Genetic and epidemiological analyses of infection load and its relationship with psychiatric disorders

Ron Nudel¹,², David M. Hougaard²,³, Thomas Werge²,⁴,⁵, Michael E. Benros¹,²,⁶,*

¹ CORE-Copenhagen Research Centre for Mental Health, Mental Health Centre Copenhagen, Copenhagen University Hospital, Copenhagen, Denmark
² iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Denmark
³ Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark
⁴ Institute of Biological Psychiatry, Mental Health Centre Sct. Hans, Mental Health Services Copenhagen, Roskilde, Denmark
⁵ Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
⁶ Department of Immunology and Microbiology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

* Corresponding author: Michael E. Benros. Email: Michael.Eriksen.Benros@regionh.dk.

This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.
ABSTRACT

Severe infections and psychiatric disorders have a large impact on both society and the individual. Therefore, studies investigating these conditions, and, in particular, the links between them, are of great importance. Past studies have shown that severe infections increase the risk of several psychiatric disorders, and that both types of conditions are correlated genetically. However, most studies have, in this regard, focused on binary phenotypes of particular infections or overall infection, thereby losing some information regarding susceptibility to infection as reflected in the number of specific infection types, or sites, which we term infection load. In this study we leveraged the large iPSYCH sample (N=65,534) to assess the effect of infection load on risk of psychiatric disorders and to investigate the former’s genetic architecture. Infection load was associated with increased risk of ADHD, ASD, bipolar disorder, depression, schizophrenia and overall psychiatric diagnosis. We obtain a modest but significant heritability estimate for infection load ($h^2=0.0221$), and a high degree of genetic correlation between infection load and overall psychiatric diagnosis ($r_g=0.4298$). We replicate the latter result and the result for overall infection using an external sample, where genetic correlation between susceptibility to overall infection and overall psychiatric diagnosis is $>0.4$ both within and across samples. We also find evidence supporting a genetic causality for overall infection on overall psychiatric diagnosis. Our genome-wide association study for infection load identified 138 suggestive associations. Our study thus provides further evidence for genetic links between susceptibility to infection and psychiatric disorders, and suggests that a higher infection load may have a cumulative association with psychiatric disorders, beyond what has been described for individual infections.
INTRODUCTION

Infections are one of the major global health concerns and a leading cause of early mortality (1, 2). In addition to the health risk conferred directly by severe infections, studies have shown the latter to be associated with increased risk of mental disorders, including schizophrenia (3), mood disorders such as depression (4), and neurodevelopmental disorders such as autism spectrum disorder (ASD) (5). While infections are inherently caused by external pathogens, susceptibility to infection has a genetic component (6).

In addition to replicating the epidemiological associations between psychiatric disorders and infections, we showed in a previous study that susceptibility to severe infections (i.e. infections requiring hospital contact) had a modest but significant heritability (3.2% on the observed scale), and that the genetic correlation between overall infection (defined as having at least one diagnosis from a variety of infection categories included in the study) and overall psychiatric diagnosis (ICD-10 codes F00-F99 and/or ICD-8 codes 290-315, but mostly one or more of: schizophrenia, bipolar disorder or depression (affective disorder), ASD, attention deficit/hyperactivity disorder (ADHD) and anorexia) was 0.4-0.5, depending on the method used to estimate it (7). Another study which used the iPSYCH sample found genetic overlaps between susceptibility to infections and specific psychiatric disorders, namely schizophrenia, ADHD, depression, bipolar disorder, post-traumatic stress disorder, but, interestingly, not ASD, using genetic correlation analysis and/or polygenic risk scores (8). Interestingly, a recent study has shown positive genetic correlations between blood levels of C-reactive protein, a marker for infection and inflammation, and major depressive disorder (MDD) and ADHD, as well as positive genetic correlations between white blood cell counts and MDD, ADHD and schizophrenia (9), which could provide insights into the molecular mechanisms underlying the
genetic correlation between susceptibility to infection and psychiatric disorders, although the trends were not always similar across individual psychiatric disorders in that study.

The aim of this study was to investigate these links further by employing a quantitative phenotype for infection load in the iPSYCH cohort, comprising 65,534 individuals selected as cases of major psychiatric disorders (schizophrenia, bipolar disorder or depression (affective disorder), ASD, ADHD and anorexia) or as part of a random population sample. Infection load in this study is defined as the number of site-specific infection categories (central nervous system infection, gastrointestinal infections, genital infections, hepatitis, otitis, pregnancy-related infections, respiratory infections, sepsis, skin infections, HIV/AIDS and urological infections) for which an individual received at least one diagnosis. Thus, infection load in this sense makes use of register-based diagnoses, similar to other studies (10), to get a quantitative measure for an individual’s overall susceptibility to severe infections using binary trait data (the presence or absence of a diagnosis). This extends our previous study on overall infection as a binary trait. Additionally, we replicate our previous results and results from this study using an external sample and perform genetic correlation analyses both within and across the two samples.
METHODS

Data sources for diagnoses and study sample

The sample and phenotypes used in this study has been described in our previous publications (7, 11-15), and we repeat the details here, with changes relevant to the present study: the iPSYCH sample linked data from the Danish medical registers to biobank data via the unique civil registration number used in Denmark since 1968 (16). Biological data from the Danish Neonatal Screening Biobank include dried blood spots taken 4-7 days after birth from nearly all infants born in Denmark after 1981 (16, 17), which was used for downstream genotyping. Infection diagnoses were obtained from the Danish National Hospital Register, which, since 1977, has included records of all inpatients treated in Danish non-psychiatric hospitals, and, since 1995, has included information regarding outpatient and emergency room contacts (18). The Psychiatric Central Research Register includes data from all psychiatric inpatient facilities since 1969 and outpatient contacts since 1995 (19). In Denmark, diagnoses were based on the 8th Revision of the International Classification of Diseases (ICD-8) (20) from 1977 to 1993, and, since 1994 they have been based on ICD-10 (21). For the psychiatric phenotypes included in this study, all diagnosis types apart from henvisningsdiagnose (referral diagnosis) were included, and the types of contact included were inpatient, outpatient and emergency care unit contact. Data for psychiatric phenotypes are from the Psychiatric Central Research Register. For infection phenotypes, the included types of diagnosis were the following: ICD-8: hoveddiagnose and bidiagnose (main and auxiliary diagnosis, respectively); ICD-10: aktionsdiagnose, grundmorbus, bidiagnose (main, basic and auxiliary diagnosis, respectively), and the included types of contact were inpatient and outpatient hospitalizations and emergency room contact, all from the Danish National Hospital Register. Tillægsdiagnoser (associated diagnoses) were not considered, and
diagnoses of the following types were excluded: henvisningsdiagnose (referral), komplikation (complication). ICD-8 diagnoses with the following modifications: “Obs. Pro” and “Ej befundet” (suspected and not found, respectively) were also excluded. All individuals in this study are part of the iPSYCH 2012 sample (22), drawn from all individuals born in Denmark between 1981 and 2005 (N=1,472,762), and which included individuals diagnosed with at least one of: schizophrenia, bipolar disorder or depression (affective disorder), autism spectrum disorder, attention deficit/hyperactivity disorder and anorexia, and individuals who were selected as part of a randomly selected sample from the Danish population. The iPSYCH sample used in this study has undergone several rounds of quality control (QC) as described in our previous studies (7, 11-15) using data from high-quality genetic markers prior to the imputation. The main sample QC steps included the removal of individuals based on ancestry (individuals who did not have Danish ancestry, as determined from register data of family history and genetic principal component analyses (PCAs) had been removed), as well as relatedness (if they were first- or second-degree relatives of other individuals in the sample; this step prioritized iPSYCH cases and then individuals with a higher genotype call rate). Other QC steps involved the removal of individuals based on missingness (>1%), abnormal heterozygosity, ambiguous sex (discrepancies between annotated sex and genetic data), or if they were duplicates of other individuals. The first studies employing this QC protocol have more information about the procedures (23, 24), including the supplementary methods of a preprint of the former (25) and of another study (26).

Before QC, 78,050 individuals from 23 genotyping waves were included. Following QC, 65,534 unrelated Danish individuals were retained, of whom 34,705 were male and 30,829 were female. Data up to the end of 2012 were included for infections, and data up to the end of 2013 were included for psychiatric diagnoses. We had diagnoses for the following infection categories:
bacterial, viral, central nervous system infection, gastrointestinal, genital, hepatitis, otitis, pregnancy infection, respiratory, sepsis, skin infection, HIV/AIDS, urological and other infections (e.g. protozoan infections). ICD codes for these are provided in Supplementary Table S1. For psychiatric disorders, the following ICD codes were used: any/overall psychiatric diagnosis (ICD-8 code within the range 290-315 and/or ICD-10 code within the range F00-F99); ADHD (ICD-10: F90.0); anorexia (ICD-8: 306.50, ICD-10: F50.0); ASD (ICD-10: F84.0, F84.1, F84.5, F84.8, F84.9); bipolar disorder (ICD-8: 296.19, 296.39, 298.19, ICD-10: F30-F31); depression (single and recurring) (ICD-8: 296.09, 296.29, 298.09, 300.49, ICD-10: F32-F33); schizophrenia (ICD-8: 295.x9 (excl. 295.79), ICD-10: F20).

Defining phenotypes for infection load and psychiatric diagnoses

For specific psychiatric disorders, case (control) status was determined based on having (not having) the relevant ICD diagnosis as per the above codes. For any/overall psychiatric disorder, cases had any ICD code from: ICD-8: 290-315 and/or ICD-10: F00-F99, and controls did not have any of the codes from those ranges. For the quantitative phenotype of infection load, individuals who did not have any diagnosis from the above 14 infection categories received an infection load phenotype value of 0; for individuals with infection diagnoses, we counted the number of infection categories they had excluding the broad categories of bacterial, viral and other infections, and the total count was used as their infection load value. This approach was chosen because these broad categories include codes that are found in specific categories, and, moreover, they do not indicate specific infection sites. Individuals who were cases only for those 3 categories without any site-specific category (N=3,074) received a missing infection load value and were not used in analyses for this phenotype. The final sample used in this study therefore included 62,460 individuals for the infection load phenotype (NB: when using the binary overall
infection phenotype in some analyses, the full sample was used). The phenotype distribution figure was exported using Daniel’s XL Toolbox v7.3.4 (27).

**Statistical and epidemiological analyses**

Statistical analyses were performed in R (28) v.3.5.1. Logistic regressions of psychiatric diagnosis on infection load were performed with the *glm* function (family = binomial(link = "logit")) with covariates for age (in years), age squared and sex, resulting in two-sided p-values from a Wald test for the infection load coefficient’s being different from zero. Confidence intervals were calculated with the *confint* function. The regressions were performed in the random population sample (N=21,706 before exclusions; N=20,822 after exclusions for the infection load phenotype) to avoid potential biases and obtain population estimates, and controls for psychiatric disorders were “normal controls” *i.e.* they did not have the diagnosis which was being tested but they could have other (psychiatric) diagnoses. Note that the age covariates in this study were censored for a minority of individuals who emigrated (0.43%), died (1.02%) or lost contact with the Danish authorities (0.02%) by July 2013 (*i.e.* we set their age to what it was when their status changed, whereas, for other individuals, the age at the end of 2013 was used). These figures apply to the full sample. Tetrachoric correlations for overall infection and overall psychiatric diagnosis in the full sample and in the random population sample were calculated with the *polycor* R package v.0.8-1 (29).

**Genetic dataset and genome-wide association study for infection load**

We performed a discovery genome-wide association study (GWAS) for infection load, and GWASs for other phenotypes (overall psychiatric diagnosis; overall infection) for downstream heritability and genetic correlation analyses. The description of the genetic dataset used in this
study has appeared in several of our previous publications (11, 12), but we repeat it briefly here, with further information relevant to the present study: the marker dataset used in this GWAS has undergone several rounds of quality control (QC). The starting point for the QC was a dataset which had 78,050 samples genotyped in 23 of the original 25 waves. This dataset is described in detail elsewhere (30). Further QC of this dataset including a description of the procedures for sample and marker QC is provided in the first study which utilized it (23). Note, however, that that study had a minor allele frequency threshold (for dosage data) different from the one below. Before imputation of more markers, markers with rare alleles and non-autosomal markers were removed. The pre-imputation QC is described in more detail in other studies which used this sample (23, 24, 26). For the imputation, genotypes were phased with SHAPEIT3 (31) and the imputation was performed with IMPUTE2 (32). Following the imputation, markers were excluded if they had: an INFO score (calculated with QCTOOL) <0.2; a minor allele frequency (MAF) <0.001; best-guess genotypes missing in >10% of subjects, whereby the missingness in imputed genotypes was determined by treating individual genotypes with probability <0.9 as missing; Hardy-Weinberg equilibrium (HWE) P<1×10^{-6}, and/or a genome-wide significant association with the genotyping wave or with the imputation batch itself (in controls in a homogeneous European subset of the sample). Lastly, markers with differential missingness between psychiatric cases and controls (P<1×10^{-6}) were also removed. In this study we used only best-guess genotypes, or hard calls, and only genotypes with probability of 0.9 or above (i.e. the imputation resulted in a best-guess genotype with a probability of at least 0.9 for a given marker and a given individual) were retained (others were set to missing). Furthermore, markers were retained only if they had an INFO score (following the imputation) of at least 0.8 and MAF of at least 0.01 in the QCed sample (these steps were additional steps not included in some of the
previous studies mentioned (23, 24, 26)). Ultimately, 7,071,055 markers were used in the GWASs. The genome build for our dataset was hg19. The infection load discovery GWAS was performed with PLINK v.1.90b6.18 using a linear regression model (--linear) and included covariates for age, age squared, sex, the first ten principal components and overall psychiatric diagnosis status. The Manhattan plot and the QQ plot were generated with the “qqman” R scripts by Stephen Turner and Daniel Capurso (with the (major update) version from April 19, 2011 for the former plot and the version from June 10, 2013 for the latter, available from: https://github.com/stephenturner/qqman/blob/v0.0.0/qqman.r). Other GWASs (with logistic regression in the case of binary phenotypes, with --logistic) were performed for downstream analysis with LDSC, as described below. Those GWASs included covariates for age, age squared, sex, and the first ten principal components. A covariate for overall psychiatric diagnosis status was included in some analyses for the heritability estimates and not in others (as indicated in the Results section), and it was not included in GWASs used in downstream genetic correlation analyses. For the top markers, a post hoc association test was run in R using a Poisson regression with the glm function (family = poisson(link = "log")) and the same covariates as in the GWAS, as the infection load phenotype consisted of counts and was not normally distributed.

Heritability estimates and genetic correlations

Using the same dataset of markers and covariates as the ones used in the GWAS, we estimated the heritability of infection load in our full sample. This was achieved with GCTA (33) as well as LDSC (34, 35). For GCTA, the genetic relationship matrix was calculated for each autosomal chromosome separately with --make-grm and merged with --mgrm with GCTA v1.91.1 beta as previously described (14). The heritability of infection load was estimated with --reml in GCTA
v1.93.2 beta (covariates included age, age squared, sex, the first ten principal components and overall psychiatric diagnosis status). For LDSC, LD score files were generated using the QC-passing random population sample (from the marker dataset with best-guess genotype hard calls) updated to have genetic positions using the genetic map from 1000 Genomes phase 3, with a 1 cM window (\texttt{--ld-wind-cm 1}), as described previously (7). The summary statistics (PLINK output) from the GWAS were processed with the \texttt{munge\_sumstats.py} script with the default parameters (after adding A2 from the PLINK bim file and using the NMISS column as the N for LDSC), and LDSC (\texttt{ldsc.py}) v.1.0.1 was used (with the default parameters) in the estimation of the heritabilities (with \texttt{--h2}) and genetic correlations (with \texttt{--rg}). We used the same LD score dataset as both reference and regression (\texttt{--ref-ld-chr} and \texttt{--w-ld-chr}) datasets, as recommended in the LDSC tutorial for non-partitioned LD Score regression (https://github.com/bulik/ldsc/wiki/Heritability-and-Genetic-Correlation, version from Jul 13, 2017). In the iPSYCH datasets, 5,464,859 markers remained after processing with \texttt{munge\_sumstats.py}. We tested the heritability (h²) as being different from 0 using a Wald test and the χ² distribution with 1 degree of freedom (d.f.) in the following way: χ² = (h²/SE)², where h² is the heritability on the observed scale from LDSC, and SE is the standard error for the heritability, and p-values were calculated with \texttt{0.5*pchisq(\chi^2, df=1, lower.tail=F)} in R (we multiply by 0.5 since h² should be non-negative). For GCTA, the likelihood ratio test statistic from the GCTA output was used to derive p-values using the χ² distribution with one degree of freedom, as described above for the LDSC h² estimates. Genetic correlations were estimated only with LDSC, as it is robust to sample overlap for the studied traits and allowed the use of only summary statistics-level data from the replication sample. Note that when testing for genetic correlations between infection phenotypes and psychiatric disorder in iPSYCH, the summary
statistics for the infection phenotypes in iPSYCH were from a GWAS that did not include a covariate for overall psychiatric diagnosis status, as stated previously. For heritability estimation with LDSC, a covariate for overall psychiatric diagnosis status was included in the corresponding GWAS, or not, as indicated in the Results. All GWASs used in downstream LDSC analyses included covariates for age, age squared, sex and the first ten principal components. Genetic correlation ($r_g$) was tested for being different from 0 using a Wald test as well: $\chi^2 = (r_g/SE)^2$, where $r_g$ is the genetic correlation from LDSC, and SE is the standard error for the genetic correlation. P-values were calculated in R using `pchisq(\chi^2, df=1, lower.tail=F)`. Note that in the case of the heritabilities and genetic correlation for or between any infection and any psychiatric diagnosis, which we reported previously (7), we include in this study results which used the censored age covariates, so as to be in line with the present study’s methodology. For the binary traits, heritabilities were transformed to the liability scale (36) using the proportion of cases in each GWAS and a population prevalence estimate from the iPSYCH random population sample (0.1 for overall psychiatric diagnosis and 0.357 for overall infection).

**Replication sample**

To replicate the results from the present study and our previous study with a binary overall infection phenotype, we used summary statistics from FinnGen (37) Release 7 (309,154 individuals). We used two phenotypes which were the closest possible to our any/overall infection and any/overall psychiatric diagnosis phenotypes: Certain infectious and parasitic diseases (AB1_INFECTIONS, based on ICD-10 codes starting with A or B) and Any mental disorder (KRA_PSY_ANYMENTAL, based on ICD-10 codes F00-F03, F051, G30, F1-F9, or ICD-8/9 equivalent codes) with the following numbers of cases;controls: 86,892;222,262 and 76,073;233,081, respectively. The FinnGen summary statistics were processed in the following...
way: first, they were cross-referenced with the HRC reference panel (ftp://ngs.sanger.ac.uk/production/hrc/HRC.r1-1/HRC.r1-1.GRCh37.wgs.mac5.sites.tab.gz) based on marker ID. From that, the chromosome and position in genome build hg19 was added to the summary statistics. Based on chromosome and position in hg19, marker IDs were changed to corresponding marker IDs in the iPSYCH dataset, and only overlapping markers were retained. This made it possible to use the same LD score files and to have greater marker overlap between the iPSYCH and FinnGen datasets for the genetic correlation analyses. Duplicate markers and markers with mismatched alleles (identified using PRSice (38) v2.3.5) were removed from the summary statistics. The QCed summary statistics were then processed with munge_sumstats.py (as they did not contain sample sizes per marker, the total sample size as per the above numbers was used). After QC and processing, 4,887,803 markers were included in the FinnGen datasets. It should be noted that our LD score files were based on a homogenous Danish-Europeans sample; while there exist LD differences between Finns and Danes, studies have used the same (European) LD weights when analyzing both Finnish and Danish samples for genetic correlation estimation with LDSC (39). However, we also checked the heritabilities and genetic correlations within FinnGen using European LD scores (https://data.broadinstitute.org/alkesgroup/LDSCORE/eur_w_ld_chr.tar.bz2). In these analyses, we did not convert marker IDs, as both FinnGen datasets and the LD score files included SNP rsIDs. Preprocessing with munge_sumstats.py in this case included adding the total sample size as per the above; otherwise, the default thresholds we used. After processing, 12,809,821 markers in each dataset (AB1_INFECTIONS and KRA_PSY_ANYMENTAL) were retained (excluding markers with no rsID in the summary statistics files). The number of overlapping markers across the FinnGen datasets and the European LD score dataset was 1,169,856.
Heritabilities were transformed to the liability scale using the proportion of cases for each trait as reported by FinnGen, and the same values were used for the population prevalences (since FinnGen is a full population cohort).

**Post hoc analyses for genetic correlations**

We performed several post hoc analyses and sensitivity analyses to evaluate whether various factors affected the genetic correlation analyses and/or provide further insight into the genetic relationship between infection and mental illness. The first of these was a test for cross-trait assortative mating, which could influence the $r_g$ estimate. It was shown that, for a single trait, assortative mating would induce a correlation between a polygenic risk scores (PRSs) generated using only odd- or only even-numbered chromosomes (in the target sample) at a time, and that, under random mating, these scores should not be correlated (40). This approach was later extended to cross-trait assortative mating (41), whereby the PRS for trait 1 generated using only odd-numbered chromosomes is tested for correlation with the PRS for trait 2 generated using only even-numbered chromosomes, and vice versa. We used a simplified version of this procedure, testing for both correlations at both an inclusive p-value threshold ($p_T=1$) and the genome-wide significance threshold ($p_T=5 \times 10^{-8}$), with FinnGen summary statistics (as processed above for genetic correlations with iPSYCH) as the training dataset and iPSYCH as the target dataset. The PRSs were generated with PRSice v.2.3.5 with an $r^2$ threshold of 0.1 in a window of 250 kbp and using the --score sum method. Otherwise the default parameters were used. The correlation was tested with the `cor.test` function in R, using Pearson correlation as the method. The tests were two-sided.
To test for genetic causality between overall infection and overall psychiatric diagnosis in iPSYCH, we used the latent causal variable (LCV) model (42). This model tests whether trait 1 (in our case, overall infection) is causal of trait 2 (overall psychiatric diagnosis) using a latent causal variable which is assumed to mediate the genetic correlation between the two traits. If trait 1 is strongly genetically correlated with this variable, then it is partially genetically causal of trait 2. This relationship is quantified using the genetic causality proportion (GCP). GCP ranges from 0 (no genetic causality) to 1 (full genetic causality; if it is negative, it means that trait 2 is partially or fully genetically causal of trait 1. LCV uses LD scores and summary statistics (we used the same LD scores as used with LDSC, and the summary statistics processed with LDSC), as LDSC does, but it uses them in a modified procedure. We used both the full dataset of summary statistics, as well as markers with MAF\(\geq0.05\), as recommended on the LCV GitHub page. The scripts used were

https://github.com/lukejoconnor/LCV/blob/master/R/MomentFunctions.R and
https://github.com/lukejoconnor/LCV/blob/master/R/RunLCV.R. The versions used were from July 1 2020 for the first script and March 14 2019 for the last two scripts.

Lastly, we performed sensitivity analyses for the LDSC results after removing the major histocompatibility complex (MHC) region on chromosome 6 from the dataset used to generate the LD scores: we removed markers on chromosome 6 between coordinates 25,392,021 and 33,392,022, based on the coordinates from a GWAS protocol (43) converted to genome build hg19 using the UCSC Genome Browser LiftOver tool, and regenerated the LD scores with LDSC v1.0.1 using the dataset of the random population sample (as described earlier). The LDSC analyses were then repeated using these new LD scores. The numbers of markers included
in the $h^2$ and $r_g$ analyses after the removal of the MHC region were 5,424,093 for the iPSYCH datasets and 4,857,454 for the FinnGen datasets (when used with iPSYCH LD scores). Note that the external European LD score dataset did not include the MHC region to begin with.
RESULTS

In the full iPSYCH sample, 24,161 individuals had a psychiatric diagnosis but no infection diagnosis, 6,744 had an infection diagnosis but no psychiatric diagnosis, 21,728 had both, and 12,901 had neither. In the random population sample, 1,113 had a psychiatric diagnosis but no infection diagnosis, 6,693 had an infection diagnosis but no psychiatric diagnosis, 1,062 had both, and 12,838 had neither. The tetrachoric correlation between overall infection and overall psychiatric diagnosis was 0.2006 (standard error (SE)=0.0063) in the full sample and 0.1923 (SE =0.0143) in the random population sample. Sample sizes for psychiatric disorders and the binary infection phenotype are shown in Table 1. To examine the age relations between the two disease classes, we used individuals who had both diagnoses and defined the following variable: age at (first) infection category diagnosis minus age at (first) psychiatric diagnosis (both in years). This variable had a mean of -8.45, a median of -8.21 and a standard deviation of 8.17. In the random population sample, these were: -8.79, -8.89 and 8.24, respectively. This suggests that on average, the first severe infection was diagnosed before the first psychiatric diagnosis in our sample. For infection load, the counts of individuals per infection load value are shown in Table 2. Among individuals with an infection load of 0 in the full sample, the proportion of individuals with overall psychiatric diagnosis was 65%, which rose to 85% for individuals with an infection load of 5 or above. Among individuals with an infection load of 0 in the random population sample, the proportion of individuals with overall psychiatric diagnosis was 8%, which rose to 27% for individuals with an infection load of 5 or above. Supplementary Figure S1 shows the proportion of individuals with any psychiatric diagnosis across infection load values.

Comorbidity analyses for infection load and psychiatric disorders
Infection load was significantly associated with increased risk of all individual psychiatric disorders except for anorexia. It was especially associated with overall psychiatric diagnosis, with an odds ratio (OR) of 1.4578 (P=1.10×10^{-41}). Table 3 shows the results of the regression analyses.

**Heritabilities and genetic correlations**

Infection load showed a significant non-zero heritability, but it was not high, at h^2=0.0221 (SE=0.0055, P=2.93×10^{-5}). The heritability estimate for infection load from GCTA was similar: h^2=0.0432 (SE=0.0053, LRT statistic=77.298, P=7.35×10^{-19}). GCTA used a genomic relationship matrix which was based on more markers than included in the summary statistics used by LDSC, which could contribute to the difference between these estimates, in addition to the methodological differences between GCTA and LDSC. The binary overall infection and overall psychiatric diagnosis phenotypes in iPSYCH and the corresponding phenotypes in FinnGen were also significantly different from zero, with the psychiatric phenotypes showing higher heritabilities. Table 4 shows the results of all LDSC heritability analyses.

For the binary phenotypes, our genetic correlation analyses replicated our previous results. Namely, overall infection was highly correlated with overall psychiatric diagnosis within iPSYCH, within FinnGen, and across FinnGen and iPSYCH, with estimates mostly being higher than 0.4. For infection load, the genetic correlation with overall psychiatric diagnosis was slightly higher than that between overall infection and overall psychiatric diagnosis. For the corresponding phenotypes across iPSYCH and FinnGen, the genetic correlations were above 0.6. All estimates remained significant after Bonferroni correction for the number of genetic correlations estimated in this study. The full results are shown in Table 5.
Top results from the discovery GWAS for infection load

Our GWAS found 138 associations at the suggestive threshold (1×10^{-5}), which are shown in Supplementary Table S2. Figure 1 shows the Manhattan plot for the GWAS, and the corresponding QQ plot is shown in Supplementary Figure S2. The top association in our study was with marker rs12361013 on chromosome 11 (β=-0.0323, P=4.88×10^{-7}, effect allele: A, other allele: G), which was the leading SNP in the suggestive association peak on chromosome 11. However, there was also a second suggestive association peak on chromosome 9, with the leading SNP being rs35711908 (β=0.0242, P=5.33×10^{-7}, effect allele: A, other allele: G). These two markers manually tested in a Poisson regression in R, due to the fact that the phenotype consisted of count data and therefore might not be ideal for linear regression. This type of regression model is not implemented in common genetic association tools, but it may be more suitable for the infection load phenotype. For rs12361013, we obtained an estimate (beta) of -0.05848 for the A allele count, with P=7.64×10^{-8}. For rs35711908, we obtained an estimate of 0.04231 for the A allele count, with P=1.14×10^{-7}. Thus, at least for the top results, the linear regression analysis in PLINK underestimated the effects, but the differences were not large.

Tests for assortative mating and causal relationship between overall infection and overall psychiatric diagnosis, and sensitivity analyses for genetic correlations

All cross-trait PRS correlations between odd- and even-numbered chromosomes were close to zero and non-significant: for infection PRS for odd-numbered chromosomes and psychiatric diagnosis PRS for even-numbered chromosomes the correlations were -0.0018 (P=0.6388) and 0.0042 (P=0.2875) for pT=1 and pT=5×10^{-8}, respectively. For infection PRS for even-numbered
chromosomes and psychiatric diagnosis PRS for odd-numbered chromosomes the correlations were 0.0002 (P=0.9560) and 0.0044 (P=0.2643) for those two p-value thresholds.

The LCV analyses showed some evidence for a causal relationship between overall infection and overall psychiatric diagnosis with GCP=0.71 (P=6.41×10⁻⁵) with all markers and GCP=0.77 (P=4.65×10⁻⁸) using markers with MAF≥0.05. The LCV Z scores for the heritability of overall infection were 5.96 and 5.59, respectively. The h² Z scores for overall psychiatric diagnosis were >7.

Our LDSC sensitivity analyses after the removal of the MHC region obtained results that were similar to the original results in terms of the sizes of the correlations, with the average change across all h² estimates being ~0.001 and the average change across all rᵉ estimates being ~0.051. The full results can be found in the Supplementary Notes for this paper.
DISCUSSION

Our study identified strong epidemiological and genetic associations between infection load and psychiatric disorders. In both cases the association was positive, i.e., an odds ratio greater than 1 in a regression of a psychiatric diagnosis on infection load or a positive genetic correlation. While our previous study obtained higher odds ratios for the effect of the (binary) overall infection phenotype (7) on psychiatric diagnosis, in this study we report the increase in odds per infection category/site (Table 3). Thus, having been diagnosed with, for example, infections of 5 different categories would increase the odds of having a psychiatric diagnosis by a factor of 6.6.

In our sample, individuals tended to have been diagnosed with the (first) infection before the (first) psychiatric diagnosis. When looking at the quantitative phenotype of infection load, we observe a general positive association trend between infection load and proportion of psychiatric cases (Supplementary Figure S1). While the temporal relationship between the two diseases classes is obscured with the infection load phenotype, we know from the literature that some somatic diseases have a bidirectional relationship with psychiatric disorders, and that the latter are also associated with risk of somatic diseases even a decade later (44, 45). Infections can also have a bidirectional relationship with psychiatric disorders (46).

The heritability of infection load is similar to the heritability of overall infection on the liability scale, and, as expected, they are strongly correlated genetically. Infection load shows a slightly higher genetic correlation (than overall infection) with overall psychiatric diagnosis in iPSYCH but also a lower one with FinnGen; however, in all cases the genetic correlation is positive and

---

1 The results of our previous study do not change much after censoring the age for individuals who died, emigrated or lost contact with the Danish authorities, as employed in the present study. For example, for overall psychiatric diagnosis, before age-censoring, the OR for overall infection was 1.74 (95%CI: 1.59-1.90), and after age censoring it was 1.73 (95%CI: 1.58-1.90).
remains significant after Bonferroni correction. The genetic correlation between overall infection and overall psychiatric diagnosis replicated between iPSYCH and FinnGen (using infections in iPSYCH and psychiatric diagnosis in FinnGen and vice versa) and was about 0.4. It should be noted that the genetic correlations between the corresponding phenotypes across iPSYCH and FinnGen were not 1, but they were >0.6. This is not uncommon when the phenotypes were ascertained differently and come from different studies (e.g. as in the case of ASD in iPSYCH and in PGC (30)), and in our case there were differences in sample composition and (mostly for infections) differences in the included diagnoses. The highest genetic correlation between overall infection and overall psychiatric diagnosis was obtained when using FinnGen summary statistics for both phenotypes, with European LD scores. Combined, the results of all analyses suggest that there is a high degree of positive genetic correlation between susceptibility of severe infections and psychiatric disorders. We found no evidence that the genetic correlation between overall infection and overall psychiatric diagnosis was inflated by potential cross-trait assortative mating between susceptibility to infection and psychiatric disorders. We found some evidence that susceptibility to overall infection is partially genetically causal of overall psychiatric diagnosis using the LCV model. We note that LCV Z scores for the heritability of overall infection were below 7, the threshold LCV recommends; however, the LCV script specifies that it is “a very stringent threshold”. These results do not mean that specific psychiatric disorders are necessarily linked to infections via genetics in the same way as overall psychiatric diagnosis. For example, we have shown in a previous study that susceptibility to infection is not genetically correlated with autism spectrum disorder (ASD), but genetics does play a role in the interplay between infections and ASD (11). Namely, ASD cases with a history of maternal pregnancy-related infections were genetically different, as a group, from ASD cases without a history of maternal
infections (the genetic correlation between the two phenotypes was \( r_g = 0.3811, P = 0.0033 \) when testing against a null of 0, \( P = 1.89 \times 10^{-6} \) when testing against a null of 1). ASD cases with a history of maternal infections occurring more than 2 months following birth were not genetically distinct from the group of ASD cases without a history of maternal infections (i.e. their genetic correlation was not significantly different from 1). These estimates do not change much when the MHC region is excluded from the analysis (discussed in more detail below). As ASD cases form a large part of the iPSYCH case subset, this may suggest that the underlying factors in the high genetic correlation between overall infection and overall psychiatric diagnosis could be non-disorder-specific with regards to psychiatric disorders, or that they are driven by other disorders. Lastly, our LDSC analyses included markers in the MHC region. Most studies which use LDSC remove these due to the high level of LD in the MHC region, because such markers could be outliers in the regression model and influence the LDSC regression. However, in the iPSYCH GWASs in this study and our previous studies (7, 11), there was no strong signal in the MHC region (for infection phenotypes, this was true both for GWASs with and without a covariate for overall psychiatric diagnosis). The problem with LDSC and the MHC region having extremely strong associations pertains mostly to autoimmune diseases (47), and we did not have any such signal in our results. Also, given the relevance of the MHC to infection-related phenotypes, we decided to retain it in the analyses. However, we also provide results for the LDSC analyses repeated after removing the MHC region from the LD score datasets, for \( h^2 \) and \( r_g \) estimates reported in this study or relevant studies which employed the iPSYCH-generated LD scores. These can be found in the Supplementary Notes. The removal of the MHC region affected mostly analyses involving the FinnGen infections phenotype (when using the iPSYCH LD scores). This is likely due to the fact that, unlike the iPSYCH overall infection GWAS, the
FinnGen infections GWAS did obtain a genome-wide significant signal within the MHC region, and, therefore, the analyses without the MHC region could be more appropriate when this dataset is used. A previous study which used the iPSYCH sample examined genetic correlations between infections and specific psychiatric disorders, and positive genetic correlations were found between infections and anorexia, ADHD, bipolar disorder, depression and schizophrenia (8).

Our GWAS for infection load identified two suggestive association peaks. Interestingly, the top association from our GWAS for overall infection, with rs6447952, was not among the suggestive associations for infection load. In this context, one important point to consider is what our phenotype of infection load captures. By definition, all individuals with infection load >0 would be cases for our overall infection phenotype, and, therefore, we expect some overlap in genetic associations, as also reflected in the high genetic correlation (which, it should be noted, could be very high in part due to the low heritabilities of the individual phenotypes). However, a high infection load, as defined in our study, suggests a high predisposition to infection regardless of the site of infection. In other words, it may indicate a deeper cause of immune dysfunction. This is not the same as simply having an infection or even being prone to a specific type of infection: an individual with 20 skin infections of the same kind and no other infection would have an infection load of 1, whereas an individual with a skin infection, a gastrointestinal infection and a central nervous system infection would have an infection load of 3, and both individuals would have the same affection status for overall infection. This also makes biological sense: repeated infections of the same type may suggest a narrower genetic risk, or a primarily non-genetic risk factor e.g. personal hygiene in the case of skin infection, or diet in the case of gastrointestinal infection, whereas an individual with multiple infections at different sites of the body is likely to
have a more general immune deficit. Thus, there could be genetic factors that might be captured by one phenotype but not the other.

Our top GWAS association was with rs12361013, which is located in the Contactin 5 (CNTN5) gene on chromosome 11. Contactin 5 is an immunoglobulin cell adhesion molecule which is involved in neural development and has been implicated in ASD (48). It has also been implicated in inflammatory diseases such as gout (49) and in the response to anti-tumor necrosis factor alpha medication in Crohn’s disease patients (50). Moreover, a variant in CNTN5 showed a genome-wide significant association with secreted Interleukin-2 in response to vaccinia virus stimulation in individuals who had received a smallpox vaccine (51). The second highest association was with rs35711908 on chromosome 9. This variant showed genome-wide significant associations with several phenotypes in FinnGen (Release 7, accessed through the web portal on October 9 2022), including various forms of hypothyroidism and disorders of the thyroid gland, which, in turn, may be associated with infections (52, 53). On the GTEx web portal (Version 8, accessed on October 9 2022), rs35711908 is associated with the expression of the Proteasome 20S Subunit Beta 7 (PSMB7) and NIMA Related Kinase 6 (NEK6) genes in several tissues. PSMB7 encodes a subunit of the 20S proteasome, which is responsible for protein degradation; upon stimulation by pro-inflammatory cytokines, this subunit is down-regulated and replaced by one encoded by PSMB10, which forms part of the immunoproteosome, which is involved in MHC class I antigen presentation (54, 55). NEK6 is involved in mitosis and, if overexpressed, may result in human B cell lymphoma (56).

Interestingly, both PSMB7 and NEK6 interact or are predicted to interact with B-cell specific enhancers (56, 57). Thus, our top GWAS results do show some links to immune-related functions, but they are not genome-wide significant.
CONCLUSIONS

Our study identified positive associations between infection load and psychiatric disorders both epidemiologically, using the random population subset of the iPSYCH sample, and genetically, using the full iPSYCH sample as well as an external replication sample, FinnGen. We have also replicated our previous result for the genetic correlation between overall infection and overall psychiatric diagnosis using the FinnGen sample. Our results also suggest that infection susceptibility could have a causal genetic relationship with psychiatric disorders. Our GWAS for infection load identified suggestive associations that were linked to genes involved in neurodevelopmental or inflammatory diseases or immune-related phenotypes. Our study thus provides further support to the notion of links between the immune system and psychiatric disorders.
REFERENCES


(44) Wium-Andersen MK, Wium-Andersen IK, Prescott EIB, Overvad K, Jorgensen MB, Osler M. An attempt to explain the bidirectional association between ischaemic heart disease,


DECLARATIONS

Ethics approval and consent to participate

The Danish Scientific Ethics Committee approved this study (ESDH 1-10-72-287-12). The following institutions also approved the study: the Danish Health Data Authority, the Danish data protection agency and the Danish Neonatal Screening Biobank Steering. All personal information from the registers is anonymized when used for research purposes, according to Danish legislation; informed consent from participants was not required.

Availability of data and material

The iPSYCH initiative is committed to providing access to these data to the scientific community, in accordance with Danish law. Researchers may be granted access upon request to the iPSYCH management.

Competing interests

All researchers had full independence from the funders. The authors report no biomedical financial interests or potential conflicts of interest. TW states that he has acted as a lecturer and scientific counselor to H. Lundbeck A/S.

Funding

This study was funded by The Lundbeck Foundation, Denmark (grant numbers R268-2016-3925, R102-A9118 and R155-2014-1724), the Independent Research Fund Denmark (grant number 7025-00078B), the Mental Health Services Capital Region of Denmark, University of Copenhagen, Aarhus University and the University Hospital in Aarhus. The genotyping of the
iPSYCH samples was supported by grants from the Lundbeck Foundation, the Stanley
Foundation, the Simons Foundation (SFARI 311789), and NIMH (5U01MH094432-02).

Authors’ contributions

RN conceived and designed the study, performed the analyses, interpreted the results, wrote the
manuscript; MEB supervised the study, interpreted the results, revised the manuscript; DMH and
TW are the principal investigators of iPSYCH.

Acknowledgements

This research has been conducted using the Danish National Biobank resource supported by the
Novo Nordisk Foundation. The analyses in this study were performed on GenomeDK. The
iPSYCH Consortium was established through the work of six principal investigators in
Denmark: from Aarhus University: Preben B. Mortensen, Ole Mors, Anders D. Børglum; from
the University of Copenhagen, Mental Health Services of the Capital Region of Denmark and/or
Statens Serum Institut: Thomas Werge, Merete Nordentoft, David M. Hougaard. We want to
acknowledge the participants and investigators of the FinnGen study, from which we obtained
summary statistics for use in replication analyses in this study.
Table 1: Sample sizes for psychiatric disorders and overall infection phenotype

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Number of individuals in the full sample</th>
<th>Number of individuals in the random population sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full sample size</td>
<td>65,534</td>
<td>21,706</td>
</tr>
<tr>
<td>Overall infection</td>
<td>28,472</td>
<td>7,755</td>
</tr>
<tr>
<td>Overall psychiatric diagnosis</td>
<td>45,889</td>
<td>2,175</td>
</tr>
<tr>
<td>Autism spectrum disorder</td>
<td>12,331</td>
<td>277</td>
</tr>
<tr>
<td>Attention deficit/hyperactivity disorder</td>
<td>14,397</td>
<td>338</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>2,401</td>
<td>70</td>
</tr>
<tr>
<td>Bipolar disorder</td>
<td>1,391</td>
<td>36</td>
</tr>
<tr>
<td>Depression</td>
<td>18,511</td>
<td>435</td>
</tr>
<tr>
<td>Anorexia</td>
<td>2,551</td>
<td>50</td>
</tr>
</tbody>
</table>
Table 2: Distribution of infection load in the full sample and in the random population sample

<table>
<thead>
<tr>
<th>Infection load</th>
<th>Number of individuals in the full sample</th>
<th>Number of individuals in the random population sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>37,062</td>
<td>13,951</td>
</tr>
<tr>
<td>1</td>
<td>17,682</td>
<td>5,166</td>
</tr>
<tr>
<td>2</td>
<td>5,858</td>
<td>1,366</td>
</tr>
<tr>
<td>3</td>
<td>1,460</td>
<td>271</td>
</tr>
<tr>
<td>4</td>
<td>324</td>
<td>53</td>
</tr>
<tr>
<td>5-7*</td>
<td>74</td>
<td>15</td>
</tr>
</tbody>
</table>

* Due to the data information policy of the iPSYCH consortium, exact numbers that are lower than five (excluding zero) are not specified, which is why we combined the counts for infection load 5 to 7; this was done only for reporting purposes, and the actual analyses used the individual counts.
Table 3: Results of regressions of psychiatric disorder on infection load in the random population sample

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Estimate from the regression</th>
<th>SE (estimate)</th>
<th>P-value</th>
<th>OR</th>
<th>OR 95% CI lower bound</th>
<th>OR 95% CI upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autism spectrum disorder</td>
<td>0.3084</td>
<td>0.0750</td>
<td>3.94×10^{-05}</td>
<td>1.3613</td>
<td>1.1698</td>
<td>1.5703</td>
</tr>
<tr>
<td>Attention deficit/hyperactivity disorder</td>
<td>0.4250</td>
<td>0.0626</td>
<td>1.16×10^{-11}</td>
<td>1.5297</td>
<td>1.3488</td>
<td>1.7247</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>0.4309</td>
<td>0.1241</td>
<td>0.000519</td>
<td>1.5386</td>
<td>1.1890</td>
<td>1.9375</td>
</tr>
<tr>
<td>Bipolar disorder</td>
<td>0.4548</td>
<td>0.1684</td>
<td>0.006911</td>
<td>1.5758</td>
<td>1.1038</td>
<td>2.1444</td>
</tr>
<tr>
<td>Depression</td>
<td>0.4102</td>
<td>0.0530</td>
<td>9.96×10^{-13}</td>
<td>1.5072</td>
<td>1.3560</td>
<td>1.6694</td>
</tr>
<tr>
<td>Anorexia</td>
<td>0.1325</td>
<td>0.1915</td>
<td>0.488794</td>
<td>1.1417</td>
<td>0.7584</td>
<td>1.6132</td>
</tr>
<tr>
<td>Overall psychiatric diagnosis</td>
<td>0.3770</td>
<td>0.0279</td>
<td>1.10×10^{-41}</td>
<td>1.4578</td>
<td>1.3800</td>
<td>1.5393</td>
</tr>
</tbody>
</table>

OR: odds ratio; CI: confidence interval. The p-values are not corrected.
Table 4: Heritability estimates from LDSC

<table>
<thead>
<tr>
<th>Trait</th>
<th>Observed scale h²</th>
<th>SE</th>
<th>P-value</th>
<th>Liability scale h²</th>
</tr>
</thead>
<tbody>
<tr>
<td>iPSYCH infection load (with covariate for overall psychiatric diagnosis)</td>
<td>0.0221</td>
<td>0.0055</td>
<td>2.93×10⁻⁵</td>
<td>-</td>
</tr>
<tr>
<td>iPSYCH infection load (without covariate for overall psychiatric diagnosis)</td>
<td>0.0252</td>
<td>0.0055</td>
<td>2.30×10⁻⁶</td>
<td>-</td>
</tr>
<tr>
<td>iPSYCH overall infection (with covariate for overall psychiatric diagnosis)</td>
<td>0.0157</td>
<td>0.0048</td>
<td>0.005362</td>
<td>0.0242</td>
</tr>
<tr>
<td>iPSYCH overall infection (without covariate for overall psychiatric diagnosis)</td>
<td>0.0178</td>
<td>0.0048</td>
<td>0.0001043</td>
<td>0.0274</td>
</tr>
<tr>
<td>iPSYCH overall psychiatric diagnosis</td>
<td>0.0763</td>
<td>0.0061</td>
<td>3.37×10⁻⁶</td>
<td>0.0956</td>
</tr>
<tr>
<td>FinnGen infections (iPSYCH LD scores)</td>
<td>0.0118</td>
<td>0.0030</td>
<td>4.19×10⁻⁵</td>
<td>0.0210</td>
</tr>
<tr>
<td>FinnGen infections (European LD scores)</td>
<td>0.0159</td>
<td>0.0020</td>
<td>9.33×10⁻¹⁶</td>
<td>0.0283</td>
</tr>
<tr>
<td>FinnGen any mental disorder (iPSYCH LD scores)</td>
<td>0.0260</td>
<td>0.0019</td>
<td>6.31×10⁻⁴³</td>
<td>0.0486</td>
</tr>
<tr>
<td>FinnGen any mental disorder (European LD scores)</td>
<td>0.0437</td>
<td>0.0028</td>
<td>3.25×10⁻⁵⁵</td>
<td>0.0817</td>
</tr>
</tbody>
</table>

The p-values are not corrected. “Liability scale h²” is relevant only to binary traits; in all cases the h² in the “observed scale h²” column is the one reported by LDSC. SE: standard error.
Table 5: Genetic correlation estimates from LDSC

<table>
<thead>
<tr>
<th>Trait 1</th>
<th>Trait 2</th>
<th>r&lt;sub&gt;g&lt;/sub&gt;</th>
<th>SE</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>iPSYCH overall infection</td>
<td>iPSYCH overall psychiatric diagnosis</td>
<td>0.4061</td>
<td>0.1049</td>
<td>0.000108</td>
</tr>
<tr>
<td>iPSYCH overall infection</td>
<td>FinnGen any mental disorder</td>
<td>0.4131</td>
<td>0.0956</td>
<td>1.55×10&lt;sup&gt;-5&lt;/sup&gt;</td>
</tr>
<tr>
<td>iPSYCH overall infection</td>
<td>FinnGen infections</td>
<td>0.6189</td>
<td>0.1394</td>
<td>9.01×10&lt;sup&gt;-6&lt;/sup&gt;</td>
</tr>
<tr>
<td>iPSYCH overall psychiatric diagnosis</td>
<td>FinnGen any mental disorder</td>
<td>0.6660</td>
<td>0.0444</td>
<td>7.34×10&lt;sup&gt;-51&lt;/sup&gt;</td>
</tr>
<tr>
<td>iPSYCH overall psychiatric diagnosis</td>
<td>FinnGen infections</td>
<td>0.4062</td>
<td>0.1118</td>
<td>0.000280</td>
</tr>
<tr>
<td>iPSYCH infection load</td>
<td>iPSYCH overall psychiatric diagnosis</td>
<td>0.4298</td>
<td>0.0901</td>
<td>1.84×10&lt;sup&gt;-6&lt;/sup&gt;</td>
</tr>
<tr>
<td>iPSYCH infection load</td>
<td>FinnGen infections</td>
<td>0.6347</td>
<td>0.1279</td>
<td>6.96×10&lt;sup&gt;-7&lt;/sup&gt;</td>
</tr>
<tr>
<td>iPSYCH infection load</td>
<td>FinnGen any mental disorder</td>
<td>0.3388</td>
<td>0.0733</td>
<td>3.80×10&lt;sup&gt;-6&lt;/sup&gt;</td>
</tr>
<tr>
<td>iPSYCH infection load</td>
<td>iPSYCH overall infection</td>
<td>0.9837</td>
<td>0.0364</td>
<td>7.57×10&lt;sup&gt;-161&lt;/sup&gt;</td>
</tr>
<tr>
<td>FinnGen infections</td>
<td>FinnGen any mental disorder</td>
<td>0.5603</td>
<td>0.1110</td>
<td>4.47×10&lt;sup&gt;-6&lt;/sup&gt;</td>
</tr>
<tr>
<td>(iPSYCH LD scores)</td>
<td>(iPSYCH LD scores)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FinnGen infections</td>
<td>FinnGen any mental disorder</td>
<td>0.6465</td>
<td>0.0555</td>
<td>2.33×10&lt;sup&gt;-31&lt;/sup&gt;</td>
</tr>
<tr>
<td>(European LD scores)</td>
<td>(European LD scores)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

When at least one of the traits was measured in iPSYCH, iPSYCH-generated LD scores were used. The p-values are not corrected. SE: standard error.
FIGURE LEGENDS

Figure 1: Manhattan plot for the GWAS for infection load. The blue line represents the threshold for suggestive association (P=1×10^-5), and the red line represents the threshold for genome-wide significance (P=5×10^-8).
ADDITIONAL FILES

Supplementary Table S1: ICD-8 and ICD-10 codes for infection categories.

Supplementary Table S2: Suggestive association from the infection load discovery GWAS.

Supplementary Figure S1: Distributions of the infection load phenotype and proportion of psychiatric cases.

Supplementary Figure S2: QQ plot for the infection load GWAS.

Supplementary Notes: Results for LDSC analyses without the MHC region for analyses using the iPSYCH-generated LD scores.