Cross-disorder analysis of schizophrenia and 19 immune diseases reveals genetic correlation

Short Title: Genetic correlation between schizophrenia and immune diseases

Jennie G Pouget[1,2,3], Schizophrenia Working Group of the Psychiatric Genomics Consortium[^], Buhm Han[4,5,6,7], Emmanuel Mignot[8], Hanna M Ollila[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], Soumya Raychaudhuri[4,5,6,19,20,21¶*], Jo Knight[1,2,3,22,23¶*]

[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 Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA, [5]Partners HealthCare Center for Personalized Genetic Medicine, Boston, MA, USA, [6]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA, [7]Department of Convergence Medicine, University of Ulsan College of Medicine and Asan Institute for Life Sciences, Asan Medical Center, Seoul, Republic of Korea, [8]Center for Sleep Sciences and Medicine, Department of Psychiatry and Behavioral Sciences, Stanford University, School of Medicine, Palo Alto, CA, USA, [9]Division of Genetics and Molecular Medicine, Faculty of Life Sciences & Medicine, King's College London, UK, [10]St. John's Institute of Dermatology, Faculty of Life Sciences & 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 and [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]Division of Genetics, and [20]Division of Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA, [21]Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK, [22]Biostatistics Division, Dalla Lana School of Public Health, University of Toronto, ON, Canada, [23]Lancaster Medical School and Data Science Institute, Lancaster University, Lancaster, UK.

*Corresponding authors
Email: soumya@broadinstitute.org (SR)
Email: jo.knight@lancaster.ac.uk (JK)

[¶]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

Epidemiological studies have revealed that schizophrenia and immune diseases co-occur in the general population at higher than expected rates. Here, we evaluated whether the epidemiologic correlation between immune diseases and schizophrenia might be explained by shared genetic risk factors. We used data from a large genome-wide association study (GWAS) of schizophrenia (N=35,476 cases and 46,839 controls) 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. Next, we investigated genome-wide sharing of common variants using polygenic risk scores for immune diseases. We identified nine variants with potential pleiotropic effects, located in regions associated with both schizophrenia and autoimmune disease. Five of these variants were located outside of the human leukocyte antigen region, and mapped to genes with known roles in calcium signaling. Polygenic risk scores revealed significant genetic overlap with schizophrenia for narcolepsy ($p=4.1x10^{-4}$), primary biliary cirrhosis ($p=1.4x10^{-8}$), psoriasis ($p=3.6x10^{-5}$), systemic lupus erythematosus ($p=2.2x10^{-8}$), type 1 diabetes ($p=2.0x10^{-6}$), and ulcerative colitis ($p=4.3x10^{-4}$). Genetic correlation between these immune diseases and schizophrenia, estimated using cross-trait LD Score regression, ranged from 0.10 to 0.18. We also observed suggestive evidence of sex-dependent genetic correlation between schizophrenia and multiple sclerosis (interaction p=0.02), with genetic risk scores for multiple sclerosis associated with greater risk of schizophrenia among males but not females. Our findings reveal the presence of significant genetic correlation between schizophrenia and several immune diseases, which in some cases may be sex-dependent.

Author Summary

Immune diseases occur at different rates among patients with schizophrenia compared to the general population. While the reasons for this are unclear, shared genetic risk (**genetic correlation**) has been proposed as a contributing factor. Prior studies have used GWAS data to estimate the genetic correlation between schizophrenia and a handful of immune diseases, with conflicting results. Here, we performed a comprehensive cross-disorder investigation of schizophrenia and 19 immune diseases. We identified nine individual genetic variants associated with both schizophrenia and immune diseases, including four variants close to genes involved in calcium signalling (*PLCL1*, *BANK1*, EPO, *PITPNM2*). We demonstrated significant genome-wide genetic correlation between schizophrenia and narcolepsy, primary biliary cirrhosis, psoriasis, systemic lupus erythematosus, type 1 diabetes, and ulcerative colitis. Finally, we identified a potential sex-dependent pleiotropic effect between schizophrenia and multiple sclerosis, with genetic risk scores for multiple sclerosis associated with greater risk of schizophrenia among males but not females. Our findings point to a shared genetic basis between schizophrenia and at least a subset of immune diseases. These results raise the possibility that the same genetic variants may exert their effects on neurons or immune cells to influence the development of psychiatric and immune disorders, respectively.

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 an immune-mediated disease 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 discovered role of *complement component 4* (*C4*) variation in schizophrenia (9) provides a compelling illustration of a 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 increases the risk of developing schizophrenia, likely by increased tagging of neuronal synapses for elimination by microglia leading to excessive synaptic pruning (9). 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. Recently, 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 (12–17). While the effect sizes identified in cross-disorder studies are often modest, they 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 increased HDL cholesterol levels do not reduce the risk of myocardial infarction (14). In psychiatry cross-disorder analyses have identified significant pleiotropy 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). While previous studies have investigated genetic correlation between schizophrenia and immune disorders, results have been inconsistent (**S1 Table**). Genetic correlation has been reported between schizophrenia and Crohn's disease (18–20), multiple sclerosis (21), rheumatoid arthritis (18, 19), systemic lupus erythematosus (19, 20), type 1 diabetes (18), and ulcerative colitis (19, 20) in some studies, but not in others (8, 13, 16, 19, 22). Interestingly, negative genetic correlation (whereby genetic risk protects against developing schizophrenia) has also been reported for RA (23), in keeping with the inverse epidemiological association (5, 8). 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 the largest available genetic studies.

Results

Identification of immune risk variants

We identified immune diseases with robust genome-wide association study (GWAS) findings using ImmunoBase (http://www.immunobase.org), 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, IBD included the union of genetic variants associated with both CRO and UC. For 14 of these immune diseases (see **Table 1**), we also obtained full GWAS or Immunochip summary statistics allowing us to conduct additional polygenic risk scoring (PRS) (22, 24) and 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 diseases (25), 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 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**.

Genome-wide significant immune disease loci are associated with schizophrenia

To evaluate the association of immune risk variants with schizophrenia, 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 SCZ-52 study. 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. Five immune risk variants showed potential pleiotropic effects, with the risk allele for immune disease also conferring risk for schizophrenia (**Table 3**). These variants have been previously implicated in CRO (rs6738825, rs13126505, rs1734907 (26, 27)), MS (rs7132277 (28)), and CEL (rs296547 (29)). To evaluate the pleiotropic potential of these variants, we compared the genetic architecture of the region using regional plots (**S1 Fig**). In particular, the association signals for rs13126505 looked very similar in schizophrenia and CRO.

Next, 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. 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 $p>0.05$, **S2 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 429 LD-independent, non-HLA immune risk variants associated with at least one

of the 19 immune-mediated diseases. We found significant deviation from the theoretical null in schizophrenia for immune risk SNPs (λ=1.53). However, when we compared the association of immune risk SNPs to that of similar SNP sets (**Supplementary Methods**) we observed no evidence of enrichment (**S3 Fig**, p=0.66), indicating that immune risk SNPs were not associated with schizophrenia more than expected by chance given the polygenic nature of schizophrenia.

Finally, we evaluated the association of the top HLA variant implicated in each of the 19 immune diseases with schizophrenia. Notably, the association results for HLA variants were obtained using a refined subset of 31 cohorts from the SCZ-52 study with high-quality coverage of the MHC region (20,253 cases and 25,011 controls), as previously described (11). Overall, the effect sizes of these HLA variants were substantially greater for immune diseases (median OR = 2.92, minimum OR = 1.49 for UC, maximum OR = 251 for NAR) than schizophrenia (median OR = 1.06, minimum OR = 1.01, maximum OR = 1.19). We observed four HLA risk alleles associated with both immune disease and schizophrenia, particularly in the class II HLA region (**Table 2, Fig 1**). The four HLA variants showing potential pleiotropic effects in schizophrenia were the strongest HLA risk variants for alopecia (HLA-DRB1 #37 Asn), CEL (HLA-DQB1 #74 Ala), PSC (HLA-B*08:01), and SJO (HLA-DQB1*02:01). There was very low LD between these four HLA variants and rs1233578, the strongest associated variant in the region in schizophrenia (r^2=0.04 – 0.20, **Table 2**). The presence of HLA-DRB1 #37 Asn conferred a protective association in both alopecia and schizophrenia, but the remaining HLA variants showed the opposite direction of effect in schizophrenia compared to immune disease (**Table 2**, **Fig 1**).

Polygenic risk for six immune diseases is associated with schizophrenia

Our immune risk variant set captured only those variants associated with immune diseases at genome-wide significance in current GWASs. Given their polygenicity, there are

100s to 1,000s of additional variants associated with each immune disease which have not yet been identified (31). To evaluate sharing of risk alleles between immune diseases and schizophrenia using a broader set of variants, we used PRS (22, 24) and LDSC (16).

For each of the 14 immune diseases with available genome-wide summary statistics, we constructed genetic risk scores (GRSs) at a range of p-value thresholds (p_T) as in previous studies (12) and tested for the association of these GRSs with schizophrenia in a refined subset of the SCZ-52 study (17,000 cases and 20,655 controls) which excluded samples shared with the immune disease GWASs. To benchmark our findings in immune diseases, we also analyzed human height (32) and included previously published PRS results for bipolar disorder (12). We considered immune diseases with PRS $p<0.002$ at any p_T to show significant genetic overlap with schizophrenia (Bonferroni correction for 14 immune diseases tested in both sexes, 0.05/(14*2)=0.002). Importantly, commonly used goodness-of-fit estimates obtained from PRS (such as β_{GRS} and Nagelkerke's pseudo-R^2) lack meaningful interpretation and cannot be compared across studies (33). For these reasons we used β_{GRS} to interpret the direction of effect (i.e. positive or negative correlation), but not to interpret or compare the degree of genetic sharing between immune diseases and schizophrenia. For further details of our PRS approach, see **Materials and Methods**.

Using PRS, we had over 80% power to detect genetic covariance with schizophrenia ranging from 0.02-0.03 for most of the immune diseases, although some showed less than 80% power in this range (PSO, SLE, VIT; **S4 Fig**).

As previously described, bipolar disorder PRSs were significantly associated with schizophrenia ($p<1x10^{-50}$ at $p_T<1$) (12). Surprisingly, human height PRSs were also significantly associated with schizophrenia ($p=1x10^{-11}$ at $p_T<1$, **S2 Table**). Height was analyzed as a negative control based on its previously reported lack of genetic correlation with schizophrenia using LDSC (16). Using PRS, we observed that genetic liability for increased height protected

against schizophrenia (β_{GRS}=-0.11 at p_T<1). The significant inverse association of height PRSs with schizophrenia case-status we observed may reflect the greater sensitivity of this approach to subtle population stratification, sample sharing, and/or true genetic overlap.

Genetic scores including the HLA region were significant for CEL, NAR, PBC, PSO, RA, SLE, SSC, T1D, and UC (p<0.002 at multiple p_T, **S3 Table**). Height was not included in these analyses, given that HLA variants have not been previously reported in GWAS of this phenotype (32). With the exception of CEL (β_{GRS}≈-0.04 at p_T<$5x10^{-8}$, $1x10^{-4}$, and $1x10^{-3}$), all immune diseases exhibited a positive association with schizophrenia case-status (all β_{GRS}>0, **S3 Table**). For CEL, RA, SLE, and SSC only those PRSs constructed using the most stringent p-value cutoffs ($5x10^{-8}$, $1x10^{-4}$, $1x10^{-3}$) were significantly associated with schizophrenia. To evaluate whether the HLA region alone was driving the observed genetic sharing, we constructed PRSs excluding this region. After excluding HLA variants, genetic scores for NAR, PBC, PSO, SLE, T1D, and UC remained significantly associated with schizophrenia (**S5 Fig**, **S2 Table**). Because the genetic overlap between these six immune diseases and schizophrenia was not driven by a single HLA variant of large effect, we focused on these findings for the remainder of our analyses.

Given the potential sensitivity of PRS to artificial genetic overlap highlighted in our analysis of height, we wanted to assess whether cryptic sample sharing between the immune and schizophrenia GWASs could be driving the shared genetic liability that we observed. To do this, we conducted leave-half-out analyses. If the observed genetic overlap was driven by samples shared between certain schizophrenia cohorts and the immune disease GWASs, the PRS association should not be consistently observed across subsamples leaving out half of the schizophrenia cohorts. Across 1,000 subsamples (N_{cases} ranging from 3,985-13,074) leaving out a randomly selected 14 cohorts, we observed a high proportion of subsamples with PRSs significantly associated with schizophrenia (p<0.05 at p_T<1) for height (0.99), NAR (0.72), PBC

(0.95), PSO (0.84), SLE (0.97), T1D (0.95), and UC (0.70) suggesting our findings were not driven by sample sharing.

To further validate our PRS results we applied LDSC, an independent method for estimating genome-wide genetic correlation between traits that is robust to sample sharing (16), using summary statistics from the 49 European-ancestry cohorts in the SCZ-52 study (31,335 cases and 38,765 controls) (1). Unlike PRS, LDSC provides an interpretable and comparable estimation of genetic sharing between two traits in the form of genetic correlation (r_g) values. Notably, LDSC is less sensitive than PRS and is not robust when applied to genetic data obtained from specialty chips (e.g. Immunochip) (16). Given that this was a secondary analysis, we considered immune diseases with r_g $p<0.05$ to show significant genetic overlap with schizophrenia.

As expected, our positive control (bipolar disorder) showed significant genetic overlap with schizophrenia ($r_g=0.75\pm0.05$, $p=8.5\times10^{-60}$; **Fig 2, S4 Table**). In contrast to our PRS results, but in agreement with previous findings (16), our negative control (height) showed no such overlap using LDSC ($r_g=-0.004\pm0.02$, $p=0.84$; **Fig 2**, **S4 Table**). With respect to immune diseases, LDSC confirmed significant genetic overlap with schizophrenia for PBC, PSO, SLE, and UC ($r_g=0.10$-0.18, **Fig 2**, **S4 Table**) indicating the association of PRSs for these diseases was not driven by shared samples. Notably, genetic correlations for PSO and SLE did not survive correction for the 14 tests performed (**S4 Table**). We also observed significant genetic overlap with schizophrenia for NAR and T1D using LDSC, with the caveat that these datasets were genotyped using Immunochip and did not survive multiple testing correction (**Fig 2**, **S4 Table**). Overall, LDSC provided consistent results for the six immune diseases showing significant genetic sharing with schizophrenia by PRS. Interestingly, both CRO and IBD showed significant genetic correlation with schizophrenia using LDSC ($r_g=0.10$ and 0.12 for CRO and IBD, respectively; **Fig 2**, **S4 Table**), while PRS for these diseases were not significant (**S2**

278 **Table**). The genetic correlations observed between immune diseases and schizophrenia were
279 moderate, with r_g values about a fifth of that previously reported for bipolar disorder, and a
280 quarter of that for major depressive disorder (16).
281
282 **Exploration of sex-dependent genetic correlation between schizophrenia and immune**
283 **diseases**
284 Given the significant sex bias of autoimmune diseases, with women at greater risk
285 overall (34), we hypothesized that there may be sex-dependent genetic overlap between
286 schizophrenia and some immune-mediated diseases. We therefore performed sex-stratified
287 PRS, testing the association of height and immune disease GRSs with schizophrenia separately
288 in males and females of the SCZ-52 study. Genetic scores for height showed significant
289 association with schizophrenia in both males and females. Three of the immune diseases (PBC,
290 PSO, T1D) with significant main effects showed sex-dependent effects, with greater signal
291 among males (**S5 Table**). Additionally, although genetic scores for MS were not significantly
292 associated with schizophrenia in the total sample there was significant association among
293 males (R^2=0.03, p=1.26x10^{-3} at p_T<1; **S5 Table**).
294 Given the greater statistical power for the male subset of the schizophrenia GWAS, we
295 performed simulations by selecting random subsamples of male cases and controls equal in
296 size to the female sample (5,321 cases and 9,094 controls). If the stronger genetic overlap
297 between schizophrenia and MS, PBC, PSO, and T1D among males was driven by the larger
298 sample size rather than a true sex-dependent effect, there should be no consistent association
299 of PRSs with schizophrenia in these subsamples. Across 1,000 subsamples, the proportion with
300 significant PRS (p<0.002 at p_T<1) was high for PBC (0.94) and T1D (0.87), suggesting our
301 finding of a greater pleiotropic effect among males for these diseases was not driven solely by
302 lower statistical power among females; this was not the case for PSO (0.59) or MS (0.21).

Next, we performed formal statistical tests for an interaction between sex and genetic scores for these four immune diseases. We observed a nominally significant interaction for MS ($p<0.05$ at several p_T; **S5 Table**), noting that this finding did not survive correction for multiple testing. The remaining immune diseases did not show significant sex interactions, although the direction of effect was consistent with a greater pleiotropic effect in males (**S5 Table**).

Discussion

Using a variety of statistical approaches, we provide evidence of shared genetic risk for schizophrenia and several immune diseases. Outside of the MHC region, we identified five SNPs with potential pleiotropic effects - influencing risk for both immune disease and schizophrenia. Interestingly, nearby genes for four of the pleiotropic variants have been implicated in calcium signaling, suggesting this may be a shared risk pathway (12, 35). Parallel to its role in synaptic transmission, calcium signaling plays an important role in lymphocyte activation (36). However, we note that two of the pleiotropic variants are reportedly eQTLs for more distant genes not necessarily involved in calcium signaling, and further work is required to confirm their biological function.

Within the HLA region, we identified four potentially pleiotropic variants. An important caveat is that these four variants were not the top SNPs in their respective regions of association with schizophrenia, and were not primary drivers of the MHC association in schizophrenia in stepwise conditional analyses (9). Therefore, the biological significance of these particular HLA variants in schizophrenia is likely limited.

Using PRS we observed shared genetic liability with schizophrenia for six immune diseases (NAR, PBC, PSO, SLE, T1D, and UC), all of which have been previously reported to co-occur with schizophrenia (3, 5, 37). Thus, currently available genetic data suggest that shared genetic risk may contribute to the co-occurrence of these immune diseases in

schizophrenia - particularly PBC, PSO, SLE, and UC, which also showed robust genetic correlation with schizophrenia using LDSC. The magnitude of genetic correlations (r_g values) observed between these immune diseases and schizophrenia was about a fifth of that previously reported for bipolar disorder, and a quarter of that for major depressive disorder (16). Possible explanations for this sharing of genetic risk include the presence of a hidden subgroup of “autoimmune-like” schizophrenia cases and/or sharing of specific biological 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. We found nominal evidence of male-specific genetic correlation for MS, and a stronger pleiotropic effect among males for PBC, PSO, and T1D although the latter were not statistically significant. 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 (38), while androgens appear to protect against the development autoimmune diseases (39, 40). We emphasize that our sex-dependent findings require validation in independent samples. If replicated, one possibility is that sex hormones modulate pathogenesis among genetically vulnerable individuals, making males more likely to develop schizophrenia and females more likely to develop autoimmune diseases.

Our work was subject to several important limitations. First, PRSs for human height – analyzed as a negative control – showed stronger association with schizophrenia than any of the immune diseases. An inverse epidemiological relationship between height and schizophrenia has been reported (41, 42), consistent with our PRS findings. The reasons for the discrepancy between PRS and LDSC, which showed no genetic correlation between height and schizophrenia (as previously reported (16)) are unclear. One possibility is that PRS, which uses individual-level genotype data as opposed to summary statistics, is a more sensitive method to detect genome-wide sharing of risk alleles. If this is the case, it raises a broader question

354 regarding how much genetic overlap is expected across complex traits in general using the PRS
355 approach. An alternative explanation that must be considered is that PRS may be more
356 vulnerable to confounding by cryptic population stratification or sample sharing. Second,
357 genome-wide summary statistics were not available for all of the immune diseases, resulting in
358 a more limited analysis of 14 diseases. Furthermore, for five of these diseases (CEL, JIA, MS,
359 NAR, T1D) summary statistics were obtained from Immunochip rather than GWAS, providing
360 incomplete coverage of the genome for comparison with schizophrenia and biasing the genetic
361 correlation estimates obtained by LDSC.

362 Despite these limitations, our work adds to a growing body of evidence suggesting that
363 schizophrenia and immune diseases share genetic risk factors. There are conflicting reports in
364 the literature with respect to the specific immune diseases demonstrating genetic overlap with
365 schizophrenia, and the direction of effect (positive or negative genetic correlation). Genetic
366 overlap with schizophrenia has been previously investigated for nine of the immune diseases
367 studied here. Genome-wide genetic correlation with schizophrenia has been previously reported
368 for CRO (18–20), MS (21), PBC (20), PSO (20), RA (both positive (18, 19) and negative (23)
369 genetic correlations), SLE (19, 20), T1D (18), and UC (19, 20) (see **S1 Table** for a summary of
370 previous studies). Our results are consistent with previously reported genetic overlap between
371 schizophrenia and PBC (20), PSO (20), SLE (19, 20), T1D (18), and UC (19, 20). While we did
372 not observe genetic correlation between schizophrenia and MS in the total sample, there was a
373 significant sex-dependent effect with genetic correlation observed among males. We provide
374 new evidence of genetic correlation with NAR (not previously investigated). Notably, we did not
375 find any significant genetic correlation between schizophrenia and RA. Despite the robust
376 inverse epidemiological association between schizophrenia and RA (8), the genetic association
377 is less consistent. Using methods based on summary statistics (including PRS and LDSC), four
378 previous studies reported no evidence of pleiotropy between schizophrenia and RA (8, 16, 20,
379 22), while two studies reported positive genetic correlation (18, 19). 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 (23). Given the modest and potentially sex-dependent 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 that common genetic variants influencing the risk of immune diseases – in particular NAR, PBC, PSO, SLE, T1D, and UC – are also involved in schizophrenia. Studies identifying the cell types and biological pathways that may be driving this genetic overlap are underway, and will hopefully provide further insights into the pathophysiology of schizophrenia. In the meantime, our work provides further support for ongoing functional studies investigating parallel pathogenic mechanisms between the immune and the central nervous systems.

Materials and Methods

Samples and quality control

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

Schizophrenia dataset

We used data from the SCZ-52 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 or Immunochip summary statistics for 14 of the 19 immune diseases (five immune diseases were not included in PRS analyses due to lack of available summary statistics). We obtained publicly available summary statistics for ten immune diseases (see URLs): CEL (43), CRO (44), IBD (44), JIA (45), MS (28), NAR (46), RA (47), SLE (48), T1D (49), and UC (44). For the following four immune diseases, we obtained summary statistics with permission from the authors: PBC (50), PSO (51), SSC (52), and VIT (53).

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 (for details, see **Supplementary Methods**). Notably, we included the union of risk loci for CRO and UC as IBD loci. 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 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.

We evaluated the association of these 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.

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.

To evaluate the collective association of SNPs associated with any immune disease, we evaluated the p-values of a pruned set of 429 LD-independent, non-HLA immune risk SNPs in the schizophrenia dataset. We quantified enrichment of 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 minor allele frequency (MAF) and linkage disequilibrium (LD) score as these parameters are correlated with GWAS test statistics (see **Supplementary Methods** for details).

Testing the association of polygenic risk scores for 14 immune diseases in schizophrenia

To evaluate whether common variants influencing risk of immune diseases collectively contribute to schizophrenia, we used PRS (22, 24). To benchmark the amount of genetic overlap between schizophrenia and immune disease, we included previously published results for bipolar disorder as a positive control (12). We used human height (32) as a negative control because – despite the inverse epidemiological relationship between height and schizophrenia

455 previously reported (41, 42) – a prior study using cross-trait LDSC reported no genetic
456 correlation with schizophrenia (16).

457 For 14 immune diseases with available genome-wide summary statistics we performed
458 PRS at a range of p-value thresholds (p_T) as in previous studies (12): $5x10^{-8}$, $1x10^{-4}$, $1x10^{-3}$,
459 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, and 1.0 (which included all LD-independent SNPs, **Table 1**).
460 Due to extensive LD in the HLA region, we performed analyses both including the top HLA
461 variant and excluding the HLA region. At each p_T, we constructed PRSs for each individual *i* in
462 the schizophrenia cohort for each immune disease *h* by calculating the sum of risk-allele
463 dosages (*g*) weighted by their effect sizes (β) for that immune disease:

$$PRS_{i,h} = \sum_{M} \beta_{M,h} g_{M,i}$$

464 where *M* iterates over all known risk alleles for disease *h*, $\beta_{M,h}$ is the effect size (log odds ratio)
465 of *M* in disease *h*, and $g_{M,i}$ is the risk-allele dosage of *M* in individual *i*. We then performed
466 logistic regression in R (54) using the stats package (54) to evaluate the association between
467 schizophrenia case-status and PRSs for each immune disease. As in previous studies,
468 statistical significance of the PRSs was estimated based on their logistic regression coefficient
469 (12, 22). Variance in schizophrenia case-status explained by the PRSs was estimated using the
470 deviation in liability-scale R^2 between a null model (including 10 ancestry-informative principal
471 components and study site) and the full model (including PRSs in addition to these covariates),
472 calculated as previously described (33) assuming a population prevalence of schizophrenia of
473 1%. We also estimated Nagelkerke's pseudo-R^2 using the fmsb package (55). We considered
474 immune diseases with PRS $p<0.002$ at any p_T to show significant genetic overlap with
475 schizophrenia (Bonferroni correction for 14 immune diseases tested in both sexes,
476 0.05/(14*2)=0.002). As in previous studies (12, 22) we did not use Bonferroni correction for the
477 number of p-value thresholds, as these tests are highly correlated.

We excluded eight schizophrenia cohorts using Wellcome Trust Case Control Consortium (WTCCC) controls, due to the use of these samples in the immune disease GWASs. The total schizophrenia sample analyzed by PRS included 37,655 subjects (28 cohorts; 17,000 cases and 20,655 controls). Sex-stratified and formal sex-PRS interaction analyses were performed among the subset of subjects with known sex (9,787 male cases and 9,284 male controls; 5,231 female cases and 9,094 female controls). For details of PRS, see **Supplementary Methods** and **Table 1**.

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

To validate our PRS results and obtain genetic correlation (r_g) estimates, we performed a secondary analysis using cross-trait LDSC (16). This method estimates the genetic correlation between two traits using GWAS summary statistics. The method is computationally fast, and genetic correlation results are consistent with those obtained using individual-level data (16). The statistical framework 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 (56). 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 SNP*j*, ℓ_j is the LD score (56), 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 (57).

Because LDSC is robust to sample sharing across GWAS (16), we used summary statistics from the 49 European-ancestry cohorts in the SCZ-52 study (31,335 cases and 38,765 controls) (1). Similar to the PRS analyses described above, we benchmarked the genetic correlations observed for immune diseases by including bipolar disorder as a positive control (58) and human height (32) as a negative control. We used LD Scores from the "eur_w_ld_chr/" files available from https://data.broadinstitute.org/alkesgroup/LDSCORE, computed using 1000 Genomes Project (59) Europeans as a reference panel as previously described (57). To ensure we were using well-imputed SNPS we filtered all GWAS as previously described (16), including limiting the analysis to HapMap 3 (60) SNPs as implemented in the LDSC script munge_sumstats.py (https://github.com/bulik/ldsc). Given that this was a secondary analysis, we considered traits with r_g $p<0.05$ to have significant genetic correlation with schizophrenia.

Statistical power

Power to detect association of individual non-HLA and HLA immune risk variants in schizophrenia was calculated using the Genetic Power Calculator (61) assuming a risk allele frequency (RAF) of 0.05, disease prevalence of 1%, and significance threshold (α) of $8.6\text{x}10^{-5}$. Power for PRS was evaluated using AVENGEME (62, 63), assuming disease and genetic parameters detailed in **S6 Table**.

URLs

LD Score database:

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

GWAS summary statistics:

- CEL
  https://www.immunobase.org/downloads/protected_data/iChip_Data/hg19_gwas_ic_cel_trynka_4_19_1.tab.gz
- CRO, IBD, UC
  ftp://ftp.sanger.ac.uk/pub/consortia/ibdgenetics/iibdgc-trans-ancestry-filtered-summary-stats.tgz
- JIA
  https://www.immunobase.org/downloads/protected_data/iChip_Data/hg19_gwas_ic_jia_hinks_UK_4_19_1.tab.gz
- MS
  https://www.immunobase.org/downloads/protected_data/GWAS_Data/hg19_gwas_ms_imsgc_4_19_1.tab.gz
- NAR
  https://www.immunobase.org/downloads/protected_data/iChip_Data/hg19_gwas_ic_nar_faraco_4_19_1.tab.gz
- 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

- T1D

https://www.immunobase.org/downloads/protected_data/iChip_Data/hg19_gwas_ic_t1d_onengut_meta_4_19_1.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 International Multiple Sclerosis Genetics Consortium (IMSGC) and the investigators leading the GWASs for CEL, JIA, RA, SLE, and T1D 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 Asan Institute for Life Sciences, Asan Medical Center, Seoul, Republic of Korea (grant 2016-0717) and the Korean Health Technology R&D Project, Ministry of Health and Welfare, Republic of Korea (grant HI14C1731). HM Ollila is supported by the Finnish Cultural Foundation. J Martin is supported by grants SAF2015-66761-P from the Spanish Ministry of Economy and Competitiveness and P12-BIO-1395 from 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. J Knight is the Joanne Murphy Professor in Behavioural Science. 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.

References

1. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014;511:421-427.
2. Chan MK, Krebs MO, Cox D et al. Development of a blood-based molecular biomarker test for identification of schizophrenia before disease onset. *Transl Psychiatry* 2015;5:e601.
3. Benros ME, Nielsen PR, Nordentoft M, Eaton WW, Dalton SO, Mortensen PB. Autoimmune diseases and severe infections as risk factors for schizophrenia: a 30-year population-based register study. *Am J Psychiatry* 2011;168:1303-1310.
4. Chen SJ, Chao YL, Chen CY et al. Prevalence of autoimmune diseases in in-patients with schizophrenia: nationwide population-based study. *Br J Psychiatry* 2012;200:374-380.
5. Benros ME, Pedersen MG, Rasmussen H, Eaton WW, Nordentoft M, Mortensen PB. 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* 2013
6. Eaton WW, Byrne M, Ewald H et al. Association of schizophrenia and autoimmune diseases: linkage of Danish national registers. *Am J Psychiatry* 2006;163:521-528.
7. Eaton WW, Pedersen MG, Nielsen PR, Mortensen PB. Autoimmune diseases, bipolar disorder, and non-affective psychosis. *Bipolar Disord* 2010;12:638-646.
8. Euesden J, Breen G, Farmer A, McGuffin P, Lewis CM. The relationship between schizophrenia and rheumatoid arthritis revisited: genetic and epidemiological analyses. *Am J Med Genet B Neuropsychiatr Genet* 2015;168B:81-88.
9. Sekar A, Bialas AR, de Rivera H et al. Schizophrenia risk from complex variation of complement component 4. *Nature* 2016;530:177-183.
10. Yih Chen J, Ling Wu Y, Yin Mok M et al. 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* 2016;68:1442-1453.
11. Pouget JG, Goncalves VF, Spain SL et al. Genome-wide association studies suggest limited immune gene enrichment in schizophrenia compared to 5 autoimmune diseases. *Schizophr Bull* 2016;42:1176-1184.
12. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 2013;381:1371-1379.
13. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* 2013;45:984-994.
14. Voight BF, Peloso GM, Orho-Melander M et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* 2012;380:572-580.
15. Do R, Willer CJ, Schmidt EM et al. Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat Genet* 2013;45:1345-1352.
16. Bulik-Sullivan B, Finucane HK, Anttila V et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet* 2015;47:1236-1241.
17. Ellinghaus D, Jostins L, Spain SL et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet* 2016;48:510-518.
18. Stringer S, Kahn RS, de Witte LD, Ophoff RA, Derks EM. Genetic liability for schizophrenia predicts risk of immune disorders. *Schizophr Res* 2014;159:347-352.

19. Wang Q, Yang C, Gelernter J, Zhao H. Pervasive pleiotropy between psychiatric disorders and immune disorders revealed by integrative analysis of multiple GWAS. *Hum Genet* 2015;134:1195-1209.
20. Tylee DS, Sun J, Hess JL et al. Genetic correlations among psychiatric and immune-related phenotypes based on genome-wide association data. *bioRxiv* 2018;doi: https://doi.org/10.1101/070730
21. Andreassen OA, Harbo HF, Wang Y et al. Genetic pleiotropy between multiple sclerosis and schizophrenia but not bipolar disorder: differential involvement of immune-related gene loci. *Mol Psychiatry* 2015;20:207-214.
22. Purcell SM, Wray NR, Stone JL et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009;460:748-752.
23. Lee SH, Byrne EM, Hultman CM et al. New data and an old puzzle: the negative association between schizophrenia and rheumatoid arthritis. *Int J Epidemiol* 2015;44:1706-1721.
24. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res* 2007;17:1520-1528.
25. Fernando MM, Stevens CR, Walsh EC et al. Defining the role of the MHC in autoimmunity: a review and pooled analysis. *PLoS Genet* 2008;4:e1000024.
26. Franke A, McGovern DP, Barrett JC et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* 2010;42:1118-1125.
27. Jostins L, Ripke S, Weersma RK et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012;491:119-124.
28. Beecham AH, Patsopoulos NA, Xifara DK et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat Genet* 2013;45:1353-1360.
29. Dubois PC, Trynka G, Franke L et al. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet* 2010;42:295-302.
30. Zhu Z, Zheng Z, Zhang F et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun* 2018;9:224.
31. Stahl EA, Wegmann D, Trynka G et al. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat Genet* 2012;44:483-489.
32. Wood AR, Esko T, Yang J et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* 2014;46:1173-1186.
33. Lee SH, Goddard ME, Wray NR, Visscher PM. A better coefficient of determination for genetic profile analysis. *Genet Epidemiol* 2012;36:214-224.
34. Fairweather D, Frisancho-Kiss S, Rose NR. Sex differences in autoimmune disease from a pathological perspective. *Am J Pathol* 2008;173:600-609.
35. Purcell SM, Moran JL, Fromer M et al. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* 2014;506:185-190.
36. Feske S. Calcium signalling in lymphocyte activation and disease. *Nat Rev Immunol* 2007;7:690-702.
37. Canellas F, Lin L, Julia MR et al. Dual cases of type 1 narcolepsy with schizophrenia and other psychotic disorders. *J Clin Sleep Med* 2014;10:1011-1018.
38. Arad M, Weiner I. Disruption of latent inhibition induced by ovariectomy can be reversed by estradiol and clozapine as well as by co-administration of haloperidol with estradiol but not by haloperidol alone. *Psychopharmacology (Berl)* 2009;206:731-740.
39. Zhu ML, Bakhru P, Conley B et al. Sex bias in CNS autoimmune disease mediated by androgen control of autoimmune regulator. *Nat Commun* 2016;7:11350.
40. Markle JG, Frank DN, Mortin-Toth S et al. Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. *Science* 2013;339:1084-1088.

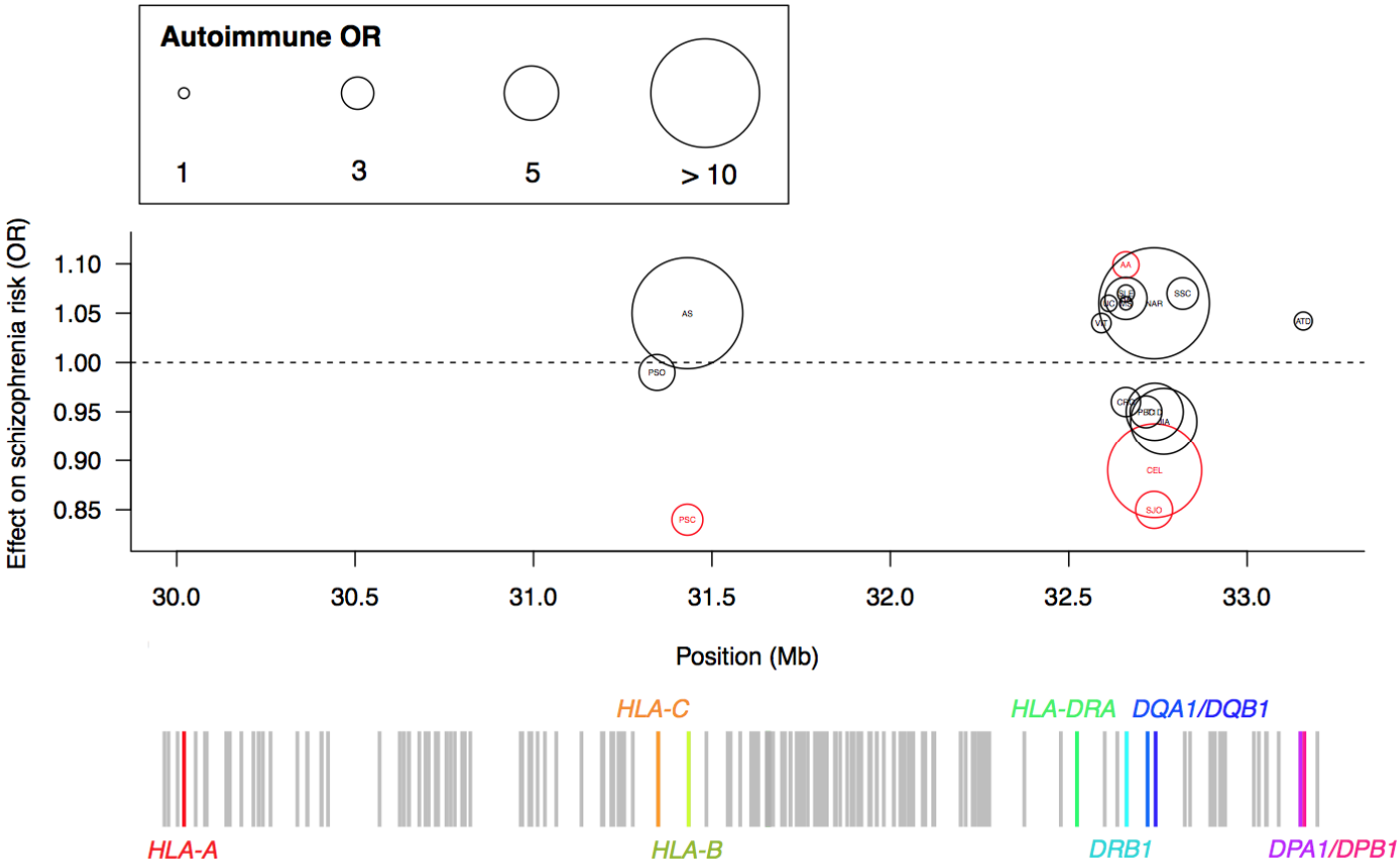
41. Zammit S, Rasmussen F, Farahmand B et al. Height and body mass index in young adulthood and risk of schizophrenia: a longitudinal study of 1 347 520 Swedish men. *Acta Psychiatr Scand* 2007;116:378-385.
42. Takayanagi Y, Petersen L, Laursen TM et al. Risk of schizophrenia spectrum and affective disorders associated with small for gestational age birth and height in adulthood. *Schizophr Res* 2014;160:230-232.
43. Trynka G, Hunt KA, Bockett NA et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet* 2011;43:1193-1201.
44. Liu JZ, van Sommeren S, Huang H et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* 2015;47:979-986.
45. Hinks A, Cobb J, Marion MC et al. Dense genotyping of immune-related disease regions identifies 14 new susceptibility loci for juvenile idiopathic arthritis. *Nat Genet* 2013;45:664-669.
46. Faraco J, Lin L, Kornum BR et al. ImmunoChip study implicates antigen presentation to T cells in narcolepsy. *PLoS Genet* 2013;9:e1003270.
47. Stahl EA, Raychaudhuri S, Remmers EF et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet* 2010;42:508-514.
48. Bentham J, Morris DL, Cunninghame Graham DS et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat Genet* 2015;47:1457-1464.
49. Onengut-Gumuscu S, Chen WM, Burren O et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat Genet* 2015;47:381-386.
50. Cordell HJ, Han Y, Mells GF et al. International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. *Nat Commun* 2015;6:8019.
51. Strange A, Capon F, Spencer CC et al. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. *Nat Genet* 2010;42:985-990.
52. Radstake TR, Gorlova O, Rueda B et al. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. *Nat Genet* 2010;42:426-429.
53. Jin Y, Andersen G, Yorgov D et al. Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. *Nat Genet* 2016;48:1418-1424.
54. R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing* 2012;Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.
55. Nakazawa M. fmsb: Functions for medical statistics book with some demographic data. *R package version 0.5.1.* 2014http://CRAN.R-project.org/package=fmsb.
56. Bulik-Sullivan BK, Loh PR, Finucane HK et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015;47:291-295.
57. Finucane HK, Bulik-Sullivan B, Gusev A et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* 2015;47:1228-1235.
58. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet* 2011;43:977-983.

59. Abecasis GR, Auton A, Brooks LD et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491:56-65.
60. International HapMap 3 Consortium, Altshuler DM, Gibbs RA et al. Integrating common and rare genetic variation in diverse human populations. *Nature* 2010;467:52-58.
61. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003;19:149-150.
62. Palla L, Dudbridge F. A fast method that uses polygenic scores to estimate the variance explained by genome-wide marker panels and the proportion of variants affecting a trait. *Am J Hum Genet* 2015;97:250-259.
63. Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet* 2013;9:e1003348.
64. Betz RC, Petukhova L, Ripke S et al. Genome-wide meta-analysis in alopecia areata resolves HLA associations and reveals two new susceptibility loci. *Nat Commun* 2015;6:5966.
65. Cortes A, Hadler J, Pointon JP et al. Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat Genet* 2013;45:730-738.
66. Chu X, Pan CM, Zhao SX et al. A genome-wide association study identifies two new risk loci for Graves' disease. *Nat Genet* 2011;43:897-901.
67. Gutierrez-Achury J, Zhernakova A, Pulit SL et al. Fine mapping in the MHC region accounts for 18% additional genetic risk for celiac disease. *Nat Genet* 2015;47:577-578.
68. Goyette P, Boucher G, Mallon D et al. 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* 2015;47:172-179.
69. Patsopoulos NA, Barcellos LF, Hintzen RQ et al. Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. *PLoS Genet* 2013;9:e1003926.
70. Tafti M, Hor H, Dauvilliers Y et al. DQB1 locus alone explains most of the risk and protection in narcolepsy with cataplexy in Europe. *Sleep* 2014;37:19-25.
71. Liu JZ, Almarri MA, Gaffney DJ et al. Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. *Nat Genet* 2012;44:1137-1141.
72. Liu JZ, Hov JR, Folseraas T et al. Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. *Nat Genet* 2013;45:670-675.
73. Okada Y, Han B, Tsoi LC et al. Fine mapping major histocompatibility complex associations in psoriasis and its clinical subtypes. *Am J Hum Genet* 2014;95:162-172.
74. Raychaudhuri S, Sandor C, Stahl EA et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet* 2012;44:291-296.
75. Lessard CJ, Li H, Adrianto I et al. Variants at multiple loci implicated in both innate and adaptive immune responses are associated with Sjogren's syndrome. *Nat Genet* 2013;45:1284-1292.
76. Kim K, Bang SY, Lee HS et al. The HLA-DRbeta1 amino acid positions 11-13-26 explain the majority of SLE-MHC associations. *Nat Commun* 2014;5:5902.
77. Mayes MD, Bossini-Castillo L, Gorlova O et al. Immunochip analysis identifies multiple susceptibility loci for systemic sclerosis. *Am J Hum Genet* 2014;94:47-61.
78. Hu X, Deutsch AJ, Lenz TL et al. Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. *Nat Genet* 2015;47:898-905.
79. Terunuma M, Jang IS, Ha SH et al. GABAA receptor phospho-dependent modulation is regulated by phospholipase C-related inactive protein type 1, a novel protein phosphatase 1 anchoring protein. *J Neurosci* 2004;24:7074-7084.

80. Lin X, Varnai P, Csordas G et al. Control of calcium signal propagation to the mitochondria by inositol 1,4,5-trisphosphate-binding proteins. *J Biol Chem* 2005;280:12820-12832.
81. Yokoyama K, Su Ih IH, Tezuka T et al. BANK regulates BCR-induced calcium mobilization by promoting tyrosine phosphorylation of IP(3) receptor. *EMBO J* 2002;21:83-92.
82. Mitkovski M, Dahm L, Heinrich R et al. Erythropoietin dampens injury-induced microglial motility. *J Cereb Blood Flow Metab* 2015;35:1233-1236.
83. Tong Q, Hirschler-Laszkiewicz I, Zhang W et al. TRPC3 is the erythropoietin-regulated calcium channel in human erythroid cells. *J Biol Chem* 2008;283:10385-10395.
84. Chang CL, Liou J. Phosphatidylinositol 4,5-bisphosphate homeostasis regulated by Nir2 and Nir3 proteins at endoplasmic reticulum-plasma membrane junctions. *J Biol Chem* 2015;290:14289-14301.
85. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015;348:648-660.
86. Westra HJ, Peters MJ, Esko T et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* 2013;45:1238-1243.

Figure Legends

Fig 1. Identification of pleiotropic HLA variants



The most strongly associated HLA variant for each of the 19 immune-mediated diseases was evaluated for association with schizophrenia, using summary statistics from HLA imputation and association testing in the schizophrenia dataset as previously described (9). x-axis corresponds to position in the classical HLA region, shown in megabase pairs (Mb); y-axis corresponds to the HLA variant's odds ratio (OR) in schizophrenia. Dashed line indicates OR=1. Size of the circle corresponds to the HLA variant's OR in the immune disease of interest, which is indicated in the circle center. Red circles indicate HLA variants associated with schizophrenia above the Bonferroni significance threshold ($p<8.6x10^{-5}$). Gene map below indicates location of coding HLA genes (coloured lines); gray lines correspond to non-HLA genes in the region. Disease abbreviations as defined in **Table 1**.

815 **Fig 2. Genetic correlation between schizophrenia and other traits**

816 Genetic correlation between schizophrenia, bipolar disorder, height, and 14 immune diseases
817 was estimated using cross-trait LDSC (16). Colour of square indicates strength of genetic
818 correlation (red, negative correlation; blue, positive correlation). Size of square indicates
819 statistical significance (larger, more significant *p*-value). Asterisks indicate genetic correlations
820 that are statistically significant at $p < 0.05$ threshold. BP, bipolar disorder; CEL, celiac disease;
821 CRO, Crohn's disease; HGT, height; IBD, inflammatory bowel disease; JIA, juvenile idiopathic
822 arthritis; MS, multiple sclerosis; NAR, narcolepsy; PBC, primary biliary cirrhosis; PSO, psoriasis,
823 RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; SSC, systemic sclerosis; T1D,
824 type 1 diabetes; UC, ulcerative colitis; VIT, vitiligo.

Tables

Table 1. Description of datasets analyzed

Disease	Abr	Genome-wide significant SNPs[a]	Polygenic risk scoring[b]	Cases	Controls	Total number of SNPs		
						Full GWAS	*Merged with SCZ*[c]	*Pruned*[d]
Schizophrenia	SCZ	-	Target (1)	35,476	46,839	-	-	-
Height	HGT	-	Negative control (32)	253,288	-	2,085,602	2,035,446	124,888
Alopecia areata	AA	11	-	-	-	-	-	-
Ankylosing spondylitis	AS	23	-	-	-	-	-	-
Autoimmune thyroid disease	ATD	7	-	-	-	-	-	-
Celiac disease	CEL	38	Training (43)	12,041	12,228	133,352	90,922	19,698
Crohn's disease	CRO	119	Training (44)	5,956	14,927	12,276,506	4,990,991	114,950
Inflammatory bowel disease	IBD	145	Training (44)	12,882	21,770	12,716,150	5,095,448	116,346
Juvenile idiopathic arthritis	JIA	22	Training (45)	772[e]	8,530[e]	122,330	98,477	20,337
Multiple sclerosis	MS	103	Training (28)	14,498	24,091	155,756	108,118	21,818
Narcolepsy	NAR	3	Training (46)	1,886	10,421	109,768	92,859	19,866
Primary biliary cirrhosis	PBC	19	Training (50)	2,764	10,475	1,038,537	1,041,977	97,806
Primary sclerosing cholangitis	PSC	12	-	-	-	-	-	-
Psoriasis	PSO	34	Training (51)	2,178	5,175	7,586,779	3,701,354	107,002
Rheumatoid arthritis	RA	77	Training (47)	5,539	20,169	2,090,825	2,087,383	126,049
Sjögren's syndrome	SJO	6	-	-	-	-	-	-
Systemic lupus erythematosus	SLE	19	Training (48)	4,036	6,959	7,915,251	6,539,217	264,374
Systemic sclerosis	SSC	4	Training (52)	1,486[f]	3,477[f]	253,179	251,441	66,402
Type 1 diabetes	T1D	56	Training (49)	9,340[g]	12,835	123,081	98,418	20,835
Ulcerative colitis	UC	96	Training (44)	6,968	20,464	12,255,263	5,167,266	120,720
Vitiligo	VIT	16	Training (53)	1,381	14,518	8,790,155	6,223,502	257,654

[a]We obtained lists of genome-wide significant SNPs for each autoimmune disease from ImmunoBase, and processed them as described in **Supplementary Methods**; [b]The following columns provide details for datasets used in the polygenic risk scoring analysis. We used effect sizes obtained from the height (negative control) and autoimmune disease GWASs (training datasets) to construct polygenic risk scores in the

schizophrenia sample (target dataset). Because genome-wide summary statistics were required for this analysis, we were unable to perform polygenic risk scoring for five autoimmune diseases for which these data were not available (AA, AS, ATD, PSC, SJO); [c]Prior to merging the training dataset SNP set with the target schizophrenia dataset SNP set, the following quality control steps were performed: SNPs on non-autosomal chromosomes (X, Y, M) were removed, SNPs with MAF<0.01 were removed if MAF was available in the training dataset, SNPs with INFO<0.90 were removed if INFO was available in the training dataset, SNPs with missing p-value or OR were removed, symmetrical SNPs were removed; [d]Pruning was performed by clumping using PLINK to retain SNPs with r^2<0.1 within 1,000 kb windows, while filtering for the highest significance levels within LD blocks (using options --clump-p1 1 --clump-p2 1 --clump-r2 0.1 --clump-kb 1000); [e]only the UK cohort from this study was available for analysis; [f]only the US cohort from this study was available for analysis; [g]includes cases from 2,601 affected sibling pairs and 69 trios, which were analyzed using the Generalized Disequilibrium Test (GDT) method and combined with case-control results by meta-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 (64)	**HLA-DRB1#37Asn**	**$4.99x10^{-73}$**	**0.42**	**$4.85x10^{-9}$**	**0.91**	**0.04**
AS (65)	HLA-B*27	$<1x10^{-100}$	46	0.13	1.05	0
ATD (66)	rs2281388 (tags HLA-DPB1*05:01)	$1.50x10^{-65}$	1.64	0.39	1.04[b]	0
CEL (67)	**HLA-DQB1#74Ala**	**n.r.**	**2.14**	**$2.16x10^{-12}$**	**0.89**	**0.11**
CRO (68)	HLA-DRB1*01:03	$3.00x10^{-62}$	2.51	0.61	0.96	0
IBD (68)	HLA-DRB1*01:03	$1.93x10^{-112}$	3.01	0.61	0.96	0
JIA (45)	rs7775055	$3.14x10^{-174}$	6.01	0.12	0.94	0
MS (69)	HLA-DRB1*15:01	$1.40x10^{-234}$	2.92	$5.10x10^{-3}$	1.06	0
NAR (70)	HLA-DQB1*06:02	$1.04x10^{-120}$	251	$7.30x10^{-3}$	1.06	0
PBC (71)	HLA-DQA1*04:01	$5.90x10^{-45}$	3.06	0.20	0.95	0
PSC (72)	**HLA-B*08:01**	**$3.70x10^{-246}$**	**2.82**	**$5.65x10^{-16}$**	**0.84**	**0.2**
PSO (73)	HLA-C*06:02	$2.10x10^{-201}$	3.26	0.55	0.99	0
RA (74)	HLA-DRB1#11Val	$<1x10^{-581}$	3.80	$2.68x10^{-4}$	1.07	0
SJO (75)	**HLA-DQB1*02:01**	**$1.38x10^{-95}$**	**3.36**	**$3.84x10^{-15}$**	**0.85**	**0.11**
SLE (76)	HLA-DRB1#13Arg	$7.99x10^{-10}$	1.55[c]	$5.81x10^{-4}$	1.07	0
SSC (77)	rs17500468 (TAP2)	$5.87x10^{-62}$	2.87	$6.76x10^{-4}$	1.07	0
T1D (78)	HLA-DQB1#57Ala	$<1x10^{-1000}$	5.17	$7.80x10^{-4}$	0.95	0.06
UC (68)	rs6927022	$8.00x10^{-154}$	1.49	$3.37x10^{-4}$	1.06	0.03
VIT (53)	rs9271597 (4.7kb upstream of HLA-DQA1)	$3.15x10^{-89}$	1.77	0.01	1.04	0

[a]r^2 with rs12333578, 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. AA, alopecia areata; AS, ankylosing spondylitis; ATD, autoimmune thyroid disease; CEL, celiac disease; CRO, Crohn's disease; IBD, inflammatory bowel disease; JIA, juvenile idiopathic arthritis; MS, multiple sclerosis; NAR, narcolepsy; n.r., not reported; PBC, primary biliary cirrhosis; PSC, primary sclerosing cholangitis; PSO, psoriasis, RA, rheumatoid arthritis; SJO, Sjögren's syndrome; SLE, systemic lupus erythematosus; SSC, systemic sclerosis; T1D, type 1 diabetes; UC, ulcerative colitis; VIT, vitiligo.

Table 3. Immune disease risk SNPs showing pleiotropic effect in schizophrenia

SNP (chr:bp)	Immune Disease	Risk Allele/ Non-Risk Allele	Immune OR (95% CI); p[a]	Schizophrenia OR (95% CI); p	Closest Gene	eQTL[b]	Function
rs296547 (chr1: 200892137)	CEL (29)	G/A	1.12 (1.09-1.16); $4.11x10^{-9}$	1.04 (1.02-1.07); $6.17x10^{-5}$	*C1orf106* (7.3kb 3')	*C1orf106*, inconsistent direction of effect across tissues Other genes	Unknown function
rs6738825 (chr2: 198896895)	CRO (26)	A/G	1.06 (1.02-1.11); $3.50x10^{-9}$	1.05 (1.03-1.07); $3.02x10^{-6}$	*PLCL1* (intronic)	*PLCL1*, decreased expression *RFTN2*, decreased expression	Regulates $GABA_A$ receptor signaling (79), inhibits IP_3 mediated calcium signalling (80)
rs13126505 (chr4: 102865304)	CRO[b] (27)	A/G	1.17 (1.10-1.25); $2.33x10^{-10}$	1.14 (1.10-1.19); $1.19x10^{-8}$	*BANK1* (intronic)	*BANK1*, decreased expression[c]	B-cell-specific scaffold protein that mediates receptor-induced calcium mobilization from intracellular stores (81)
rs1734907 (chr7: 100315517)	CRO[b] (27)	A/G	1.16 (1.11-1.21); $1.67x10^{-13}$	1.07 (1.04-1.10); $7.55x10^{-6}$	*EPO* (2.9kb 5')	*EPHB4*, decreased expression ***GIGYF1*, increased expression** Other genes	Found in plasma, regulates red cell production by promoting erythroid differentiation and initiating hemoglobin synthesis; neuroprotective activity (82); increases intracellular calcium concentration (83)
rs7132277 (chr12: 123593382)	MS (28)	A/G	1.12 (n.r.); $1.90x10^{-13}$	1.07 (1.04-1.09); $2.52x10^{-6}$	*PITPNM2* (intronic)	*ABCB9*, increased expression Other genes	Replenishes PIP_2 in plasma membrane, the major substrate for IP_3-related calcium channel activation (84)

[a]Effect sizes and p-values reported based on Immunobase curation, which reports statistics from meta-analysis of discovery and replication datasets where available; [b]cis-eQTL data from The GTEx Consortium (85) (http://gtexportal.org) and Westra *et al.* (86) (http://www.genenetwork.nl/bloodeqtlbrowser/); all genes with FDR<0.05 are listed, effect on expression (increased/decreased) corresponds to the risk allele; [b]Also associated with inflammatory bowel disease; [c]eQTL results presented are for proxy SNP rs13127398, r^2=0.89 with rs13126505; Disease abbreviations as defined in **Table 1**; bold fond indicates brain eQTLs.