleagues demonstrated that polygenic risk scores reflecting additive risk for several autoimmune disease can explain a small proportion of variance in schizophrenia case-control status, yet the genome-wide significant SNPs from the autoimmune GWASs were not over-represented among schizophrenia’s genome-wide significant hits when permutation-based analysis was performed. While other methods are better suited to characterizing the specific loci contributing to shared genetic risk, the approach taken here attempts to quantitate similarities and differences in association signals across the entire genome. Using this approach, we identified several relationships between brain/behavioral phenotypes and immune/inflammatory phenotypes that were robust to multiple testing; we focus our discussion on those. Some of these phenotype pairs have assessed previously using the LD score regression method and we compare our results with the findings of those previous studies.

Prominent among our findings was a positive genetic correlation between bipolar and celiac disease. To our knowledge, this relationship has not previously been quantitatively reported. Several large-scale epidemiological studies have reported that a history of autoimmune disease increases the subsequent risk for bipolar disorder (along with other psychiatric phenotypes), and a history of celiac disease appears to significantly increase the subsequent risk for mood disorders (including bipolar), though well-powered reports linking a history of celiac disease with bipolar disorder (in particular) have
Genetic correlations among behavioral and immune phenotypes - 12

not yet emerged. More comprehensive reviews of the neurological and psychiatric findings reported in association with celiac disease or gluten insensitivity can be found elsewhere, along with more comprehensive reviews of immune-related findings in bipolar disorder. Here, we focus on a small body of clinical literature examining celiac disease or gluten hypersensitivity in connection with bipolar disorder. One cross-sectional study of patients ascertained for celiac disease identified a higher incidence of bipolar disorder than comparison subjects without celiac disease. Conversely, studies of patients ascertained for bipolar disorder have examined serological markers in order to gain insight into this relationship.

Celiac disease is typically distinguished from the apparently more common condition of gluten sensitivity or gluten-related allergy based on the presence of histological damage upon intestinal biopsy and the presence of serological markers indicating an autoimmune response involving the adaptive immune system. With respect to serological markers, anti-endomysial and anti-tissue transglutaminase antibodies tend to be considered more specific for celiac disease, while anti-deamidated gluten peptide may be assessed in certain situations (e.g. IgA deficiency) to support diagnosis, and anti-gliadin antibodies could be considered compatible with both celiac or gluten sensitivity without intestinal damage. Similarly, IgA autoantibodies tend to be considered more specific, but IgG autoantibodies may also be considered, particularly in the setting of IgA deficiency.

In one study of patients ascertained for bipolar disorder, higher levels of IgG anti-gliadin and suggestive evidence of higher levels of IgG to de-amidated gliadin peptide were observed compared to comparison subjects without a psychiatric diagnosis, but no significant differences were observed in levels of IgA to either antigen or in levels of IgA or IgG to tissue transglutaminase enzyme. The same group examined individuals hospitalized with acute mania, and reported significantly increased levels of IgG anti-gliadin compared with non-psychiatric control subjects, but did not replicate the finding of increased IgG to de-amidated gliadin peptide. As a whole, this group of acutely hospitalized individuals showed no significant difference in IgG anti-gliadin levels compared with controls at a six-month
Genetic correlations among behavioral and immune phenotypes - 13

longitudinal follow-up. However, individuals showing persistently elevated levels of IgG anti-gliadin at follow-up were more likely to have been hospitalized during the intervening period. Another study screened a sample of 104 psychiatric inpatients and 41 non-psychiatric comparison subjects for concentrations of 14 different autoantibodies. IgG anti-gliadin was the most frequently detectable antibody in the patients, while anti-double-stranded DNA was the most frequently detected among controls. With respect to group differences in concentrations, they observed significantly higher levels of IgG anti-gliadin (but no other antibodies) among the psychiatric patients. The sample reflected a mixture of diagnostic phenotypes, with approximately 15.5% classified as bipolar and another 22.3% classified as schizoaffective. The titers of these antibodies were reported to be within the low range, with respect to the threshold typically used to inform the diagnosis of autoimmune disease. Taken together, the clinical literature provides some support for the ideas that: 1) celiac disease incidence may be higher in bipolar-affected individuals compared with non-psychiatric comparison subjects; and 2) a subset of bipolar disorder patients demonstrate serological patterns that could be consistent with gluten hypersensitivity. The present findings of a positive genetic correlation between celiac disease and bipolar disorder suggests that similarities in the common variant genetic risk signal for these two conditions may underlie these clinical associations. The phenomena of autoantibodies in individuals with bipolar disorder has found mixed support over the past few decades. A recent meta-analysis concluded that there was significantly higher odds of NMDA receptor antibody seropositivity among those with major psychiatric disorders (including schizophrenia or schizoaffective, bipolar, or major depressive disorders). Taken together, the present study suggests an understanding of the genetic diathesis (or gene-x-environment interactions) that dispose individuals to autoantibody production in celiac disease may help shed light on the phenomenon of autoantibody production in bipolar disorder.

We also observed an apparently robust positive genetic correlation between bipolar disorder and ulcerative colitis. A previously mentioned epidemiological study found increased prevalence of mood disorders (including bipolar) among individuals with ulcerative colitis (as well as a number of other
genetic correlations among behavioral and immune phenotypes)\textsuperscript{72} Additionally, another study reported a positive association between ulcerative colitis (as well as Crohn’s disease) and risk for a concurrent diagnosis of bipolar disorder\textsuperscript{87}. The literature regarding the genetic similarities between bipolar and ulcerative colitis appears mixed. The first study implementing the LD score regression method reported no significant correlation ($r_g = 0.08 \pm 0.08$, uncorrected $p = .33$) between these phenotypes; the present study used the same bipolar GWAS summary data, but a different version of the ulcerative colitis GWAS data, and we restricted our analysis to SNPs with a minor allele frequency $\geq 5\%$ in the 1000 Genomes Project Phase 3 European samples. In keeping with this initial report, a similar genetic correlation is also reported in LD-Hub (\url{http://ldsc.broadinstitute.org/}), using what appear to be the same data sets. Thus, it seems data set selection and preprocessing can have a considerable effect on the result of the LD score regression method. On the other hand, Wang \textit{et al.}\textsuperscript{14} used several approaches that were not dependent on the directionality of a given SNP’s effect, and they concluded that many (24 of 35) pairwise combinations of psychiatric and immune-related phenotypes shared a statistically significant proportion of risk-associated loci; among these findings was a significant genetic correlation between bipolar disorder (as well as schizophrenia) and ulcerative colitis. However, many of the other relationships identified by Wang and colleagues were not significant in the present study. Finally, one previous study implementing a REML-based approach did not find significant SNP-based co-heritabilities between Crohn’s disease and the major psychiatric phenotypes\textsuperscript{68}. Taken together, the existing literature seems to support the possibility that a genetic relationship may play a role in the increased comorbidity between bipolar and ulcerative colitis.

We observed several robustly significant genetic correlations among brain-behavioral phenotypes (particularly psychiatric phenotypes), all of which have been reported previously using related sets of summary statistics reflecting published versions of PGC Working Group primary studies (used presently)\textsuperscript{66}, unreleased versions of PGC Working Group results\textsuperscript{71}, or previously published results from the PGC Cross Disorder GWAS study.\textsuperscript{65} These included positive inter-correlations between
schizophrenia, bipolar disorder, major depression, and obsessive compulsive disorder. Differences in estimated heritabilities, genetic correlation point estimates, and error estimates across these studies could be accounted for by utilization of different versions of the GWAS summary data, differences in pre-processing methods, or differences in the implementation of the genetic correlation with constrained model intercept based on known proportion of sample overlap; this latter feature was not performed in the present study, because only association summary-level data could be provided for our study and we did not want to risk the introduction of bias through miscalculation of sample overlap. Furthermore, we noticed that several apparently modestly powered ordinal phenotypes with high estimated heritability (e.g. opioid and cocaine dependence) consistently produced genetic correlations that were above or below the accepted boundaries of the LDSC software. Despite these differences, we observed many of the same robustly significant genetic correlations that were previously reported among brain-behavioral phenotypes, and these correlations maintained similar significance and correlation strength in the presence and absence of the extended MHC region.

Genetic relationships among many of the immune and inflammatory phenotypes had not been quantified previously using this method, though previous studies employing alternative methods have suggested considerable genetic pleiotropy\textsuperscript{70} and genetic correlation.\textsuperscript{88} We observed robust positive genetic correlations between celiac disease and type 1 diabetes. Type 1 diabetes is caused by the autoimmune destruction of insulin-secreting pancreatic beta cells. The co-segregation of type 1 diabetes and celiac disorder (or its serological markers) is a well-documented observation.\textsuperscript{89–93} Notably, genetic risk for both celiac and type 1 diabetes are mediated by some of the same Class II HLA-DBQ1 alleles,\textsuperscript{91,94,95} however the present study and others support the idea that these phenotypes share common variant genetic risk factors outside of the extended MHC region.\textsuperscript{92,94} We also observed a robust positive genetic correlation between celiac and Crohn’s disease. Celiac disease is relatively more rare, limited to the small intestines, and is thought to involve an inappropriate immune response to dietary constituents and their metabolic biproducts. In Crohn’s disease, inflammation can occur throughout the
gastrointestinal tract (though most commonly affects the ileum of the small intestines) and is thought to involve abnormal reactions to normal commensal flora of the bowel in genetically susceptible individuals. Within the clinical literature, only a few studies of small samples have explicitly examined the comorbidity between these two phenotypes, with seemingly equivocal results across studies. Despite the unclear clinical picture, genetic studies have employed cross-phenotype GWAS comparison and meta-analysis to identify loci that might mediate risk for both disorders.

We observed a robust positive genetic correlation between Crohn’s and ulcerative colitis. While the vast majority of published studies have examined the differential diagnosis of these two conditions, there is a smaller body of literature examining their clinical comorbidity. These two phenotypes appear to co-occur within families, though studies of co-occurrence within individual patients appear to be limited to case reports. The heritability of Crohn’s disease appears to be higher than that of ulcerative colitis, as does the family risk (for any inflammatory bowel disorder) imparted to first-degree relative. Because these two phenotypes are often considered jointly as inflammatory bowel disease in previously published studies, little direct light has been shed on whether the extent to which a family history of one increases subsequent risk for the other. Nonetheless, the apparent similarity between the phenotypes stimulated a substantial body of genetic studies highlighting shared risk loci for these conditions.

We must recognize that several of the GWAS data sets used in the present study were likely under-powered for reliable heritability estimation and genetic correlation, hence some genetic correlations between these phenotypes may have gone undetected. Our study yielded an abundance of nominally significant genetic correlations, many of which probably reflect false-positive findings, and thus the set of findings should be considered with appropriate caution. Another important critique of this exploratory examination of existing GWAS data is related to the lack of clearly identified positive and negative control comparisons; these types of comparisons would be essential in order to assess the sensitivity and specificity of the LD score regression approach to assessing genetic relationships. Studies designed to
assess the limitations of this method may be especially warranted in light of the apparent absence of genetic correlation between phenotypes in which a significant positive correlation might be presumed (e.g. blood pressure and stroke, stroke and coronary artery disease). Despite these limitations, many of the nominally significant correlations are intriguing and warrant further examination; future studies may be able to utilize GWAS data derived from larger samples or more selective test hypotheses to alleviate the burden of multiple testing.

More generally, it is important to recognize that genetic correlations detected in this study could be influenced by a number of phenomena. As discussed previously, a heritable, yet unexamined underlying trait might confer risk to two or more phenotypes, thereby giving rise to an apparent genetic correlation between them. Related to this concept, the present study identified nominally significant positive correlations between smoking behavior and several of the immune and inflammatory conditions thought to be caused or exacerbated by smoking. These findings suggest that GWAS signal of some immune or inflammatory diseases may actually be tainted with GWAS signal reflecting the propensity toward smoking behavior and that statistical control or post-hoc adjustment of results may help distinguish these two effects. We also must recognize that GWAS studies of psychiatric phenotypes typically do not screen affected cases based on general medical conditions, and studies of immune-related phenotypes may not exclude individuals with psychiatric comorbidity. Thus, if there is true co-segregation of two phenotypes in the population, it is possible that a psychiatric sample may unwittingly over-represent the cases of a given immune-related phenotype, and vice-versa, contributing to the appearance of positive genetic correlation. Alternatively, it is unclear the extent to which explicit exclusion of individuals from a control group could bias the estimated genetic correlation in certain circumstances (e.g. the exclusion of individuals with type 1 diabetes from the control group of a study of type 2 diabetes and vice-versa). Another caveat to consider in any cross-phenotype genetic study involving clinically similar conditions is that the estimate of genetic similarities could be influenced by a subset of misdiagnosed cases. General limitations of this method (in comparison with other
approaches) have been discussed previously,\textsuperscript{65,71} and include the requirement for large sample sizes, the current inability to extend to samples of recently admixed ancestry, the present inability to account for the effects of other factors (\textit{e.g.} assortative mating),\textsuperscript{109} and reliance on the assumption that traits examined are sufficiently polygenic. Other methods may be better suited to examine or quantitate genetic relationships between phenotypes where common variation is known to be associated with large effect sizes.\textsuperscript{65}

Additionally, consideration of expression quantitative trait loci, differentially expressed or methylated genes, or enriched ontological and functional terms may provide a broader context for assessing biological similarities between phenotypes.

With the recent launch of a web-based platform for performing LD-score regression analyses (\url{http://ldsc.broadinstitute.org/}) and for sharing GWAS summary-level data,\textsuperscript{66} investigators can perform their own analyses for phenotypes of interest with relative ease. We hope that the availability of well-powered GWAS summary data will increase as tools like these gain more widespread use. Nevertheless, we identified several phenotypes (multiple sclerosis, psoriasis, systemic lupus erythematosus, glioma, and circulating IgE and C-reactive Protein) for which well-powered GWAS summary data exist, but could not be provided for the present analysis. The broader research community should commend the monumental efforts of research teams and consortia that make possible large-scale GWAS and GWAS meta-analyses. Because of the immense amount of work required for subject ascertainment, genotyping, imputation, quality control, and association testing, there is considerable incentive to delay the release of GWAS summary data into the public domain. As such, groups electing to disseminate their summary-level data should be commended for this additional level of commitment to the goal of advancing the scientific understanding of human disease. The present study, as well as its predecessors,\textsuperscript{65,66,71} and other approaches for cross-phenotype GWAS integration,\textsuperscript{70,110–112} justify the impetus for open sharing of GWAS results in order to accelerate the discovery of previously undetected genetic relationships that extend across medical disciplines and to support the identification and prioritization of multi-disease-associated loci.
While there is clear and historic support for the assertion that psychiatric and neurodevelopmental disorders primarily involve alterations to tissues, circuits, and resident cells of the brain (i.e. neurons and glia), a growing body of literature suggests that the cells and signaling molecules of the immune system are poised to influence the early development of these entities, as well as their plasticity and function in fully developed organisms. As plausible mechanisms linking altered immune functions with neurobiological changes are identified in model systems, it will provide an increasingly clear context for understanding how altered immune and inflammatory states might be contributing to the development or symptomatic maintenance of disorder in subsets of psychiatric patients. In support of these future goals, the present study leveraged existing data and identified immune-related phenotypes that appear to share genetic diatheses with the psychiatric and neurodevelopmental phenotypes.

**Acknowledgements**

The authors gratefully acknowledge the contributions all of the individuals (patients, families, clinicians and diagnosticians, research associates, and data analysts) and consortia whose efforts made possible the GWAS studies and meta-analyses featured in the present study. For most of the phenotypes examined in the present study, clinical and genetic data were collected across numerous sites, each with their own unique patients, staff, and funding sources. While we attempted to provide more thorough recognition of the required acknowledgments for each individual phenotype in our supplementary note, we realize that it is not possible to recognize every individual and funding mechanism that made these studies possible, and we apologize for this. We also gratefully acknowledge the developers of the LDSC software. The authors declare no conflicts of interest related to this study. Daniel S. Tylee, Jonathan L. Hess, and Stephen J. Glatt were supported by grants from the U.S. National Institutes of Health (R01MH101519, P50MH081755), the Gerber Foundation, the Sidney R. Baer, Jr. Foundation, NARSAD: The Brain & Behavior Research Foundation. Daniel S. Tylee was also supported by the Autism Speaks foundation (Weatherstone Pre-doctoral Training Grant #9645). Rainer Malik was supported by the Vascular Dementia Research Foundation. Andrew Levine was supported by the UCLA-AIDS Institute.
Genetic correlations among behavioral and immune phenotypes - 20

and the UCLA Center for AIDS Research AI28697 (Freimer & Levine); as well as NIH R03DA026099 (Levine). We gratefully acknowledge Susan Service for her assistance preparing and analyzing the data supporting the original association studies of neurocognitive impairment and dementia in HIV-affected adults. Sergey Nejenstev is a Wellcome Trust Senior Research Fellow in Basic Biomedical Science and is also supported by the NIHR Cambridge Biomedical Research Centre.

Author Lists for Contributing Consortia

Attention-Deficit Hyperactivity Disorder Working Group of the Psychiatric Genomics Consortium


Autism Spectrum Disorders Working Group of the Psychiatric Genomics Consortium


Bipolar Disorders Working Group of the Psychiatric Genomics Consortium

Genetic correlations among behavioral and immune phenotypes - 22


CORtisol NETwork (CORNET) Consortium


Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium

Patrick Sullivan, Stephanie Ripke, Danielle Posthuma, Henning Tiemeier, André G Uitterlinden, Nese Direk, Saira Saeed Mirza, Albert Hofman, Susanne Lucae, Stefan Kloiber, Klaus Berger, Jürgen Wellmann, Bertram Müller-Myhsko, Qingqin Li, Marcus Ising, Till Andlauer, Marcella Rietschel, Andreas Josef Forstner, Fabian Streit, Jana Strohmaier, Maren Lang, Josef Frank, Stefan Herms, Stephanie Witt, Jens Treutlein, Markus Nöthen, Sven Cichon, Franziska Degenhardt, Per Hoffmann, Thomas G. Schulze, Bernhard Baune, Udo Dannlowski, Tracy Air, Grant Sinnamon, Naomi Wray, Andrew McIntosh, Douglas Blackwood, Toni-Kim Clarke, Donald MacIntyre, David Porteous, Caroline Hayward, Tonu Esko, Evelin Mihailev, Lili Milani, Andres Metspalu, Hans J Grabe, Henry Völzke, Alexander Teumer, Sandra Van der Auwera, Georg Homuth, Matthias Nauck, Cathryn Lewis, Gerome Breen, Margarita Rivera, Michael Gill, Nick Craddock, John P Rice, Michael Owen, Peter McGuffin, Henriette Buttenschon, Ole Mors, Anders Børglum, Jakob Grove, Jesper Krogh, Enrico Domenici, Daniel Umbricht, Jorge A Quiroz, Carsten Horn, Enda Byrne, Baptiste Couvy-Duchesne, Scott Gordon, Andrew C Heath, Anjali Henders, Ian Hickie, Pamela AF Madden, Nicholas Martin, Sarah Medland, Grant Montgomery, Dale Nyholt, Michele Pergadia, Divya Mehta, Martin Preisig, Enrique Castelao, Zoltán Kutalik, Steven Hamilton, Patrick McGrath, Katherine Tansey, Rudolf Uher, Glyn Lewis, Michael O'Donovan, Brenda WJH Penninx, Yuri Milaneschi, Wouter Peyrot, Johannes H Smit, Rick Jansen, Aartjan TF Beekman, Robert A Schoevers, Albert van Hemert, Gerard van Grootheest, Dorret Boomsma,

22
Genetic correlations among behavioral and immune phenotypes - 23


METASTROKE consortium of the International Stroke Genetics Consortium.


Obsessive Compulsive Disorder and Tourette Syndrome Working Group of the Psychiatric Genomics Consortium


Schizophrenia Working Group of the Psychiatric Genomics Consortium

Genetic correlations among behavioral and immune phenotypes - 25

References


Genetic correlations among behavioral and immune phenotypes - 30


Genetic correlations among behavioral and immune phenotypes - 31


Genetic correlations among behavioral and immune phenotypes - 34


85. Dickerson F, Stallings C, Vaughan C, Origoni A, Khushalani S, Yolken R. Antibodies to the
Genetic correlations among behavioral and immune phenotypes - 36


| Condition                              | Morning Cortisol | Myeloid Neoplasm | Systemic Lupus Erythematosus | Pernicious Anemia | Systemic Sclerosis | Coronary Artery Disease | Sarcoidosis | Nephritis | Pulmonary Hypertension | Pneumonitis | Pneumonia | Scleroderma | Sjogren's | Sarcoidosis | Uveitis | Vasculitis | Rheumatoid Arthritis | Type 1 Diabetes | Type 2 Diabetes | Type 3 Diabetes | Type 4 Diabetes | Type 5 Diabetes | Type 6 Diabetes | Type 7 Diabetes | Type 8 Diabetes | Type 9 Diabetes | Type A Diabetes | Type B Diabetes | Type C Diabetes | Type D Diabetes | Type E Diabetes | Type F Diabetes | Type G Diabetes | Type H Diabetes | Type I Diabetes | Type J Diabetes | Type K Diabetes | Type L Diabetes | Type M Diabetes | Type N Diabetes | Type O Diabetes | Type P Diabetes | Type Q Diabetes | Type R Diabetes | Type S Diabetes | Type T Diabetes | Type U Diabetes | Type V Diabetes | Type W Diabetes | Type X Diabetes | Type Y Diabetes | Type Z Diabetes |
|----------------------------------------|------------------|------------------|-----------------------------|-------------------|-------------------|-------------------|----------------|-----------|---------------------|------------|-----------|----------------|-------------|-------------|---------------|-------------|-------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
Figure Titles and Legends

Figure 1. Genetic correlations between the set of brain/behavior phenotypes and the set of immune/inflammatory phenotypes.

Figure 1 Legend. A modified heatmap was constructed to depict the strength of genetic correlations between the set of brain/behavioral phenotypes and the set of immune/inflammatory phenotypes. The extended MHC region was included for these analyses. Brain/behavioral phenotypes are displayed in rows and immune/inflammatory phenotypes are displayed in columns. The ordering of phenotypes in rows and columns is based on hierarchical clustering. The coefficient for genetic correlation is depicted in each cell. Cells are colored for genetic correlations reaching an uncorrected \( p \)-value < 0.10. Genetic correlations reaching an uncorrected \( p \)-value < 0.05 are denoted with *, while correlations surviving Benjamini-Hochberg correction (for the total number of unique tests depicted in the Figure) are denoted with **. The full list of test statistics supporting this table are provided in Supplementary Table 3. Additionally, these test statistics were recalculated in the absence of the extended MHC region in Supplementary Table 4.

Figure 2. Genetic correlations amongst the set of brain/behavior phenotypes.

Figure 2 Legend. A modified heatmap was constructed to depict the strength of genetic correlations between amongst the set of brain/behavioral phenotypes. The extended MHC region was included for these analyses. The ordering of phenotypes in rows and columns is based on hierarchical clustering. The coefficient for genetic correlation is depicted in each cell. Cells are colored for genetic correlations reaching an uncorrected \( p \)-value < 0.10. Genetic correlations reaching an uncorrected \( p \)-value < 0.05 are denoted with *, while correlations surviving Benjamini-Hochberg correction (for the total number of unique tests depicted in the Figure) are denoted with **. The full list of test statistics supporting this table are provided in Supplementary Table 5. Additionally, these test statistics were recalculated in the absence of the extended MHC region in Supplementary Table 6.

Figure 3. Genetic correlations amongst the set of immune/inflammatory phenotypes.
Figure 3 Legend. A modified heatmap was constructed to depict the strength of genetic correlations between amongst the set of immune/inflammatory phenotypes. The extended MHC region was included for these analyses. The ordering of phenotypes in rows and columns is based on hierarchical clustering. The coefficient for genetic correlation is depicted in each cell. Cells are colored for genetic correlations reaching an uncorrected \( p \)-value < 0.10. Genetic correlations reaching an uncorrected \( p \)-value < 0.05 are denoted with *, while correlations surviving Benjamini-Hochberg correction (for the total number of unique tests depicted in the Figure) are denoted with **. The full list of test statistics supporting this table are provided in Supplementary Table 5. Additionally, these test statistics were recalculated in the absence of the extended MHC region in Supplementary Table 6.