Cross-disorder analysis of schizophrenia and 19 immune-mediated diseases identifies shared genetic risk

Jennie G Pouget[1,2,3,*], Schizophrenia Working Group of the Psychiatric Genomics Consortium[‡], Buhm Han[4], Yang Wu[5], Emmanuel Mignot[6], Hanna M Ollila[6,7,8], Jonathan Barker[9,10], Sarah Spain[9,11], Nick Dand[9], Richard Trembath[9,12], Javier Martin[13], Maureen D Mayes[14], Lara Bossini-Castillo[11], Elena López-Isac[13], Ying Jin[15,16], Stephanie A Santorico[15,17,18], Richard A Spritz[15,16], Hakon Hakonarson[19], Constantin Polychronakos[20], Soumya Raychaudhuri[21,22,23,24,25,26,†], Jo Knight[27,†]

[1]Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada

[2]Institute of Medical Sciences, University of Toronto, Toronto, ON, Canada

[3]Department of Psychiatry, University of Toronto, Toronto, ON, Canada

[4]Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Republic of Korea

[5]Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, Australia

[6]Center for Sleep Sciences and Medicine, Department of Psychiatry and Behavioral Sciences, Stanford University, School of Medicine, Palo Alto, CA, USA

[7]Finnish Institute for Molecular Medicine, Helsinki, Finland

[8]Center for Genomic Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA and Broad Institute, Cambridge, Massachusetts, USA

[9]School of Basic and Medical Biosciences, Faculty of Life Sciences and Medicine, King's College London, UK

[10]St. John's Institute of Dermatology, Faculty of Life Sciences and Medicine, King's College London, London, UK

[11]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK

[12]Queen Mary University of London, Barts and the London School of Medicine and Dentistry, London, UK

[13]Institute of Parasitology and Biomedicine López-Neyra, IPBLN-CSIC, Granada, Spain

[14]The University of Texas Health Science Center–Houston, Houston, TX, USA

[15]Human Medical Genetics and Genomics Program

[16]Department of Pediatrics, University of Colorado School of Medicine, Aurora, CO, USA

[17]Department of Mathematical and Statistical Sciences, University of Colorado Denver, Denver, CO, USA

[18]Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado, Aurora, CO, USA

[19]Center for Applied Genomics, Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA

[20]Endocrine Genetics Laboratory, Department of Pediatrics and the Child Health Program of the Research Institute, McGill University Health Centre, Montreal, QC, Canada

[21]Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

[22]Partners HealthCare Center for Personalized Genetic Medicine, Boston, MA, USA

[23]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

[24]Division of Genetics

[25]Division of Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

[26]Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK

[27]Lancaster Medical School and Data Science Institute, Lancaster University, Lancaster, UK.

*Corresponding author. Email: jennie.pouget@camh.ca (JP); P: 416-535-8501; F: 416-979-4666

[†]These authors contributed equally to this work

[‡]Membership of the Schizophrenia Working Group of the Psychiatric Genomics Consortium is provided in the **Supporting Information**

Abstract

Many immune diseases occur at different rates among people with schizophrenia compared to the general population. Here, we evaluated whether this phenomenon might be explained by shared genetic risk factors. We used data from large genome-wide association studies to compare the genetic architecture of schizophrenia to 19 immune diseases. First, we evaluated the association with schizophrenia of 581 variants previously reported to be associated with immune diseases at genome-wide significance. We identified five variants with potentially pleiotropic effects. While colocalization analyses were inconclusive, functional characterization of these variants provided the strongest evidence for a model in which genetic variation at rs1734907 modulates risk of schizophrenia and Crohn's disease via altered methylation and expression of *EPHB4* – a gene whose protein product guides the migration of neuronal axons in the brain and the migration of lymphocytes towards infected cells in the immune system. Next, we investigated genome-wide sharing of common variants between schizophrenia and immune diseases using cross-trait LD Score regression. Of the 11 immune diseases with available genome-wide summary statistics, we observed genetic correlation between six immune diseases and schizophrenia: inflammatory bowel disease (r_g=0.12±0.03, p=2.49x10^{-4}), Crohn's disease (r_g=0.097±0.06, p=3.27x10^{-3}), ulcerative colitis (r_g=0.11±0.04, p=4.05x10^{-3}), primary biliary cirrhosis (r_g=0.13±0.05, p=3.98x10^{-3}), psoriasis (r_g=0.18±0.07, p=7.78x10^{-3}), and systemic lupus erythematosus (r_g=0.13±0.05, p=3.76x10^{-3}). With the exception of ulcerative colitis, the degree and direction of these genetic correlations were consistent with the expected phenotypic correlation based on epidemiological data. Our findings suggest shared genetic risk factors contribute to the epidemiological association of certain immune diseases and schizophrenia.

Introduction

Despite recent advances in identifying key biomarkers and genetic loci for schizophrenia, its pathophysiology remains poorly understood (1, 2). One interesting epidemiological observation is that the risk of developing many immune-mediated diseases is increased among patients with schizophrenia (3–5), and vice versa (6, 7). Here, we use the term **immune disease** to broadly encompass both autoimmune and inflammatory disorders. While there are discrepancies among studies regarding which immune diseases are most strongly correlated with schizophrenia, there is converging evidence that these diseases co-occur at a greater rate than is expected by chance (3–7). A notable exception is rheumatoid arthritis (RA), where a consistent inverse association with schizophrenia has been observed (5, 8).

Genetic factors have long been proposed as an explanation for the differing prevalence of immune diseases among patients with schizophrenia compared to the general population (5, 6). The recently reported role of *complement component 4* (*C4*) variation in schizophrenia (9) illustrates a potential shared genetic mechanism in the development of immune and psychiatric disorders. Genetic variants conferring increased *C4* expression protect against developing systemic lupus erythematosus (SLE), possibly by increased tagging of apoptotic cells – which are the trigger for autoantibody development in SLE – leading to more effective clearance by macrophages (10). The same genetic mechanism may increase the risk of developing schizophrenia, by increased tagging of neuronal synapses for elimination by microglia leading to excessive synaptic pruning (9). We hypothesize that similar shared genetic mechanisms may occur throughout the genome, with cellular manifestations in immune cells and neurons influencing the development of immune and psychiatric disorders, respectively. Previously, we found that susceptibility to schizophrenia does not appear to be driven by the broad set of loci harboring immune genes (11). However, not all genetic variants conferring risk of immune disease fall within immune loci. Here, we evaluated whether common genetic variants influencing the risk of 19 different immune diseases may also be involved in schizophrenia.

Our cross-disorder genetic approach is supported by recent successes in identifying shared genetic risk variants (**pleiotropy**) across a variety of human diseases (12–18). The biological interpretation of pleiotropy is challenging, given the various molecular mechanisms that can drive shared genetic risk variants across complex traits. For instance, pleiotropy can result from a single-nucleotide polymorphism (SNP) independently influencing two unrelated traits (horizontal pleiotropy), a SNP influencing one trait which is related to or a risk factor for additional traits (vertical pleiotropy), or a SNP influencing one trait which has a high rate of misclassification with or represents a subgroup of a second trait (clinical heterogeneity) (19, 20). Here, we use the term **pleiotropy** to refer broadly to a single genetic variant affecting multiple traits, regardless of the underlying molecular basis and biological implications. Pleiotropy in this broad sense is emerging as a pervasive phenomenon in the human genome (21–23), and cross-disorder studies characterizing the nature of genotype-phenotype relationships have the potential to yield significant insights into disease etiology. For instance, cross-trait genetic analyses have shed new light on cardiovascular disease and lipid biology – and shifted attention away from HDL as a potential treatment target – by demonstrating that genetically increased HDL cholesterol levels do not reduce the risk of myocardial infarction (14). In psychiatry, cross-disorder analyses have identified pleiotropic variants between schizophrenia, bipolar disorder, and major depressive disorder, indicating that these diseases are not as distinct at a pathophysiological level as current diagnostic criteria suggest (12, 13, 24).

While previous studies have investigated genome-wide pleiotropy between schizophrenia and immune disorders, results have been inconsistent (**S1 Table**). Genetic correlation has been reported between schizophrenia and Crohn's disease (25–29), multiple sclerosis (30), primary biliary cirrhosis (27), psoriasis (31), rheumatoid arthritis (25, 26), systemic lupus erythematosus (26, 27), and type 1 diabetes (25, 26, 28, 29) in some studies, but not in others (8, 13, 16, 26, 32). Interestingly, negative genetic correlation (whereby genetic risk protects against developing schizophrenia) has also been reported for RA (33), in keeping with the inverse epidemiological association (5, 8). Potential explanations for the differences across studies include differences in statistical power due to diverse methodologies and sample sizes (**S1 Table**), and varying degrees of influence by confounding variables

such as population stratification and linkage disequilibrium. Furthermore, although immune diseases have a significant sex bias with women at greater risk overall (46), potential sex-specific effects have not been explored in cross-trait analyses to date. If sex-specific effects are present, differences in the proportion of male:female samples across genome-wide association studies (GWAS) may also contribute to some of the differences across studies.

Additional studies are needed to reconcile the inconsistencies in existing cross-trait analyses of schizophrenia and immune disorders, with careful attention towards potential confounding variables (e.g. population stratification, linkage disequilibrium, non-independence of GWAS samples, and sex-specific effects). To this end we have performed a comprehensive cross-disorder analysis of schizophrenia and 19 immune diseases, using data from large genetic studies in European samples. Our findings add to a growing body of literature supporting pervasive pleiotropy between schizophrenia and immune diseases. We extend existing literature by including 10 immune diseases that have not previously been compared with schizophrenia, prioritizing pleiotropic genes through integrative analyses of multi-omics data, and estimating how much of the phenotypic correlation between schizophrenia and immune diseases was explained by the genetic correlations we observed.

Results

Defining immune risk variants

We identified immune-mediated diseases with robust GWAS findings using ImmunoBase (http://www.immunobase.org; accessed 7 June 2015), an online resource providing curated GWAS data for immune-related human diseases. These included the following 19 diseases: alopecia areata (AA), ankylosing spondylitis (AS), autoimmune thyroid disease (ATD), celiac disease (CEL), Crohn's disease (CRO), inflammatory bowel disease (IBD), juvenile idiopathic arthritis (JIA), multiple sclerosis (MS), narcolepsy (NAR), primary biliary cirrhosis (PBC), primary sclerosing cholangitis (PSC), psoriasis (PSO), rheumatoid arthritis (RA), Sjögren's syndrome (SJO), systemic lupus erythematosus (SLE),

systemic sclerosis (SSC), type 1 diabetes (T1D), ulcerative colitis (UC), and vitiligo (VIT). Notably, the majority of IBD risk variants were also risk variants for CRO and/or UC. For 11 of these immune diseases (see **Table 1**), we also obtained full GWAS summary statistics allowing us to conduct additional cross-trait Linkage Disequilibrium Score regression (LDSC) analyses (16).

Given that human leukocyte antigen (HLA) alleles within the major histocompatibility complex (MHC) region (chromosome 6: 25-34 Mb) account for a significant proportion of heritability of immune and inflammatory disorders (34), we considered HLA and non-HLA risk variants separately in our analyses. Within the MHC region we considered only the most strongly associated HLA variant (including SNPs, imputed HLA amino acid sites, and classical alleles) for each disease based on univariate analysis in previously published studies (see **Table 2**), because multivariate conditional analyses reporting adjusted effect sizes of independent HLA variants were not available for all immune diseases. Outside of the MHC region, we considered all non-HLA variants curated in ImmunoBase for each of the 19 immune diseases.

The number of genome-wide significant non-HLA risk loci for each of the 19 immune diseases varied from three (NAR) to 144 (IBD). Several variants were associated with more than one immune disease. In total we identified 581 unique variants (563 non-HLA variants and 18 HLA variants) associated with any immune disease at genome-wide significance. We refer to these variants as **immune risk variants**. For the complete list of non-HLA immune risk variants, see **S2 Table**.

Identifying pleiotropic variants implicated in both immune disease and schizophrenia

First, we evaluated whether there was any evidence of overall risk allele sharing between each of the 19 immune diseases and schizophrenia using a binomial sign test. To do this, we used previously published findings from a GWAS conducted by the Schizophrenia Working Group of the Psychiatric Genomics Consortium (1, 11). This GWAS represented a meta-analysis of 52 cohorts, comprising a total of 35,476 cases and 46,839 controls, and the full dataset is referred to here as the **PGC2 study**. Overall, the direction of effect for the sets of non-HLA SNPs associated with each of the 19 immune diseases at

genome-wide significance was not shared with schizophrenia more than expected by chance (all binomial sign test p>0.05, **S1 Fig**). Thus, we did not observe evidence of risk allele sharing between any immune disease and schizophrenia when using a stringent genome-wide significance threshold to define immune risk variants. We also evaluated the collective association of 261 LD-independent, non-HLA immune risk variants associated with at least one of the 19 immune-mediated diseases, for which linkage disequilibrium (LD) Score and minor allele frequency (MAF) information were available in the European LD Score database (16). We found significant deviation from the theoretical null in schizophrenia for immune risk SNPs (λ=1.46). However, when we compared the association of immune risk SNPs to that of similar randomly selected SNP sets (**Supplementary Methods**) we observed no evidence of enrichment (p=0.66, **S2 Fig**), indicating that immune risk SNPs were not associated with schizophrenia more than expected by chance given the polygenic nature of schizophrenia.

Next, we identified potential pleiotropic variants by evaluating the association of individual immune risk variants with schizophrenia. We considered SNPs associated with schizophrenia at $p<8.6x10^{-5}$ (Bonferroni correction for 581 tests, 563 non-HLA and 18 HLA variants) to have pleiotropic effects. Given the size of the schizophrenia GWAS, we had over 80% power to detect pleiotropic SNPs assuming an OR≥1.12 in schizophrenia.

Within the MHC region, we observed four HLA risk alleles associated with both immune disease and schizophrenia, particularly in the class II HLA region (**Table 2, S3 Fig**). These HLA risk alleles were the strongest MHC region associations for AA (HLA-DRB1 #37 Asn), CEL (HLA-DQB1 #74 Ala), PSC (HLA-B*08:01), and SJO (HLA-DQB1*02:01). The presence of HLA-DRB1 #37 Asn conferred a protective association in both AA and schizophrenia, but the remaining HLA variants showed the opposite direction of effect in schizophrenia compared to immune disease (**Table 2**, **S3 Fig**). Notably, none of these four HLA variants were significantly associated with schizophrenia in previous conditional analyses (9, 11), suggesting that their association with schizophrenia may be driven by LD with other causal variants in the region rather than true pleiotropy. Thus, we did not focus additional analyses on these variants.

Outside of the MHC region, five immune risk variants showed potential pleiotropic effects, with the risk allele for immune disease also conferring risk for schizophrenia (for regional association plots in schizophrenia and the immune diseases of interest, see **S4 Fig**). These variants have been previously implicated in CRO (rs6738825, rs13126505, rs1734907), MS (rs7132277), and CEL (rs296547). To evaluate the pleiotropic potential of these non-HLA variants, we performed Bayesian colocalization analysis (35) to estimate the posterior probability (PP) that the corresponding regional associations were driven by the same SNP in both schizophrenia and the immune disease of interest. We used available summary statistics from the PGC2 GWAS and the GWAS reporting these immune risk variants as genome-wide significant in CRO (36, 37), MS (38), and CEL (39). Notably, the immune disease GWAS summary statistics were not imputed (total number of SNPs: 953,241 (36), 25,075 (37), 155,756 (38), and 523,398 (39)). Low density SNP coverage across the genome results in a loss of statistical power to detect colocalization, as the true causal variants are unlikely to be genotyped (35). We observed statistically significant colocalization (PP>0.50) in the regions of association on chromosomes 2, 4, and 12. Only the region on chromosome 2 colocalized to the immune SNP of interest, rs6738825 (PP=0.23, **Table 3**); this region also colocalized to rs7587251 (PP=0.25), a SNP in high LD with rs6738825 (r^2=0.98 in 1000 Genomes Phase 3 CEU Population (40)). We did not observe statistically significant colocalization in the region of association on chromosome 1 (PP=0.06) or chromosome 7 (PP=0.05, **Table 3**). Given that our colocalization analyses were susceptible to type II error for the reasons discussed above, we cannot exclude the possibility of a shared variant underlying the associations seen in schizophrenia and immune diseases for these regions.

Next, we used conditional and joint analysis (COJO) (41) to perform association analyses in the PGC2 schizophrenia GWAS conditioning on each of the five immune risk variants (**S5 Fig**). If the immune risk variants (or SNPs in high LD with them) accounted for the regional associations seen in schizophrenia, no significant associations should remain after conditioning on these variants (statistically, all $p>8.6x10^{-5}$). We observed no remaining associations with schizophrenia after conditioning on rs296547 (top SNP after conditioning: rs111530734, $p=1.19x10^{-3}$), rs1734907 (top SNP after

conditioning: rs11768688, $p=9.79x10^{-4}$), and rs13126505 (top SNP after conditioning: rs112786981, $p=4.58x10^{-4}$). Significant associations with schizophrenia remained after conditioning on rs6738825 (top SNP after conditioning: rs111744017, $p=8.03x10^{-6}$) and rs7132277 (top SNP after conditioning: rs74240770, $p=1.37x10^{-8}$), suggesting there may be additional independent causal variants contributing to the associations in these regions for schizophrenia.

In order to prioritize genes underlying the five potentially pleiotropic SNPs, we performed an integrative analysis of GWAS summary statistics with methylation quantitative trait loci (mQTL) and expression quantitative trait loci (eQTL) studies using SMR and HEIDI (42, 43) (**Materials and Methods**). Briefly, we obtained summary-level mQTL and eQTL SNP data described in Wu et al. (40). The mQTL data were from 1,980 individuals with DNA methylation measured in peripheral blood (44, 45), and and eQTL data were from 2,765 individuals with gene expression levels measured in peripheral blood (82). Notably, rs296547 was not genotyped in the eQTL dataset, and we used rs404339 as a proxy SNP ($r^2=0.85$ in 1000 Genomes Phase 3 CEU Population (40)) in SMR analyses of gene expression analyses for rs296547.

We observed that rs1734907, a SNP associated with both Crohn's disease and schizophrenia, was an mQTL ($\beta=0.47$, $p=2.13x10^{-26}$) and eQTL ($\beta=-0.24$, $p=3.54x10^{-10}$) for *EPHB4* in peripheral blood (**S3 Table, Fig 1**). rs1734907 showed consistent pleiotropic associations with schizophrenia and *EPHB4* DNAm ($\beta_{SMR}=-0.14$, $p_{SMR}=3.58x10^{-5}$, $p_{HEIDI}=0.12$), schizophrenia and *EPHB4* expression ($\beta_{SMR}=-0.28$, $p_{SMR}=2.63x10^{-4}$, $p_{HEIDI}=0.17$), and *EPHB4* DNAm and *EPHB4* expression ($\beta_{SMR}=1.98$, $p_{SMR}=6.56x10^{-8}$, $p_{HEIDI}=0.011$). These consistent associations across molecular phenotypes and schizophrenia at the *EPHB4* locus suggest *EPHB4* may be driving the association of rs1734907 in schizophrenia (**Fig 1**). Notably,*TRIP6* is also a candidate functional gene underlying the association of rs1734907 with schizophrenia. We observed pleiotropic association for rs1734907 with schizophrenia and *TRIP6* DNAm with inconsistent direction of effect ($\beta_{SMR}=0.15$, $p_{SMR}=5.00x10^{-5}$, $p_{HEIDI}=0.17$ for probe cg18683606; $\beta_{SMR}=-0.12$, $p_{SMR}=2.32x10^{-5}$, $p_{HEIDI}=0.18$ for probe cg27396824), a trend for association with

schizophrenia and *TRIP6* expression (β_{SMR}=-0.33, p_{SMR}=6.38x10^{-4}, p_{HEIDI}=0.14), but no significant association with *TRIP6* DNAm and *TRIP6* expression. Colocalization analyses were inconclusive with respect to rs1734907 showing pleiotropic association with schizophrenia and Crohn's disease as described above, while conditional analyses suggested this variant explained the regional association seen on chromosome 7 in schizophrenia (**Table 3**).

We also observed that rs7132277, a SNP associated with both multiple sclerosis and schizophrenia (**Table 3**), was an mQTL (β=0.27, p=2.87x10^{-11}) and eQTL (β=0.32, p=5.23x10^{-19}) for *ABCB9* in peripheral blood (**S3 Table**). Furthermore, we observed consistent pleiotropic associations for rs7132277 with schizophrenia and *ABCB9* DNAm (β_{SMR}=-0.24, p_{SMR}=1.20x10^{-4}, p_{HEIDI}=0.55), schizophrenia and *ABCB9* expression (β_{SMR}=0.20, p_{SMR}=3.10x10^{-5}, p_{HEIDI}=0.17), and *ABCB9* DNAm and *ABCB9* expression (β_{SMR}=-0.83, p_{SMR}=9.81x10^{-8}, p_{HEIDI}=0.48). Thus, there was consistent association across molecular phenotypes and schizophrenia at the *ABCB9* locus, suggesting this gene may be driving the association of rs1734907 in schizophrenia. Although the region of association on chromosome 12 showed statistically significant evidence of colocalization in schizophrenia and multiple sclerosis, rs7132277 was not the pleiotropic variant driving this association (**Table 3**). Thus, our analyses highlight *ABCB9* as a candidate gene underlying the association of rs7132277 in schizophrenia, but do not implicate this SNP as a pleiotropic immune variant.

The other potentially pleiotropic SNPs did not demonstrate consistent localization to a particular gene across traits and molecular phenotypes (**Table 3, S3 Table**).

Detecting genetic correlations between immune disease and schizophrenia

Our immune risk variant set captured only those variants associated with immune diseases at genome-wide significance in current GWAS. Given the polygenicity of immune-related diseases, there are 100s to 1,000s of additional variants associated with each disease which have not yet been identified

(46). To evaluate sharing of risk alleles between immune diseases and schizophrenia using a broader set of variants, we used LDSC (16).

LDSC provides an interpretable and comparable estimation of genetic sharing between two traits in the form of genetic correlation (r_g) values; the method uses genome-wide summary statistics, effectively accounts for linkage-disequilibrium, and is robust to non-independence of GWAS samples (16). We therefore used LDSC to estimate pairwise genome-wide genetic correlations between schizophrenia and immune diseases. In addition to the 11 immune diseases with available genome-wide summary statistics, we included bipolar disorder as a positive control and height as a negative control. We used summary statistics from the 49 European-ancestry cohorts in the PGC2 study (31,335 cases and 38,765 controls) to ensure comparable LD structure across samples (1). Notably, LDSC is less sensitive than other methods of estimating genetic correlation (e.g.polygenic risk scoring, GREML), and is not robust when applied to genetic data obtained from specialty chips (e.g. Immunochip) (16). We considered immune diseases with r_g $p<0.05$ to show genetic overlap with schizophrenia.

As previously reported (16), our positive control (bipolar disorder) showed significant genetic overlap with schizophrenia ($r_g=0.75\pm0.05$, $p=8.5\times10^{-60}$; **Fig 2, Table 4**) and our negative control (height) showed no such overlap ($r_g=-0.004\pm0.02$, $p=0.84$; **Fig 2**, **Table 4**). With respect to immune diseases, we observed genetic overlap with schizophrenia for CRO, IBD, PBC, PSO, SLE, and UC ($r_g=0.10$-0.18, **Fig 2**, **Table 4**). Notably, genetic correlations for PSO and UC did not survive correction for the 11 tests performed (**Table 4**). Unsurprisingly, the genetic correlations between schizophrenia and immune diseases were smaller in magnitude than those of commonly overlapping immune diseases (e.g. RA and SLE: $r_g=0.55\pm0.08$, $p=3.60\times10^{-11}$).

Given the significant sex bias of immune diseases, with women at greater risk overall (46), we hypothesized that there may be sex-dependent genetic overlap between schizophrenia and some immune-mediated diseases. We therefore performed exploratory sex-stratified LDSC, estimating genetic correlation between immune diseases and schizophrenia separately in males (33,097 schizophrenia cases

and 35,190 controls) and females (17,760 schizophrenia cases and 36,903 controls) of European ancestry from the PGC2 study and additional PGC samples. We found consistent genetic correlation estimates across male and female subsamples compared to the total sample for nine of the 11 immune diseases investigated (**Table 4**). For SLE we observed significant genetic correlation with schizophrenia in the total sample and in the male sample, but not in the female schizophrenia sample. For SSC we observed genetic correlation with schizophrenia only in the male schizophrenia sample. For both SLE and SSC, there was a trend for the same direction of effect (positive genetic correlation) among females which did not reach statistical significance.

Benchmarking genetic correlations between immune disease and schizophrenia with epidemiological data

To determine how much of the phenotypic correlation between schizophrenia and immune-mediated diseases was explained by the genetic correlations we observed, we benchmarked significant genetic correlations between schizophrenia and immune-mediated disorders relative to the expected phenotypic correlations from epidemiological data (**Materials and Methods**). Using incidence of immune diseases in schizophrenia reported in a large population-based study (3), we estimated phenotypic correlations between schizophrenia and PBC, PSO, SLE, and UC. For PBC, PSO, and SLE we observed small positive genetic correlations with schizophrenia that were consistent with the epidemiological data (**Table 4**). For UC we observed a small positive estimate of genetic correlation (r_g=0.106 ± 0.04) while there was no strong evidence for any correlation between UC and schizophrenia in the epidemiological data (r_p=-0.001). Importantly, while the MHC region contains risk factors for both schizophrenia and immune-mediated diseases, our genetic correlation estimates were obtained considering only SNPs outside of the MHC region due to unusual LD in this region (47).

Discussion

Using a variety of statistical approaches, we provide evidence of shared genetic risk for schizophrenia and immune diseases. Within the MHC region, we identified four HLA variants showing statistically significant association with schizophrenia. An important caveat is that these four variants were not the top variants in their respective regions of association with schizophrenia, and were not primary drivers of the MHC association in schizophrenia in stepwise conditional analyses (9, 11). Therefore, the biological significance of these particular HLA variants in schizophrenia is likely limited.

Outside of the MHC region, we identified five SNPs with potential pleiotropic effects - influencing risk for both schizophrenia and celiac disease (CEL) (rs296547), Crohn's disease (CRO) (rs1734907, rs13126505, rs6738825), or multiple sclerosis (MS) (rs7132277) and schizophrenia. These variants do not appear to be broadly pleiotropic across human traits, as there were no phenome-wide significant results reported for any of these SNPs in the PheWAS Catalog (all $p>1.2x10^{-8}$; all FDR>0.1 (48)). Integration of GWAS, mQTL, and eQTL data implicated *ABCB9* and *EPHB4/TRIP6* as functional candidates underlying the association of rs7132277 and rs1734907, respectively. Although *ABCB9* emerged as a functional candidate driving the association of rs7132277 in schizophrenia, the significant colocalization of this region between schizophrenia and MS (PP=0.94) was not driven by rs7132277 (posterior probability=$1.38x10^{-39}$). Overall, our findings provide the strongest evidence for rs1734907 as a functional, pleiotropic immune variant associated with both Crohn's disease and schizophrenia. In particular, our results suggest a model in which genetic variation at rs1734907 (~85kb upstream of *EPHB4*) increases DNA methylation, upregulates *EPHB4* expression, and decreases the risk of schizophrenia. While DNA methylation is classically associated with gene silencing, the effect of methylation on transcription depends on the genomic context (49); for instance, methylation of silencers or insulators eliminates transcription-blocking activity thereby promoting gene expression (50, 51). *EPHB4* is a transmembrane tyrosine kinase receptor that coordinates cell movement via bidirectional intercellular signaling at sites of direct cell-to-cell contact (52). In the brain, ephrin signaling mediates

synaptic plasticity by initiating and stabilizing neuronal synapse formation (reviewed by (53)). An analogous role has not yet been discovered in the immune system, possibly due to the much shorter lifespan of immunological synapses between lymphocytes and antigen presenting cells (minutes) as compared to neuronal synapses (years) (54, 55). Interestingly, ephrin signaling attenuates the migration responses of both neurons and immune cells toward chemoattractants *in vitro* (56, 57). Thus, pathfinding may be a shared risk mechanism by which *EPHB4* contributes to immune disease and schizophrenia. The hypotheses raised by our findings require further validation. If the association of rs1734907 with CRO and schizophrenia is robustly replicated in future GWAS, functional studies will be needed to investigate both the genetic mechanism by which rs1734907 (or a causal variant in LD with this SNP) influences *EPHB4* transcription, and the biological mechanism by which increased *EPHB4* expression influences susceptibility to CRO and schizophrenia. With the multi-kinase inhibitor dasatinib already on the market for treatment of chronic myeloid leukemia (58) and other EphB4 inhibitors currently in Phase II trials (59–62), the potential for future drug repurposing makes *EPHB4* an attractive candidate for further investigation.

We observed genome-wide sharing of risk variants for schizophrenia and six immune diseases (inflammatory bowel disease (IBD) including both CRO and ulcerative colitis (UC), primary biliary cirrhosis (PBC), psoriasis (PSO), systemic lupus erythematosus (SLE)) using LDSC, all of which have been previously reported to co-occur with schizophrenia in epidemiological studies (3, 5, 63). With the exception of UC, the small positive genetic correlations observed between these immune diseases and schizophrenia (r_g~0.1) were consistent with phenotypic correlations observed in epidemiological data. Thus, currently available genetic data suggest that shared genetic risk contributes to the co-occurrence of CRO, PBC, PSO, and SLE in schizophrenia. We note that most of the phenotypic correlation between these immune diseases and schizophrenia appears to be captured by common genetic variation. Interestingly, phenotypic correlations for the remaining five immune diseases were similar in magnitude (r_p=0.0.04-0.15), but no genetic correlation was detected. Possible explanations for this include inadequate statistical power to detect genetic correlations for these diseases in our study, or a stronger

environmental component contributing to the epidemiological relationship of these disorders with schizophrenia. Some of the immune diseases which did show significant genetic correlation with schizophrenia (CRO, UC) are considered autoinflammatory diseases (64), and the others (PBC, PSO, SLE) have a strong inflammatory component (64–66). This raises the possibility that the genetic risk we observed between these particular immune diseases and schizophrenia reflects a subgroup of inflammation-driven schizophrenia cases, and/or sharing of specific innate immunity pathways between schizophrenia and these particular immune diseases.

To our knowledge, this is the first time that sex-dependent genetic correlation with immune diseases has been investigated in schizophrenia. Interestingly, SLE and systemic sclerosis (SSC) showed significant genetic correlation with schizophrenia only among males in sex-stratified analyses. Our findings raise the possibility of male-specific pleiotropy between schizophrenia and certain immune diseases. Interestingly, animal studies indicate that sex hormones have opposing effects on predisposition to schizophrenia and autoimmunity; estrogen has been reported to protect against the development of schizophrenia (62), while androgens appear to protect against the development autoimmune diseases (63, 64). We emphasize that our sex-dependent findings require validation in independent samples.

Our work was subject to several important limitations. Firstly, we did not have access to imputed datasets for the original immune disease GWAS reporting the potentially pleiotropic SNPs (**Table 3**). Lack of dense SNP coverage in the regions of interest resulted in low power to detect colocalizing association in schizophrenia, and thus our analyses are inconclusive with respect to pleiotropic potential for the two regions that did not show significant colocalization. Secondly, genome-wide summary statistics were not available for all of the immune diseases, resulting in a more limited analysis of 11 diseases for estimating genetic correlations. Thirdly, we used LDSC to estimate genetic correlations because of its robustness to non-independence of GWAS samples as many of the samples analyzed included Wellcome Trust Case Control Consortium samples. Compared to alternative methods for estimating genetic correlation that use individual-level genotype data, LDSC has lower statistical power. Thus, we cannot exclude the possibility of additional immune-schizophrenia genetic relationships not

identified in our analyses. Finally, we had greater statistical power to detect genetic correlation with immune diseases in the male-specific schizophrenia GWAS compared to the female-specific GWAS. This difference in power may account for the male-specific genetic correlations we observed for SLE and SSC, as opposed to sex-dependent pleiotropic effects.

Despite these limitations, our work adds to a growing body of evidence suggesting that schizophrenia and immune diseases share genetic risk factors. There are conflicting reports in the literature with respect to the specific immune diseases demonstrating genetic overlap with schizophrenia, and the direction of effect (positive or negative genetic correlation). Genetic overlap with schizophrenia has been previously investigated for nine of the 19 immune diseases studied here. Genome-wide genetic correlation with schizophrenia has been previously reported for CRO (25–27, 29), MS (30), PBC (27), PSO (27, 31), rheumatoid arthritis (RA, both positive (25, 26) and negative (33) genetic correlations), SLE (26, 27), T1D (25), and UC (26–29) (see **S1 Table** for a summary of previous studies). Our results are consistent with previously reported genetic overlap between schizophrenia and CRO (25–27, 29), PBC (27), PSO (27), SLE (26, 27), and UC (26, 27). We did not find any significant genetic correlation between schizophrenia and CEL, type 1 diabetes (T1D), rheumatoid arthritis (RA), systemic sclerosis (SSC), or vitiligo (VIT). For RA in particular, there is a robust inverse epidemiological association with schizophrenia (8). However, the genetic association is less consistent. Using methods based on summary statistics (including PRS and LDSC), four previous studies reported no evidence of pleiotropy between schizophrenia and RA (8, 16, 27, 32), while two studies reported positive genetic correlation (25, 26). Notably, Lee *et al.* reported an inverse genetic correlation – in keeping with the observed epidemiological effect – using restricted maximum likelihood (GREML), a method utilizing full genotype data which has greater statistical power to detect small pleiotropic effects than PRS or LDSC (33). Given the modest genetic correlations observed in the present study, subtle differences in statistical power across studies using different statistical methods and GWAS datasets may explain these discrepant findings. As genetic samples continue to grow, and our understanding of the degree of genetic overlap expected among complex traits evolves, it will be worthwhile to revisit these analyses.

Overall, our analyses provide statistical evidence supporting extensive pleiotropy between immune diseases and schizophrenia. Our results highlight *EPHB4*, a transmembrane receptor that coordinates cell migration and has dual roles in immune cell and neuronal pathfinding, as a promising candidate for future functional studies. More broadly, our findings indicate that common genetic variants influencing the risk of immune diseases – in particular CRO, PBC, PSO, SLE, and UC – are also involved in schizophrenia. Studies identifying the cell types and biological pathways that may be driving this genetic overlap are needed, and will hopefully provide further insights into the pathophysiology of schizophrenia. In the meantime, our work supports the emerging hypothesis that pathogenic mechanisms are shared across immune and central nervous system disorders.

Materials and Methods

Samples and quality control

We used either imputed genotype data or summary statistics generated as described in the original GWAS. For sample details, see **Table 1**.

Schizophrenia dataset

We used data from the PGC2 study (1). For analyses of non-HLA genome-wide significant risk variants for immune diseases we used publicly available summary statistics from the total dataset (52 cohorts; 35,476 cases and 46,839 controls) (1). For PRS analyses we used all 36 European ancestry case-control cohorts with available individual-level genotype data (25,629 cases and 30,976 controls). For analyses including HLA variants we used a further refined 31 European ancestry case-control cohorts (20,253 cases and 25,011 controls) with high-quality coverage of the MHC region, as previously described (11).

Immune disease datasets

To estimate the extent of genetic overlap between schizophrenia and immune diseases, we obtained full GWAS summary statistics for 11 of the 19 immune diseases (eight immune diseases were not included in PRS analyses due to lack of available summary statistics). We obtained publicly available summary statistics for five immune diseases (see URLs): CRO (67), IBD (67), RA (68), SLE (69), and UC (67). For the following six immune diseases, we obtained summary statistics with permission from the authors: CEL (39), PBC (70), PSO (71), SSC (72), T1D (73), and VIT (74).

Testing the association of genome-wide significant risk alleles for 19 immune diseases in schizophrenia

For each of the 19 immune diseases, we defined risk loci outside of the MHC region (chromosome 6: 25-34 Mb) using curated GWAS results from ImmunoBase (http://www.immunobase.org; accessed 7 June 2015. For details, see **Supplementary Methods**). Notably, the majority of IBD risk variants were also risk variants for CRO and/or UC. Within the MHC region we considered only the most strongly associated HLA variant (including SNPs, imputed HLA amino acid sites, and classical alleles) for each disease based on univariate analysis in previously published studies (see **Table 2**), because multivariate conditional analyses reporting adjusted effect sizes of independent HLA variants were not available for all immune diseases. In total there were 581 unique variants (563 non-HLA variants and 18 HLA variants) associated with any immune disease at genome-wide significance. A complete list of non-HLA and HLA immune risk variants is provided in **Table 2** and **S2 Table**, respectively.

First, we tested for shared direction of effect with schizophrenia among SNPs associated with each of the 19 immune diseases using the binomial sign test. Because some immune risk SNPs were associated with multiple diseases with inconsistent direction of effect, we could not evaluate shared direction of effect among the collective set of immune risk SNPs in schizophrenia.

Next, we evaluated the collective association of SNPs associated with any immune disease. First we extracted the p-values for a pruned set of 261 LD-independent, non-HLA immune risk SNPs with linkage disequilibrium (LD) Score and minor allele frequency (MAF) information were available in the European LD Score database (16) from the schizophrenia PGC2 GWAS. We then quantified enrichment of these immune risk SNP associations in schizophrenia using the genomic inflation value λ. We obtained an empirical enrichment p-value by comparing this to λ values from 1,000 equal-sized sets of SNPs drawn from the schizophrenia GWAS summary data, and matched to the immune SNP set for MAF and LD score as these parameters are correlated with GWAS test statistics (see **Supplementary Methods** for details).

Finally, we evaluated the association of each of the 581 variants with schizophrenia using previously published association results for non-HLA (1) and HLA variants (11). We considered SNPs associated with schizophrenia at $p<8.6x10^{-5}$ (Bonferroni correction for 581 tests, 563 non-HLA and 18 HLA variants) to have pleiotropic effects.

To evaluate the pleiotropic potential of immune risk variants significantly associated with schizophrenia, we tested for colocalization of association signals in immune diseases and schizophrenia using the Bayesian coloc2 method implemented in the R package coloc (35) . We used window sizes that captured all SNPs showing $r^2>0.2$ with each immune risk variant (chr1:200.8-201.2Mb; chr2:198.0-199.0Mb; chr4:102.6-103.4Mb; chr7:100.2-100.5Mb; chr12:123.4-123.9Mb). We used default prior probabilities for immune risk variants being associated with immune disease ($p=1x10^{-4}$), schizophrenia ($p=1x10^{-4}$), and both immune disease and schizophrenia ($p=1x10^{-5}$) as recommended. In addition to colocalization testing, we performed conditional and joint analysis (COJO) using GCTA (75). Specifically, we used COJO to perform association analyses in the PGC2 schizophrenia GWAS conditioning on the immune risk variants of interest (i.e. SNPs that were significantly associated with both an immune disease and schizophrenia). If the immune risk variants explained the regional associations in schizophrenia, no significant associations with schizophrenia should remain after

conditioning on these variants (statistically, all $p>8.6x10^{-5}$). We used the 1000 Genomes Phase 3 European dataset as a reference panel to calculate LD between SNPs.

To prioritize genes and regulatory elements driving the pleiotropic GWAS loci we identified (associated with both immune disease and schizophrenia, see **Table 3**), we followed the analytic approach described by Wu *et al.* (43) . This approach integrates summary statistics from independent -omics methylation quantitative trait loci (mQTL) studies, expression quantitative trait loci (eQTL) studies, and GWAS to identify SNPs associated with gene expression, DNA methylation, and disease through shared genetic effects.

We obtained mQTL and eQTL data used in Wu *et al.* (43) for genetic regions within a 2Mb window of each pleiotropic SNP. These data and the quality control measures applied have been described in detail elsewhere (43). Briefly, mQTL summary-level SNP data were from a meta-analysis of the Brisbane Systems Genetics Study (44) and Lothian Birth Cohorts of 1921 and 1936 (45), which comprised 1,980 individuals with DNA methylation measured in peripheral blood. eQTL summary-level SNP data were from the Consortium for the Architecture of Gene Expression (CAGE) study (76), which comprised 2,765 individuals with gene expression levels measured in peripheral blood. GWAS summary-level SNP data for schizophrenia was from the PGC2 study (1).

We applied summary data-based Mendelian randomization (SMR) using GCTA (75) to test for shared associations between the pleiotropic SNPs with DNAm probes and gene expression probes, DNAm probes and schizophrenia, and gene expression probes and schizophrenia. We included DNAm and gene expression probes within 2Mb of the pleiotropic SNPs. We considered significant associations as those with $p_{SMR}<1.30x10^{-4}$ (0.05/385 tagged genes) for mQTLs and $p_{SMR}<4.31x10^{-4}$ for eQTLs (0.05/116 tagged genes). Next, we applied the heterogeneity in dependent instruments (HEIDI) test (42) using GCTA (75) to evaluate whether significant shared associations between DNAm, gene expression and schizophrenia were driven by linkage (i.e. separate causal variants in LD exerting genetic effects on DNAm, gene expression, and schizophrenia) or a shared pleiotropic causal variant. We considered genetic effects that passed the HEIDI test ($p_{HEIDI}>0.01$) to be driven by a single causal variant. We looked for

consistent SMR and HEIDI results across GWAS, mQTL, and eQTL studies to prioritize genes for future functional studies.

Estimating the degree of genetic correlation between schizophrenia and 14 immune diseases

To evaluate whether common variants influencing risk of immune diseases collectively contribute to schizophrenia, we used cross-trait LDSC (16, 32, 77). Cross-trait LDSC estimates the genetic correlation (r_g) between two traits using GWAS summary statistics. Our main analysis included 11 immune diseases with available genome-wide summary statistics. To benchmark the amount of genetic overlap between schizophrenia and immune disease, we included bipolar disorder as a positive control (12) and human height as a negative control (78) based on their previously reported genetic correlations with schizophrenia using LDSC (16).

The statistical framework for cross-trait LDSC has been described in detail previously (16). Briefly, LDSC leverages the relationship between LD and association test statistics to estimate heritability as the slope of the regression of z-scores against LD scores (79). Cross-trait LDSC is a bivariate extension of this method which estimates genetic covariance as the slope of the regression of the products of z-scores against LD scores using the following equation (16):

$$E[z_{1j}z_2|\ell_j] = \frac{\sqrt{N_1N_2}\varrho_g}{M}\ell_j + \frac{\varrho N_S}{\sqrt{N_1N_2}}$$

where z_{ij} denotes the z score for study i and SNPj, ℓ_j is the LD score (79), N_i is the sample size for study i, ϱ_g is the genetic covariance, M is the number of SNPs in the reference panel with MAF between 5% and 50%, N_S is the number of individuals included in both studies, and ϱ is the phenotypic correlation among the N_S overlapping samples. Genetic covariance ϱ_g is estimated by regressing $z_{1j}z_2$ against $\ell_j\sqrt{N_1N_2}$, and multiplying the resulting slope by M. Statistical significance is assessed using block jackknifing over 200 equally sized blocks of SNPs (16). Importantly, the MHC region is excluded from LDSC analyses due to its unusual LD structure and genetic architecture (47).

Because LDSC is robust to sample sharing across GWAS (16), we used summary statistics from the 49 European-ancestry cohorts in the PGC2 study (31,335 cases and 38,765 controls) (1). We used LD Scores from the "eur_w_ld_chr/" files available from https://data.broadinstitute.org/alkesgroup/LDSCORE, computed using 1000 Genomes Project (80) Europeans as a reference panel as previously described (47). To ensure we were using well-imputed SNPS we filtered all GWAS as previously described (16), including limiting the analysis to HapMap 3 (81) SNPs as implemented in the LDSC script munge_sumstats.py (https://github.com/bulik/ldsc). We estimated h^2 for each trait on the observed scale (**S4 Table**). We considered traits with r_g $p<0.05$ to have significant genetic correlation with schizophrenia.

Benchmarking with epidemiological data

To determine how much of the phenotypic correlation between schizophrenia and immune-mediated diseases was explained by the genetic correlations we observed, we used the approach previously described by Lee *et al.* (33). Briefly, we benchmarked our genetic correlation estimates between schizophrenia and immune diseases relative to the expected phenotypic correlations from epidemiological data. We obtained estimates of the population risk of schizophrenia (K_{SCZ}), the population risk of each immune disease (K_{IMMUNE}), and the probability of each immune disease among patients with schizophrenia ($K_{IMMUNE\ |\ SCZ}$) from the literature as referenced in **S4 Table**. We estimated the phenotypic correlation between schizophrenia and the immune disease of interest ($R_{SCZ\text{-}IMMUNE}$) using the following formula, as derived by Lee *et al.* (33) assuming that the phenotypic liabilities of schizophrenia (l_{SCZ}) and immune disease (l_{IMMUNE}) follow a bivariate normal distribution with mean=0 and standard deviation=1:

$$R_{SCZ-IMMUNE} = \frac{i_{SCZ}t_{IMMUNE} - \sqrt{i_{SCZ}^2 t_{IMMUNE}^2 - (t_{IMMUNE\ |\ SCZ}^2 + i_{SCZ}^2)(t_{IMMUNE}^2 - t_{IMMMUNE\ |\ SCZ}^2)}}{(t_{IMMUNE\ |\ SCZ}^2 + i_{SCZ}^2)}$$

where:

t_{SCZ} is the liability threshold for schizophrenia:

Z-score of the $(1 - \mathrm{K_{SCZ}})^{\mathrm{th}}$ percentile

t_{IMMUNE} is the liability threshold for immune disease:

Z-score of the $(1 - K_{IMMUNE})^{\mathrm{th}}$ percentile

$t_{IMMUNE\ |\ SCZ}$ is the liability threshold for immune disease in those with schizophrenia:

Z-score of the $(1 - K_{IMMUNE\ |\ SCZ})^{\mathrm{th}}$ percentile

d_{SCZ} is the "height" of the normal distribution at the schizophrenia liability threshold:

probability density function of t_{SCZ}

i_{SCZ} is the mean phenotypic liability of those with schizophrenia:

d_{SCZ}/K_{SCZ}

Statistical power

Power to detect association of individual non-HLA and HLA immune risk variants in schizophrenia was calculated using the Genetic Power Calculator (82) assuming a risk allele frequency (RAF) of 0.05, disease prevalence of 1%, and significance threshold (α) of $8.6\text{x}10^{-5}$. We used a RAF of 0.05 as this was the point at which statistical power remained >80% given the prevalence of schizophrenia and size of our sample. The vast majority of immune variants had MAF >0.05 (551/563 variants).

URLs

LD Score database: ftp://atguftp.mgh.harvard.edu/brendan/1k_eur_r2_hm3snps_se_weights.RDS

Publicly available GWAS summary statistics:

- CRO, IBD, UC

  ftp://ftp.sanger.ac.uk/pub/consortia/ibdgenetics/iibdgc-trans-ancestry-filtered-summary-stats.tgz

- RA

  http://www.broadinstitute.org/ftp/pub/rheumatoid_arthritis/Stahl_etal_2010NG/

- SLE

  https://www.immunobase.org/downloads/protected_data/GWAS_Data/hg19_gwas_sle_bentham_4_20_0.tab.gz

Acknowledgements

We thank all study participants, and all research staff at the numerous study sites that made this research possible. Computations were performed on the Genetic Cluster Computer (GCC) hosted by SURFsara (www.surfsara.nl). We thank ImmunoBase for curating genetic risk factors for the immune diseases investigated. We thank the Canadian-US PBC Consortium, the Italian PBC Genetics Study Group and the UK-PBC Consortium for sharing their summary data for the PBC GWAS. We thank the International Inflammatory Bowel Disease Genetics Consortium (IIBDGC) for providing publicly available genome-wide summary statistics for CRO, UC, and IBD. We also thank the investigators leading the GWAS for RA and SLE for providing genome-wide summary statistics from their studies on ImmunoBase.

This research was supported in part by a number of funding sources. This research uses resources provided by the Genetic Association Information Network (GAIN), obtained from the database of Genotypes and Phenotypes (dbGaP) found at http://www.ncbi.nlm.nih.gov/gap through dbGaP accession number phs000021.v3.p2; samples and associated phenotype data for this study were provided by the Molecular Genetics of Schizophrenia Collaboration (PI: Pablo V. Gejman, Evanston Northwestern Healthcare (ENH) and Northwestern University, Evanston, IL, USA). JG Pouget is supported by Fulbright Canada, the Weston Foundation, and Brain Canada through the Canada Brain Research Fund - a public-private partnership established by the Government of Canada. B Han is supported by the National Research Foundation of Korea (NRF) (grant 2016R1C1B2013126) and the Bio & Medical Technology

Development Program of the NRF (grant 2017M3A9B6061852) funded by the Korean government, Ministry of Science and ICT (MSIT). HM Ollila is supported by the Finnish Cultural Foundation and Academy of Finland (grant 309643). J Martin is supported by grants SAF2015-66761-P from the Spanish Ministry of Economy and Competitiveness and P12-BIO-1395 from Consejería de Innovación, Ciencia y Tecnología, Junta de Andalucía (Spain). Y Jin, SA Santorico, and R Spritz are supported by the U.S. National Institutes of Health (NIH) grants R01AR045584, R01AR056292, X01HG007484, and P30AR057212 and by institutional research funding IUT20-46 from the Estonian Ministry of Education and Research. MD Mayes is supported by the U.S. NIH grants N01AR02251 and R01AR05528. S Raychaudhuri is supported by the U.S. NIH grants 1R01AR063759, 1R01AR062886, 1UH2AR067677-01, and U19AI111224-01 and Doris Duke Charitable Foundation grant 2013097. Funding for the GAIN schizophrenia sample was provided by the U.S. NIH grants R01 MH67257, R01 MH59588, R01 MH59571, R01 MH59565, R01 MH59587, R01 MH60870, R01 MH59566, R01 MH59586, R01 MH61675, R01 MH60879, R01 MH81800, U01 MH46276, U01 MH46289 U01 MH46318, U01 MH79469, and U01 MH79470 and the genotyping of samples was provided through GAIN. The funding sources did not influence the study design, data analysis, or writing of this manuscript.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

References

1. Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, **511**, 421-427.
2. Chan,M.K., Krebs,M.O., Cox,D., Guest,P.C., Yolken,R.H., Rahmoune,H., Rothermundt,M., Steiner,J., Leweke,F.M., van Beveren,N.J., et al. (2015) Development of a blood-based molecular biomarker test for identification of schizophrenia before disease onset. *Transl Psychiatry*, **5**, e601.
3. Benros,M.E., Nielsen,P.R., Nordentoft,M., Eaton,W.W., Dalton,S.O., Mortensen,P.B. (2011) Autoimmune diseases and severe infections as risk factors for schizophrenia: a 30-year population-based register study. *Am J Psychiatry*, **168**, 1303-1310.
4. Chen,S.J., Chao,Y.L., Chen,C.Y., Chang,C.M., Wu,E.C., Wu,C.S., Yeh,H.H., Chen,C.H., Tsai,H.J. (2012) Prevalence of autoimmune diseases in in-patients with schizophrenia: nationwide population-based study. *Br J Psychiatry*, **200**, 374-380.
5. Benros,M.E., Pedersen,M.G., Rasmussen,H., Eaton,W.W., Nordentoft,M., Mortensen,P.B. (2014) A nationwide study on the risk of autoimmune diseases in individuals with a personal or a family history of schizophrenia and related psychosis. *Am J Psychiatry*, **171**, 218-226.

6. Eaton,W.W., Byrne,M., Ewald,H., Mors,O., Chen,C.Y., Agerbo,E., Mortensen,P.B. (2006) Association of schizophrenia and autoimmune diseases: linkage of Danish national registers. *Am J Psychiatry*, **163**, 521-528.
7. Eaton,W.W., Pedersen,M.G., Nielsen,P.R., Mortensen,P.B. (2010) Autoimmune diseases, bipolar disorder, and non-affective psychosis. *Bipolar Disord*, **12**, 638-646.
8. Euesden,J., Breen,G., Farmer,A., McGuffin,P., Lewis,C.M. (2015) The relationship between schizophrenia and rheumatoid arthritis revisited: genetic and epidemiological analyses. *Am J Med Genet B Neuropsychiatr Genet*, **168B**, 81-88.
9. Sekar,A., Bialas,A.R., de Rivera,H., Davis,A., Hammond,T.R., Kamitaki,N., Tooley,K., Presumey,J., Baum,M., Van Doren,V., et al. (2016) Schizophrenia risk from complex variation of complement component 4. *Nature*, **530**, 177-183.
10. Yih Chen,J., Ling Wu,Y., Yin Mok,M., Jan Wu,Y.J., Lintner,K.E., Wang,C.M., Chung,E.K., Yang,Y., Zhou,B., Wang,H., et al. (2016) Effects of complement C4 gene copy number variations, size dichotomy, and C4A deficiency on genetic risk and clinical presentation of systemic lupus erythematosus in East Asian populations. *Arthritis Rheumatol*, **68**, 1442-1453.
11. Pouget,J.G., Goncalves,V.F., Spain,S.L., Finucane,H.K., Raychaudhuri,S., Kennedy,J.L., Knight,J. (2016) Genome-wide association studies suggest limited immune gene enrichment in schizophrenia compared to 5 autoimmune diseases. *Schizophr Bull*, **42**, 1176-1184.
12. Cross-Disorder Group of the Psychiatric Genomics Consortium (2013) Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet*, **381**, 1371-1379.
13. Cross-Disorder Group of the Psychiatric Genomics Consortium (2013) Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet*, **45**, 984-994.
14. Voight,B.F., Peloso,G.M., Orho-Melander,M., Frikke-Schmidt,R., Barbalic,M., Jensen,M.K., Hindy,G., Holm,H., Ding,E.L., Johnson,T., et al. (2012) Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet*, **380**, 572-580.
15. Do,R., Willer,C.J., Schmidt,E.M., Sengupta,S., Gao,C., Peloso,G.M., Gustafsson,S., Kanoni,S., Ganna,A., Chen,J., et al. (2013) Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat Genet*, **45**, 1345-1352.
16. Bulik-Sullivan,B., Finucane,H.K., Anttila,V., Gusev,A., Day,F.R., Loh,P.R., Duncan,L., Perry,J.R., Patterson,N., Robinson,E.B., et al. (2015) An atlas of genetic correlations across human diseases and traits. *Nat Genet*, **47**, 1236-1241.
17. Ellinghaus,D., Jostins,L., Spain,S.L., Cortes,A., Bethune,J., Han,B., Park,Y.R., Raychaudhuri,S., Pouget,J.G., Hübenthal,M., et al. (2016) Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet*, **48**, 510-518.
18. Brainstorm Consortium, Anttila,V., Bulik-Sullivan,B., Finucane,H.K., Walters,R.K., Bras,J., Duncan,L., Escott-Price,V., Falcone,G.J., Gormley,P., et al. (2018) Analysis of shared heritability in common disorders of the brain. *Science*, **360**,
19. Paaby,A.B., Rockman,M.V. (2013) The many faces of pleiotropy. *Trends Genet*, **29**, 66-73.
20. Han,B., Pouget,J.G., Slowikowski,K., Stahl,E., Lee,C.H., Diogo,D., Hu,X., Park,Y.R., Kim,E., Gregersen,P.K., et al. (2016) A method to decipher pleiotropy by detecting underlying heterogeneity driven by hidden subgroups applied to autoimmune and neuropsychiatric diseases. *Nat Genet*, **48**, 803-810.
21. Chesmore,K., Bartlett,J., Williams,S.M. (2018) The ubiquity of pleiotropy in human disease. *Hum Genet*, **137**, 39-44.
22. Verbanck,M., Chen,C.Y., Neale,B., Do,R. (2018) Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet*, **50**, 693-698.
23. Jordan,D.M., Verbanck,M., Do,R. (2018) The landscape of pervasive horizontal pleiotropy in human

genetic variation is driven by extreme polygenicity of human traits and diseases. *bioRxiv*, **doi: https://doi.org/10.1101/416545**,
24. Gandal,M.J., Haney,J.R., Parikshak,N.N., Leppa,V., Ramaswami,G., Hartl,C., Schork,A.J., Appadurai,V., Buil,A., Werge,T.M., et al. (2018) Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science*, **359**, 693-697.
25. Stringer,S., Kahn,R.S., de Witte,L.D., Ophoff,R.A., Derks,E.M. (2014) Genetic liability for schizophrenia predicts risk of immune disorders. *Schizophr Res*, **159**, 347-352.
26. Wang,Q., Yang,C., Gelernter,J., Zhao,H. (2015) Pervasive pleiotropy between psychiatric disorders and immune disorders revealed by integrative analysis of multiple GWAS. *Hum Genet*, **134**, 1195-1209.
27. Tylee,D.S., Sun,J., Hess,J.L., Tahir,M.A., Sharma,E., Malik,R., Worrall,B.B., Levine,A.J., Martinson,J.J., Nejenstev,S., et al. (2018) Genetic correlations among psychiatric and immune-related phenotypes based on genome-wide association data. *Am J Med Genet B Neuropsychiatr Genet*, **177**, 641-657.
28. Duncan,L.E., Shen,H., Ballon,J.S., Hardy,K.V., Noordsy,D.L., Levinson,D.F. (2018) Genetic correlation profile of schizophrenia mirrors epidemiological results and suggests link between polygenic and rare variant (22q11.2) cases of schizophrenia. *Schizophr Bull*, **44**, 1350-1361.
29. Pickrell,J.K., Berisa,T., Liu,J.Z., Segurel,L., Tung,J.Y., Hinds,D.A. (2016) Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet*, **48**, 709-717.
30. Andreassen,O.A., Harbo,H.F., Wang,Y., Thompson,W.K., Schork,A.J., Mattingsdal,M., Zuber,V., Bettella,F., Ripke,S., Kelsoe,J.R., et al. (2015) Genetic pleiotropy between multiple sclerosis and schizophrenia but not bipolar disorder: differential involvement of immune-related gene loci. *Mol Psychiatry*, **20**, 207-214.
31. Yin,X., Wineinger,N.E., Wang,K., Yue,W., Norgren,N., Wang,L., Yao,W., Jiang,X., Wu,B., Cui,Y., et al. (2016) Common susceptibility variants are shared between schizophrenia and psoriasis in the Han Chinese population. *J Psychiatry Neurosci*, **41**, 413-421.
32. Purcell,S.M., Wray,N.R., Stone,J.L., Visscher,P.M., O'Donovan,M.C., Sullivan,P.F., Sklar,P. (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, **460**, 748-752.
33. Lee,S.H., Byrne,E.M., Hultman,C.M., Kahler,A., Vinkhuyzen,A.A., Ripke,S., Andreassen,O.A., Frisell,T., Gusev,A., Hu,X., et al. (2015) New data and an old puzzle: the negative association between schizophrenia and rheumatoid arthritis. *Int J Epidemiol*, **44**, 1706-1721.
34. Fernando,M.M., Stevens,C.R., Walsh,E.C., De Jager,P.L., Goyette,P., Plenge,R.M., Vyse,T.J., Rioux,J.D. (2008) Defining the role of the MHC in autoimmunity: a review and pooled analysis. *PLoS Genet*, **4**, e1000024.
35. Giambartolomei,C., Vukcevic,D., Schadt,E.E., Franke,L., Hingorani,A.D., Wallace,C., Plagnol,V. (2014) Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet*, **10**, e1004383.
36. Franke,A., McGovern,D.P., Barrett,J.C., Wang,K., Radford-Smith,G.L., Ahmad,T., Lees,C.W., Balschun,T., Lee,J., Roberts,R., et al. (2010) Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet*, **42**, 1118-1125.
37. Jostins,L., Ripke,S., Weersma,R.K., Duerr,R.H., McGovern,D.P., Hui,K.Y., Lee,J.C., Schumm,L.P., Sharma,Y., Anderson,C.A., et al. (2012) Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*, **491**, 119-124.
38. Beecham,A.H., Patsopoulos,N.A., Xifara,D.K., Davis,M.F., Kemppinen,A., Cotsapas,C., Shah,T.S., Spencer,C., Booth,D., Goris,A., et al. (2013) Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat Genet*, **45**, 1353-1360.
39. Dubois,P.C., Trynka,G., Franke,L., Hunt,K.A., Romanos,J., Curtotti,A., Zhernakova,A., Heap,G.A., Adany,R., Aromaa,A., et al. (2010) Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet*, **42**, 295-302.

40. 1000 Genomes Project Consortium, Auton,A., Brooks,L.D., Durbin,R.M., Garrison,E.P., Kang,H.M., Korbel,J.O., Marchini,J.L., McCarthy,S., McVean,G.A., et al. (2015) A global reference for human genetic variation. *Nature*, **526**, 68-74.
41. Yang,J., Ferreira,T., Morris,A.P., Medland,S.E., Genetic,I.O.A.N.T.G.I.A.N.T.C., DIAbetes,G.R.A.M.-a.D.I.A.G.R.A.M.C., Madden,P.A., Heath,A.C., Martin,N.G., Montgomery,G.W., et al. (2012) Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet*, **44**, 369-75, S1.
42. Zhu,Z., Zhang,F., Hu,H., Bakshi,A., Robinson,M.R., Powell,J.E., Montgomery,G.W., Goddard,M.E., Wray,N.R., Visscher,P.M., et al. (2016) Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*, **48**, 481-487.
43. Wu,Y., Zeng,J., Zhang,F., Zhu,Z., Qi,T., Zheng,Z., Lloyd-Jones,L.R., Marioni,R.E., Martin,N.G., Montgomery,G.W., et al. (2018) Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat Commun*, **9**, 918.
44. Powell,J.E., Henders,A.K., McRae,A.F., Caracella,A., Smith,S., Wright,M.J., Whitfield,J.B., Dermitzakis,E.T., Martin,N.G., Visscher,P.M., et al. (2012) The Brisbane Systems Genetics Study: genetical genomics meets complex trait genetics. *PLoS One*, **7**, e35430.
45. Chen,B.H., Marioni,R.E., Colicino,E., Peters,M.J., Ward-Caviness,C.K., Tsai,P.C., Roetker,N.S., Just,A.C., Demerath,E.W., Guan,W., et al. (2016) DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging (Albany NY)*, **8**, 1844-1865.
46. Stahl,E.A., Wegmann,D., Trynka,G., Gutierrez-Achury,J., Do,R., Voight,B.F., Kraft,P., Chen,R., Kallberg,H.J., Kurreeman,F.A., et al. (2012) Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat Genet*, **44**, 483-489.
47. Finucane,H.K., Bulik-Sullivan,B., Gusev,A., Trynka,G., Reshef,Y., Loh,P.R., Anttila,V., Xu,H., Zang,C., Farh,K., et al. (2015) Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*, **47**, 1228-1235.
48. Denny,J.C., Bastarache,L., Ritchie,M.D., Carroll,R.J., Zink,R., Mosley,J.D., Field,J.R., Pulley,J.M., Ramirez,A.H., Bowton,E., et al. (2013) Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol*, **31**, 1102-1110.
49. Jones,P.A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet*, **13**, 484-492.
50. Bell,A.C., Felsenfeld,G. (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*, **405**, 482-485.
51. Eden,S., Constancia,M., Hashimshony,T., Dean,W., Goldstein,B., Johnson,A.C., Keshet,I., Reik,W., Cedar,H. (2001) An upstream repressor element plays a role in Igf2 imprinting. *EMBO J*, **20**, 3518-3525.
52. Davis,S., Gale,N.W., Aldrich,T.H., Maisonpierre,P.C., Lhotak,V., Pawson,T., Goldfarb,M., Yancopoulos,G.D. (1994) Ligands for EPH-related receptor tyrosine kinases that require membrane attachment or clustering for activity. *Science*, **266**, 816-819.
53. Hruska,M., Dalva,M.B. (2012) Ephrin regulation of synapse formation, function and plasticity. *Mol Cell Neurosci*, **50**, 35-44.
54. Dustin,M.L. (2012) Signaling at neuro/immune synapses. *J Clin Invest*, **122**, 1149-1155.
55. Basu,R., Huse,M. (2017) Mechanical communication at the immunological synapse. *Trends Cell Biol*, **27**, 241-254.
56. Sharfe,N., Freywald,A., Toro,A., Dadi,H., Roifman,C. (2002) Ephrin stimulation modulates T cell chemotaxis. *Eur J Immunol*, **32**, 3745-3755.
57. Lu,Q., Sun,E.E., Klein,R.S., Flanagan,J.G. (2001) Ephrin-B reverse signaling is mediated by a novel PDZ-RGS protein and selectively inhibits G protein-coupled chemoattraction. *Cell*, **105**, 69-79.

58. U.S. Food and Drug Administration, Centre for Drug Evaluation and Research Sprycel (dasatinib) NDA 21986/22072 approval letter, 2006 June 28. Retrieved 2018 November 4, from https://www.accessdata.fda.gov/drugsatfda_docs/nda/2006/021986s000_Sprycel__APPROV.pdf.,
59. ClinicalTrials.gov [Internet] Bethesda (MD): National Library of Medicine (US). 2016 March 23. Identifier: NCT02717156, Combination therapy with Pembrolizumab and sEphB4-HSA in previously treated urothelial carcinoma. Available from: https://clinicaltrials.gov/ct2/show/NCT02717156.,
60. ClinicalTrials.gov [Internet] Bethesda (MD): National Library of Medicine (US). 2016 June 15. Identifier NCT02799485, EphB4-HSA in treating patients with Kaposi sarcoma. Available from: https://clinicaltrials.gov/ct2/show/NCT02799485.,
61. ClinicalTrials.gov [Internet] Bethesda (MD): National Library of Medicine (US). 2017 May 10. Identifier NCT03146871, Recombinant EphB4-HSA fusion protein and Azacitidine or Decitabine in treating patients with relapsed or refractory myelodysplastic syndrome, chronic myelomonocytic leukemia, or acute myeloid leukemia previously treated with a hypomethylating agent. Available from: https://clinicaltrials.gov/ct2/show/NCT03146871.,
62. [Internet],C. Bethesda (MD): National Library of Medicine (US). 2017 February 10. Identifier: NCT03049618, Recombinant EphB4-HSA fusion protein and Pembrolizumab in treating patients with locally advanced or metastatic non-small cell lung cancer or locally recurrent or metastatic head and neck squamous cell cancer. Available from: https://clinicaltrials.gov/ct2/show/NCT03049618.,
63. Canellas,F., Lin,L., Julia,M.R., Clemente,A., Vives-Bauza,C., Ollila,H.M., Hong,S.C., Arboleya,S.M., Einen,M.A., Faraco,J., et al. (2014) Dual cases of type 1 narcolepsy with schizophrenia and other psychotic disorders. *J Clin Sleep Med*, **10**, 1011-1018.
64. McGonagle,D., McDermott,M.F. (2006) A proposed classification of the immunological diseases. *PLoS Med*, **3**, e297.
65. Kontaki,E., Boumpas,D.T. (2010) Innate immunity in systemic lupus erythematosus: sensing endogenous nucleic acids. *J Autoimmun*, **35**, 206-211.
66. Lleo,A., Invernizzi,P. (2013) Apotopes and innate immune system: novel players in the primary biliary cirrhosis scenario. *Dig Liver Dis*, **45**, 630-636.
67. Liu,J.Z., van Sommeren,S., Huang,H., Ng,S.C., Alberts,R., Takahashi,A., Ripke,S., Lee,J.C., Jostins,L., Shah,T., et al. (2015) Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet*, **47**, 979-986.
68. Stahl,E.A., Raychaudhuri,S., Remmers,E.F., Xie,G., Eyre,S., Thomson,B.P., Li,Y., Kurreeman,F.A., Zhernakova,A., Hinks,A., et al. (2010) Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet*, **42**, 508-514.
69. Bentham,J., Morris,D.L., Cunninghame Graham,D.S., Pinder,C.L., Tombleson,P., Behrens,T.W., Martin,J., Fairfax,B.P., Knight,J.C., Chen,L., et al. (2015) Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat Genet*, **47**, 1457-1464.
70. Cordell,H.J., Han,Y., Mells,G.F., Li,Y., Hirschfield,G.M., Greene,C.S., Xie,G., Juran,B.D., Zhu,D., Qian,D.C., et al. (2015) International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. *Nat Commun*, **6**, 8019.
71. Strange,A., Capon,F., Spencer,C.C., Knight,J., Weale,M.E., Allen,M.H., Barton,A., Band,G., Bellenguez,C., Bergboer,J.G., et al. (2010) A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. *Nat Genet*, **42**, 985-990.
72. Radstake,T.R., Gorlova,O., Rueda,B., Martin,J.E., Alizadeh,B.Z., Palomino-Morales,R., Coenen,M.J., Vonk,M.C., Voskuyl,A.E., Schuerwegh,A.J., et al. (2010) Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. *Nat Genet*, **42**, 426-429.

73. Bradfield,J.P., Qu,H.Q., Wang,K., Zhang,H., Sleiman,P.M., Kim,C.E., Mentch,F.D., Qiu,H., Glessner,J.T., Thomas,K.A., et al. (2011) A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. *PLoS Genet*, **7**, e1002293.
74. Jin,Y., Andersen,G., Yorgov,D., Ferrara,T.M., Ben,S., Brownson,K.M., Holland,P.J., Birlea,S.A., Siebert,J., Hartmann,A., et al. (2016) Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. *Nat Genet*, **48**, 1418-1424.
75. Yang,J., Lee,S.H., Goddard,M.E., Visscher,P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*, **88**, 76-82.
76. Lloyd-Jones,L.R., Holloway,A., McRae,A., Yang,J., Small,K., Zhao,J., Zeng,B., Bakshi,A., Metspalu,A., Dermitzakis,M., et al. (2017) The genetic architecture of gene expression in peripheral blood. *Am J Hum Genet*, **100**, 371.
77. Wray,N.R., Goddard,M.E., Visscher,P.M. (2007) Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res*, **17**, 1520-1528.
78. Wood,A.R., Esko,T., Yang,J., Vedantam,S., Pers,T.H., Gustafsson,S., Chu,A.Y., Estrada,K., Luan,J., Kutalik,Z., et al. (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet*, **46**, 1173-1186.
79. Bulik-Sullivan,B.K., Loh,P.R., Finucane,H.K., Ripke,S., Yang,J., Patterson,N., Daly,M.J., Price,A.L., Neale,B.M. (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*, **47**, 291-295.
80. Abecasis,G.R., Auton,A., Brooks,L.D., DePristo,M.A., Durbin,R.M., Handsaker,R.E., Kang,H.M., Marth,G.T., McVean,G.A. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56-65.
81. International HapMap 3 Consortium, Altshuler,D.M., Gibbs,R.A., Peltonen,L., Altshuler,D.M., Gibbs,R.A., Peltonen,L., Dermitzakis,E., Schaffner,S.F., Yu,F., et al. (2010) Integrating common and rare genetic variation in diverse human populations. *Nature*, **467**, 52-58.
82. Purcell,S., Cherny,S.S., Sham,P.C. (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*, **19**, 149-150.
83. Betz,R.C., Petukhova,L., Ripke,S., Huang,H., Menelaou,A., Redler,S., Becker,T., Heilmann,S., Yamany,T., Duvic,M., et al. (2015) Genome-wide meta-analysis in alopecia areata resolves HLA associations and reveals two new susceptibility loci. *Nat Commun*, **6**, 5966.
84. Cortes,A., Hadler,J., Pointon,J.P., Robinson,P.C., Karaderi,T., Leo,P., Cremin,K., Pryce,K., Harris,J., Lee,S., et al. (2013) Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat Genet*, **45**, 730-738.
85. Chu,X., Pan,C.M., Zhao,S.X., Liang,J., Gao,G.Q., Zhang,X.M., Yuan,G.Y., Li,C.G., Xue,L.Q., Shen,M., et al. (2011) A genome-wide association study identifies two new risk loci for Graves' disease. *Nat Genet*, **43**, 897-901.
86. Gutierrez-Achury,J., Zhernakova,A., Pulit,S.L., Trynka,G., Hunt,K.A., Romanos,J., Raychaudhuri,S., van Heel,D.A., Wijmenga,C., de Bakker,P.I. (2015) Fine mapping in the MHC region accounts for 18% additional genetic risk for celiac disease. *Nat Genet*, **47**, 577-578.
87. Goyette,P., Boucher,G., Mallon,D., Ellinghaus,E., Jostins,L., Huang,H., Ripke,S., Gusareva,E.S., Annese,V., Hauser,S.L., et al. (2015) High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. *Nat Genet*, **47**, 172-179.
88. Hinks,A., Cobb,J., Marion,M.C., Prahalad,S., Sudman,M., Bowes,J., Martin,P., Comeau,M.E., Sajuthi,S., Andrews,R., et al. (2013) Dense genotyping of immune-related disease regions identifies 14 new susceptibility loci for juvenile idiopathic arthritis. *Nat Genet*, **45**, 664-669.
89. Patsopoulos,N.A., Barcellos,L.F., Hintzen,R.Q., Schaefer,C., van Duijn,C.M., Noble,J.A., Raj,T., Gourraud,P.A., Stranger,B.E., Oksenberg,J., et al. (2013) Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. *PLoS Genet*, **9**, e1003926.

90. Tafti,M., Hor,H., Dauvilliers,Y., Lammers,G.J., Overeem,S., Mayer,G., Javidi,S., Iranzo,A., Santamaria,J., Peraita-Adrados,R., et al. (2014) DQB1 locus alone explains most of the risk and protection in narcolepsy with cataplexy in Europe. *Sleep*, **37**, 19-25.
91. Liu,J.Z., Almarri,M.A., Gaffney,D.J., Mells,G.F., Jostins,L., Cordell,H.J., Ducker,S.J., Day,D.B., Heneghan,M.A., Neuberger,J.M., et al. (2012) Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. *Nat Genet*, **44**, 1137-1141.
92. Liu,J.Z., Hov,J.R., Folseraas,T., Ellinghaus,E., Rushbrook,S.M., Doncheva,N.T., Andreassen,O.A., Weersma,R.K., Weismuller,T.J., Eksteen,B., et al. (2013) Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. *Nat Genet*, **45**, 670-675.
93. Okada,Y., Han,B., Tsoi,L.C., Stuart,P.E., Ellinghaus,E., Tejasvi,T., Chandran,V., Pellett,F., Pollock,R., Bowcock,A.M., et al. (2014) Fine mapping major histocompatibility complex associations in psoriasis and its clinical subtypes. *Am J Hum Genet*, **95**, 162-172.
94. Raychaudhuri,S., Sandor,C., Stahl,E.A., Freudenberg,J., Lee,H.S., Jia,X., Alfredsson,L., Padyukov,L., Klareskog,L., Worthington,J., et al. (2012) Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet*, **44**, 291-296.
95. Lessard,C.J., Li,H., Adrianto,I., Ice,J.A., Rasmussen,A., Grundahl,K.M., Kelly,J.A., Dozmorov,M.G., Miceli-Richard,C., Bowman,S., et al. (2013) Variants at multiple loci implicated in both innate and adaptive immune responses are associated with Sjogren's syndrome. *Nat Genet*, **45**, 1284-1292.
96. Kim,K., Bang,S.Y., Lee,H.S., Okada,Y., Han,B., Saw,W.Y., Teo,Y.Y., Bae,S.C. (2014) The HLA-DRbeta1 amino acid positions 11-13-26 explain the majority of SLE-MHC associations. *Nat Commun*, **5**, 5902.
97. Mayes,M.D., Bossini-Castillo,L., Gorlova,O., Martin,J.E., Zhou,X., Chen,W.V., Assassi,S., Ying,J., Tan,F.K., Arnett,F.C., et al. (2014) Immunochip analysis identifies multiple susceptibility loci for systemic sclerosis. *Am J Hum Genet*, **94**, 47-61.
98. Hu,X., Deutsch,A.J., Lenz,T.L., Onengut-Gumuscu,S., Han,B., Chen,W.M., Howson,J.M., Todd,J.A., de Bakker,P.I., Rich,S.S., et al. (2015) Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. *Nat Genet*, **47**, 898-905.

Legends to Figures

Fig 1. Prioritizing genes driving the pleiotropic association of rs1734907 in Crohn's disease and schizophrenia

Associations for SNP and SMR analyses across GWAS, eQTL, and mQTL datasets for two potentially pleiotropic loci, rs7132277 (a) and rs1734907 (b). Top plot gray circles illustrate SNP association (-$\log_{10}$ p-value) with schizophrenia in the PGC2 GWAS, while pink diamonds and blue circles indicate results of SMR tests (-$\log_{10}$ p-value) for association of gene expression and DNAm with schizophrenia, respectively, with solid shading indicating probes passing the HEIDI test. Middle plot illustrates SNP association (-$\log_{10}$ p-value) with gene expression from peripheral blood eQTL dataset. Lower plots illustrate SNP association (-$\log_{10}$ p-value) with gene methylation from peripheral blood mQTL dataset.

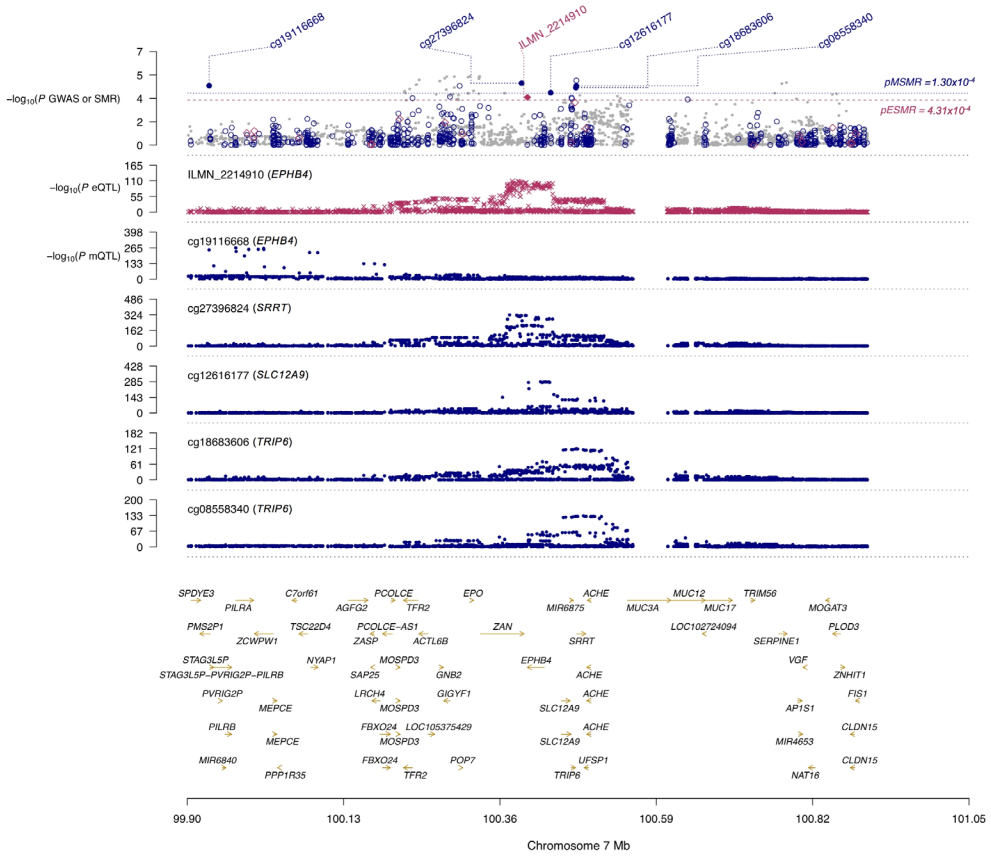
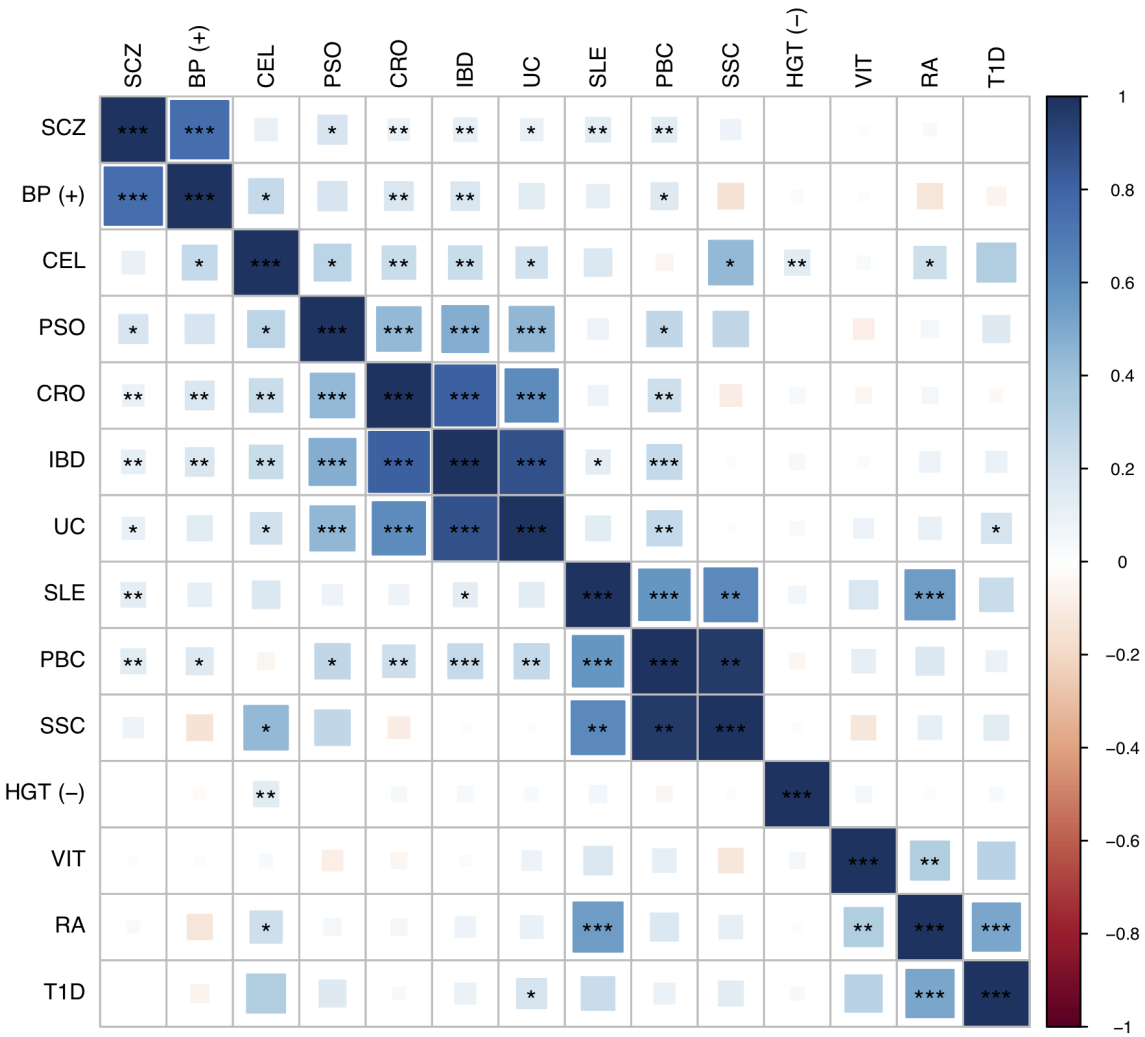


Fig 2. Genetic correlation between schizophrenia and other traits

Genetic correlation between schizophrenia, bipolar disorder, height, and 14 immune diseases was estimated using cross-trait LDSC (16). Colour intensity and size of square are proportional to strength of genetic correlation (red, negative correlation; blue, positive correlation). Asterisks indicate genetic correlations that are statistically significant at $p<0.05$ (*), $p<0.004$ (**), and $p<0.0002$ (***) thresholds. BP, bipolar disorder; CEL, celiac disease; CRO, Crohn's disease; HGT, height; IBD, inflammatory bowel disease; PBC, primary biliary cirrhosis; PSO, psoriasis, RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; SSC, systemic sclerosis; T1D, type 1 diabetes; UC, ulcerative colitis; VIT, vitiligo.



Tables

Table 1. Description of datasets analyzed

	Abr	Genome-wide significant SNPs[a]	Available GWAS summary statistics[b]	Cases	Controls	Number of SNPs with available summary statistics
Schizophrenia	SCZ	-	(1)	35,476	46,839	9,444,230
Bipolar disorder (+)	BP	-	(12)	6,990	4,820	1,233,534
Height (-)	HGT	-	(78)	253,288	-	2,085,602
Alopecia areata	AA	11	-	-	-	-
Ankylosing spondylitis	AS	23	-	-	-	-
Autoimmune thyroid disease	ATD	7	-	-	-	-
Celiac disease	CEL	38	(39)	4,533	10,750	523,398
Crohn's disease	CRO	119	(67)	5,956	14,927	12,276,506
Inflammatory bowel disease	IBD	145	(67)	12,882	21,770	12,716,150
Juvenile idiopathic arthritis	JIA	22	-	-	-	-
Multiple sclerosis	MS	103	-	-	-	-
Narcolepsy	NAR	3	-	-	-	-
Primary biliary cirrhosis	PBC	19	(70)	2,764	10,475	1,038,537
Primary sclerosing cholangitis	PSC	12	-	-	-	-
Psoriasis	PSO	34	(71)	2,178	5,175	7,586,779
Rheumatoid arthritis	RA	77	(68)	5,539	20,169	2,090,825
Sjögren's syndrome	SJO	6	-	-	-	-
Systemic lupus erythematosus	SLE	19	(69)	4,036	6,959	7,915,251
Systemic sclerosis	SSC	4	(72)	1,486[c]	3,477[c]	253,179
Type 1 diabetes	T1D	56	(73)	9,934	16,956	1,943,760
Ulcerative colitis	UC	96	(67)	6,968	20,464	12,255,263
Vitiligo	VIT	16	(74)	1,381	14,518	8,790,155

[a]We obtained lists of genome-wide significant SNPs for each autoimmune disease from ImmunoBase, and processed them as described in **Supplementary Methods**; [b]Because genome-wide summary statistics were required for the LDSC analysis, we were unable to estimate genetic

correlation with schizophrenia for eight autoimmune diseases for which these data were not available (AA, AS, ATD, JIA, MS, NAR, PSC, SJO); -, negative control; +, positive control; [c]only the US cohort from this study was available for analysis; Abr, abbreviation; -, not analyzed.

Table 2. Association of top HLA variants for immune diseases in schizophrenia

		Autoimmune		**Schizophrenia**		
Disease	HLA variant	*p*	OR	*p*	OR	r^2 with top SCZ SNP[a]
AA (83)	**HLA-DRB1#37Asn**	**$4.99x10^{-73}$**	**0.42**	**$4.85x10^{-9}$**	**0.91**	**0.04**
AS (84)	HLA-B*27	$<1x10^{-100}$	46	0.13	1.05	0
ATD (85)	rs2281388 (tags HLA-DPB1*05:01)	$1.50x10^{-65}$	1.64	0.39	1.04[b]	0
CEL (86)	**HLA-DQB1#74Ala**	**n.r.**	**2.14**	**$2.16x10^{-12}$**	**0.89**	**0.11**
CRO (87)	HLA-DRB1*01:03	$3.00x10^{-62}$	2.51	0.61	0.96	0
IBD (87)	HLA-DRB1*01:03	$1.93x10^{-112}$	3.01	0.61	0.96	0
JIA (88)	rs7775055	$3.14x10^{-174}$	6.01	0.12	0.94	0
MS (89)	HLA-DRB1*15:01	$1.40x10^{-234}$	2.92	$5.10x10^{-3}$	1.06	0
NAR (90)	HLA-DQB1*06:02	$1.04x10^{-120}$	251	$7.30x10^{-3}$	1.06	0
PBC (91)	HLA-DQA1*04:01	$5.90x10^{-45}$	3.06	0.20	0.95	0
PSC (92)	**HLA-B*08:01**	**$3.70x10^{-246}$**	**2.82**	**$5.65x10^{-16}$**	**0.84**	**0.2**
PSO (93)	HLA-C*06:02	$2.10x10^{-201}$	3.26	0.55	0.99	0
RA (94)	HLA-DRB1#11Val	$<1x10^{-581}$	3.80	$2.68x10^{-4}$	1.07	0
SJO (95)	**HLA-DQB1*02:01**	**$1.38x10^{-95}$**	**3.36**	**$3.84x10^{-15}$**	**0.85**	**0.11**
SLE (96)	HLA-DRB1#13Arg	$7.99x10^{-10}$	1.55[c]	$5.81x10^{-4}$	1.07	0
SSC (97)	rs17500468 (TAP2)	$5.87x10^{-62}$	2.87	$6.76x10^{-4}$	1.07	0
T1D (98)	HLA-DQB1#57Ala	$<1x10^{-1000}$	5.17	$7.80x10^{-4}$	0.95	0.06
UC (87)	rs6927022	$8.00x10^{-154}$	1.49	$3.37x10^{-4}$	1.06	0.03
VIT (74)	rs9271597 (4.7kb upstream of HLA-DQA1)	$3.15x10^{-89}$	1.77	0.01	1.04	0

[a]r^2 with rs1233578, the top HLA variant in schizophrenia, was obtained from the GAIN schizophrenia cohort (mgs2); [b]Effect size estimate is for HLA-DPB1*05:01; [c]Effect size estimate obtained from Asian sample. n.r., not reported; Disease abbreviations as defined in **Table 1**. Bold font indicates statistically significant association with schizophrenia.

Table 3. Immune disease risk SNPs showing potentially pleiotropic effects in schizophrenia

SNP (chr:bp)	Immune Disease	Risk/ Non-Risk Allele	Immune OR (95% CI);p[a]	Schizophrenia OR (95% CI);p	Nearby Genes	Colocalization[b]	Conditional Analysis[c]	eQTL[d]	mQTL[e]	Genomic associations co-localizing to this gene[f]
rs296547[e] (chr1:200892137)	CEL (39)	G/A	1.12 (1.09-1.16); 4.11×10^{-9}	1.04 (1.02-1.07); 6.17×10^{-5}	*CAMSAP2* *C1orf106* *KIF21B* *CACNA1S* *ASCL5*	$PP_{regional}$=0.06 $PP_{rs296547}<10^{-20}$	rs111530734, $p=1.2\times10^{-3}$	n.s.	*C1orf106*, decreased methylation	SCZ-mQTL
rs6738825 (chr2: 198896895)	CRO (36)	A/G	1.06 (1.02-1.11); 3.50×10^{-9}	1.05 (1.03-1.07); 3.02×10^{-6}	*SF3B1* *COQ10B* *HSPD1* *MOB4* *HSPE1* *RFTN2* *MARS2* *BOLL* *PLCL1*	$PP_{regional}$=0.57 $PP_{rs6738825}$=0.23	rs111744017, $p=8.0\times10^{-6}$	*PLCL1*, increased expression	*PLCL1*, decreased methylation *RFTN2*, decreased methylation	SCZ-mQTL, SCZ-eQTL SCZ-mQTL
rs13126505 (chr4:102865304)	CRO[f] (37)	A/G	1.17 (1.10-1.25); 2.33×10^{-10}	1.14 (1.10-1.19); 1.19×10^{-8}	*BANK1* *SLC39A8* *NFKB1*	$PP_{regional}$=0.99 $PP_{rs13126505}<10^{-100}$	rs35225200, $p=1.8\times10^{-3}$	*SLC39A8*, decreased expression	*SLC39A8*, increased methylation	SCZ-eQTL, SCZ-mQTL
rs1734907 (chr7:100315517)	CRO[f] (37)	A/G	1.16 (1.11-1.21); 1.67×10^{-13}	1.07 (1.04-1.10); 7.55×10^{-6}	*TFR2* *ACTL6B* *GNB2* *GIGYF1* *POP7* *EPO* *ZAN* *EPHB4* *SLC12A9*	$PP_{regional}$=0.05 $PP_{rs1734907l}<10^{-30}$	rs11539289, $p=1.8\times10^{-3}$	*EPHB4*, decreased expression *TRIP6*, decreased expression	*EPHB4*, increased methylation *TRIP6*, inconsistent effect across probes	SCZ-eQTL, SCZ-mQTL, eQTL-mQTL SCZ-eQTL, eQTL-mQTL
rs7132277 (chr12:123593382)	MS (38)	A/G	1.12 (n.r.); 1.90×10^{-13}	1.07 (1.04-1.09); 2.52×10^{-6}	*ABCB9* *ARL6IP4* *MIR4304* *OGFOD2* *PITPNM2* *MPHOSPH9* *CDK2AP1*	$PP_{regional}$=0.94 $PP_{rs7132277}<10^{-30}$	rs74240770, $p=1.0\times10^{-8}$	*ABCB9*, increased expression *ARL6IP4*, decreased expression	*ABCB9*, increased methylation	SCZ-eQTL, SCZ-mQTL, eQTL-mQTL

				SBNO1				
							MPHOSPH9, increased expression	
								PITPNM2, decreased methylation

[a]Effect sizes and p-values reported based on Immunobase curation, which reports statistics from meta-analysis of discovery and replication datasets where available; Results of Bayesian colocalization analyses using the coloc2 method (35), reported as posterior probabilities for each region of association (PP_{region}) and for each immune risk variant (PP_{rsid}); Results of conditional analyses using COJO (41), reported as the top SNP association remaining in the region after conditioning on the immune risk variant of interest. $p>8.6x10^{-5}$ indicates no significant associations remain after conditioning on the immune risk variant; [d]eQTL data was obtained from the CAGE study (76) which measured gene expression in peripheral blood. Effect on expression (increased/decreased) corresponds to the risk allele; [e]mQTL data was obtained from a meta-analysis of the Brisbane Systems Genetics Study (44) and Lothian Birth Cohorts of 1921 and 1936 (45), which measured DNA methylation in peripheral blood. Effect on expression (increased/decreased) corresponds to the risk allele; [f]Significant SMR and HEIDI (42, 43) results indicating co-localization of genomic associations with the gene of interest in schizophrenia-eQTL (SCZ-eQTL), schizophrenia-mQTL (SCZ-mQTL), and eQTL-mQTL (eQTL-mQTL) datasets; [e]eQTL data were unavailable for rs296547, and rs404339 was used as a proxy SNP (r^2=0.85 in 1000 Genomes Phase 3 CEU Population (40); [f]Also associated with inflammatory bowel disease; n.s., no statistically significant findings; Disease abbreviations as defined in **Table 1**.

Table 4. Estimated phenotypic and genome-wide genetic correlations between schizophrenia and other traits

			Total Sample		*Male-Specific*		*Female-Specific*	
Trait	$h^2 \pm SE$[a]	r_p	$r_g \pm SE$	p	$r_g \pm SE$	p	$r_g \pm SE$	p
BP (+)[b]	**0.46 ± 0.02**		**0.75 ± 0.05**	**$4.02x10^{-57}$**	**0.657 ± 0.04**	**$3.23x10^{-52}$**	**0.820 ± 0.06**	**$1.69x10^{-49}$**
HGT (-)	0.34 ± 0.02		$7.47x10^{-5}$ ± 0.02	0.99	-0.012 ± 0.02	0.53	0.012 ± 0.02	0.60
CEL	0.23 ± 0.05	0.04	0.107 ± 0.06	0.05	0.055 ± 0.06	0.34	0.075 ± 0.07	0.29
CRO	**0.37 ± 0.04**	**0.04**	**0.097 ± 0.03**	**$3.27x10^{-3}$**	**0.103 ± 0.03**	**$9.17x10^{-4}$**	**0.110 ± 0.04**	**$8.85x10^{-3}$**
IBD	**0.32 ± 0.04**	**n.a.**	**0.117 ± 0.03**	**$2.49x10^{-4}$**	**0.114 ± 0.03**	**$1.14x10^{-4}$**	**0.139 ± 0.04**	**$6.09x10^{-4}$**
PBC	**0.46 ± 0.08**	**0.11**	**0.131 ± 0.05**	**$4.00x10^{-3}$**	**0.141 ± 0.04**	**$1.50x10^{-3}$**	0.073 ± 0.05	0.15
PSO	**0.27 ± 0.09**	**0.13**	**0.182 ± 0.07**	**$7.80x10^{-3}$**	**0.205 ± 0.07**	**$4.10x10^{-3}$**	**0.212 ± 0.09**	**0.01**
RA	0.18 ± 0.03	-0.04	-0.032 ± 0.04	0.78	0.017 ± 0.04	0.70	-0.067 ± 0.05	0.19
SLE	**0.13 ± 0.05**	**0.05**	**0.130 ± 0.05**	**$3.76x10^{-3}$**	**0.153 ± 0.05**	**$1.51x10^{-3}$**	0.065 ± 0.05	0.22
SSC	0.26 ± 0.08	n.a.	0.086 ± 0.07	0.16	**0.190 ± 0.09**	**0.04**	0.011 ± 0.09	0.91

T1D	0.20 ± 0.04	0.15	-0.008 ± 0.05	0.86	0.041 ± 0.05	0.38	-0.013 ± 0.05	0.81
UC	**0.23 ± 0.03**	**-0.001**	**0.106 ± 0.04**	**$4.00x10^{-3}$**	**0.121 ± 0.04**	**$9.87x10^{-4}$**	**0.153 ± 0.05**	**$8.98x10^{-4}$**
VIT	0.86 ± 0.15	n.a.	0.011 ± 0.05	0.84	0.060 ± 0.05	0.23	0.045 ± 0.07	0.51

R^2 and h^2 are reported on the observed scale for all diseases; [a]h^2 was estimated using LDSC; [b]results reported are from previously published analyses by the Cross-Disorder Working Group of the Psychiatric Genomics Consortium (12); (+), positive control; (-), negative control; n.a., not available due to lack of data regarding incidence rate ratio of this immune disease in schizophrenia; SE, standard error; r_g, genetic correlation; r_p, expected phenotypic correlation based on epidemiological data (see **Materials and Methods** for details of r_p estimation).