Title
Genetic pleiotropy between multiple sclerosis and schizophrenia but not bipolar disorder: Differential involvement of immune-related gene loci

Permalink
https://escholarship.org/uc/item/39p9n9q8

Journal
Molecular Psychiatry, 20(2)

ISSN
1359-4184

Authors
Andreassen, OA
Harbo, HF
Wang, Y
et al.

Publication Date
2015

DOI
10.1038/mp.2013.195

Peer reviewed
ORIGINAL ARTICLE
Genetic pleiotropy between multiple sclerosis and schizophrenia but not bipolar disorder: differential involvement of immune-related gene loci

OA Andreassen1,2,3, HF Harbo4, Y Wang1,2,5,6, WK Thompson3, AJ Schork7,8, M Mattingstall1,9, V Zuber1,2,10, F Bettella1,2, S Ripke11,12, J Kelsse9, KS Kendler13, M CO14, P Sklar15, The Psychiatric Genomics Consortium (PGC) Bipolar Disorder and Schizophrenia Work Groups16, The International Multiple Sclerosis Genetics Consortium (IMSGC), LK McEvoy5,17, RS Desikan5,17, BA Lie18, S Djurovic1,2,18 and AM Dale3,5,6,17

Converging evidence implicates immune abnormalities in schizophrenia (SCZ), and recent genome-wide association studies (GWAS) have identified immune-related single-nucleotide polymorphisms (SNPs) associated with SCZ. Using the conditional false discovery rate (FDR) approach, we evaluated pleiotropy in SNPs associated with SCZ (n = 21 856) and multiple sclerosis (MS) (n = 43 879), an inflammatory, demyelinating disease of the central nervous system. Because SCZ and bipolar disorder (BD) show substantial clinical and genetic overlap, we also investigated pleiotropy between BD (n = 16 731) and MS. We found significant genetic overlap between SCZ and MS and identified 21 independent loci associated with SCZ, conditioned on association with MS. This enrichment was driven by the major histocompatibility complex (MHC). Importantly, we detected the involvement of the same human leukocyte antigen (HLA) alleles in both SCZ and MS, but with an opposite directionality of effect of associated HLA alleles (that is, MS risk alleles were associated with decreased SCZ risk). In contrast, we found no genetic overlap between BD and MS. Considered together, our findings demonstrate genetic pleiotropy between SCZ and MS and suggest that the MHC signals may differentiate SCZ from BD susceptibility.


Keywords: false discovery rate; HLA region; multiple sclerosis; polygenic pleiotropy; schizophrenia

INTRODUCTION
Schizophrenia (SCZ) and Bipolar Disorder (BD) are severe mental disorders, which are among the leading causes of disability globally. These disorders have a substantial impact on the quality of life for patients and their families, and are among the most costly societal disorders. Clinical, epidemiological and genetic findings suggest shared risk factors between BD and SCZ, and a Psychosis Continuum Model has been suggested. Despite the high heritability, most of the genetic architecture underlying susceptibility to both SCZ and BD remains to be defined, and the pathological mechanisms underlying these disorders are still largely unknown. Improved understanding of disease pathobiology and genetic risk factors may lead to major health benefits through the development of new treatment and prevention regimens.

An interesting hypothesis regarding SCZ pathology is the involvement of the immune system, which is derived from epidemiological and clinical evidence, implicating infections and cytokine abnormalities in the development of SCZ. Recent genome-wide association studies (GWAS) have robustly identified markers in the major histocompatibility complex (MHC) associated with SCZ, showing association of specific human leukocyte antigen (HLA) alleles. Although this may further support a role of the immune system in SCZ pathogenesis, it is also possible that this association reflects nonimmunological factors as several loci within the MHC are involved in neurobiological processes (for example, NOTCH4, TRIM26). With regard to BD, despite a large degree of overlapping clinical characteristics, genetic factors as well as cytokine abnormalities between BD and SCZ, there is little evidence for MHC associations in GWAS of BD.

1NORMENT, KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, Oslo, Norway; 2Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway; 3Departments of Psychiatry, University of California, San Diego, La Jolla, CA, USA; 4Department of Neurology, Oslo University Hospital, Ullevål, and Institute of Clinical Medicine, University of Oslo, Oslo, Norway; 5Multimodal Imaging Laboratory, University of California at San Diego, La Jolla, CA, USA; 6Department of Psychiatry, University of California, San Diego, La Jolla, CA, USA; 7Cognitive Sciences Graduate Program, University of California, San Diego, La Jolla, CA, USA; 8Center for Human Development, University of California, San Diego, La Jolla, CA, USA; 9Sollandet Hospital, Kristiansand, Norway; 10Centre for Molecular Medicine Norway, Nordic EMBL Partnership, University of Oslo and Oslo University Hospital, Oslo, Norway; 11Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 12Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 13Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA; 14MRC Centre for Neuropsychiatric Genetics and Genomics, School of Medicine, Cardiff University, Heath Park, Cardiff, UK; 15The Division of Psychiatric Genetics and Genomics, Mount Sinai School of Medicine, New York, NY, USA; 16PGC co-authors are listed separately in Supplementary Information; 17Department of Radiology, University of California, San Diego, La Jolla, CA, USA and 18Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway. Correspondence: Dr OA Andreassen, NORMENT, KG Jebsen Centre for Psychosis Research, Building 49, Oslo University Hospital, Ullevål, Kirkeveien 166, PO Box 4956 Nydalen, 0424 Oslo, Norway or Dr AM Dale, Department of Radiology, University of California, San Diego, 8950 Villa La Jolla Drive, Suite C101, La Jolla, CA 92037-0841, USA.
analysis of the shared genetic basis of these disorders revealed significant genetic overlap, but no common loci in the MHC. Together, this suggests differences in MHC-related genes between the two disorders.

Multiple sclerosis (MS) is a disease of the central nervous system in which inflammatory processes have an important role and associations with HLA class I and II loci are well established. The HLA-DRB1*1501 allele shows strong association with MS susceptibility in most populations with an average odds ratio (OR) of 3. A recent GWAS in MS identified more than 50 non-MHC regions associated with MS risk, and immunologically relevant genes were significantly over-represented in these regions. Further studies and meta-analyses have increased the number of detected MS risk variants. The immunological mechanisms involved in MS leads to various degrees of demyelination, and several lines of evidence point to oligodendrocyte and myelin dysfunction also in SCZ. This suggests shared risk factors between MS and SCZ. Given the overlapping phenotypes and polygenic architecture between SCZ and BD, genetic pleiotropy between BD and MS may also exist.

Combining GWAS from two disorders provides increased power to detect significant associations, and recent studies have reported overlapping single-nucleotide proteins (SNPs) between several human traits and disorders. Here, we use MS GWAS data from the International MS Genetics Consortium (IMSGC) together with SCZ GWAS and BD GWAS data to investigate pleiotropic relationships between these disorders. To date, methods to assess genetic pleiotropy have not taken full advantage of the existing GWAS data, and the majority of these studies have focused on the subset of SNPs exceeding a Bonferroni-corrected threshold of significance for each trait or disorder. However, using this approach, SNPs with nonsignificant effects in each phenotype, significant in the combined analysis (polygenic pleiotropy), cannot be detected. We recently developed a novel analytic strategy that uses all available SNPs in two or more independent GWAS to identify polygenic pleiotropy and improve gene discovery. Using this approach, we demonstrated extensive polygenic pleiotropy between SCZ and BD, and between SCZ and cardiovascular disease risk factors (CVD), and substantially increased the number of identified SCZ genetic susceptibility loci. Here, we applied these methods to independent large MS, SCZ and BD GWAS data to determine whether MS shares susceptibility loci with these psychiatric disorders.

**MATERIALS AND METHODS**

**Participant samples**

We used summary statistics from a large MS GWAS study performed by IMSGC, n = 27 148, and from two large GWAS studies from the Psychiatric GWAS Consortium (PGC), PGC Schizophrenia sample, n = 21 856, PGC Bipolar disorder sample, n = 16 731. P-values and minor allele frequencies from the discovery samples were included in the analyses. For follow-up analysis, we also investigated the PGC major depressive disorder, autism spectrum disorder, and attention deficit/hyperactivity disorder GWAS summary statistics (for additional details, see Supplementary Information).

**Statistical analyses**

Conditional Q-Q plots for pleiotropic enrichment. To visually assess pleiotropic enrichment, we used Q-Q plots conditioned on 'pleiotropic' effects (Figures 1a and 2a for BD). For a given associated phenotype, pleiotropic 'enrichment' exists if the degree of deflection from the expected null line is dependent on associations with the second phenotype (for further details see Supplementary Information). We constructed conditional Q-Q plots of empirical quantiles of nominal $-\log_{10}(P)$ values for all SNPs and for subsets of SNPs determined by the significance of their association with MS. Specifically, we computed the empirical cumulative distribution function (ecdf) of nominal $P$-values for a given phenotype for all SNPs and for SNPs with significance levels below the indicated cutoffs for the other phenotype ($-\log_{10}(P) \geq 0$, $-\log_{10}(P) > 1$, $-\log_{10}(P) > 2$, $-\log_{10}(P) > 3$) corresponding to $P < 0.1$, $P < 0.01$, $P < 0.001$, respectively. Nominal $P$-values ($-\log_{10}(P)$) are plotted on the y-axis, and empirical quantiles ($-\log_{10}(q)$, where $q = 1 - ecdf(P)$) are plotted on the x-axis. To assess pleiotropic effects below the standard GWAS significance threshold, we focused the Q-Q plots on SNPs with nominal $-\log_{10}(P) < 7.3$ (corresponding to $P > 5 \times 10^{-8}$). The same procedure was used for BD. The 'enrichment' seen in the conditional Q-Q plots can be directly interpreted in terms of the discovery rate (DTR = 1 - FDR). This is illustrated in Figures 1b and 2b for each range of $P$-values in the pleiotropic traits (see Supplementary Information for details).

**Conditional replication rate.** For each of the 17 substudies contributing to the final meta-analysis in SCZ, we independently adjusted the z-scores using an inflation factor. We randomly sampled 1000 combinations of eight and nine substudy groupings. We then calculated the eight-or-nine-study combined discovery z-score and eight-or-nine-study combined replication z-score for each SNP as the average z-score across the substudies multiplied by the square root of the number of studies. For discovery samples, the z-scores were converted to two-tailed $P$-values, whereas replication samples were converted to one-tailed $P$-values preserving the direction of effect in the discovery sample. For each of the 1000 discovery-replication pairs, cumulative rates of replication were computed over 1000 equally spaced bins spanning the range of $-\log_{10}(P$)-values observed in the discovery samples. The cumulative replication rate for any bin was the proportion of SNPs with a $-\log_{10}$ (discovery $P$-value) greater than the lower bound of the bin with a replication $P$-value < 0.05 and the same sign as the discovery sample. Cumulative replication rates were calculated independently for each of the four pleiotropic enrichment categories. For each category, the cumulative replication rate for each bin was averaged across the 1000 discovery-replication pairs, and the results are reported in Figure 1c. The vertical intercept in the figure is the overall replication rate.

**Conditional replication effect size.** Using the same z-score adjustment scheme and sampling method used for estimating cumulative replication rates (see above), we directly evaluated the relationship of replication effect size of the discovery sample versus replication samples (Figure 1d) for each SNP. The effect sizes were conditioned on various enrichment categories. For visualization, we fitted a cubic spline relating the bin midpoint of z-scores of discovery to the corresponding average replication z-scores (Figure 1d).

**Improving discovery of SNPs in SCZ and BD using conditional FDR.** To improve detection of SNPs associated with SCZ and BD, we employed a genetic epidemiology approach, leveraging the MS phenotype from an independent GWAS using conditional FDR as outlined in Andreassen et al. Specifically, conditional FDR is defined as the posterior probability that a given SNP is null for the first phenotype given that the $P$-values for both phenotypes are as small as or smaller than their observed $P$-values. We assigned a conditional FDR value for each SNP in SCZ given the $P$-value in MS (denoted as FDR$_{SCZ|MS}$). All SNPs with conditional FDR < 0.05 (−$\log_{10}(\text{FDR}_{SCZ|MS})$) in SCZ given the association with MS are listed in Supplementary Table 1. The same procedure was applied to compute FDR$_{BD|MS}$ for each SNP. To display the localization of the genetic markers associated with SCZ and BD given the MS effect, we used a 'Conditional Manhattan plot', plotting all SNPs within an LD block in relation to their chromosomal location. As illustrated for SCZ in Figure 3, the large points represent the significant SNPs (−$\log_{10}(\text{FDR}_{SCZ|MS})$ ≥ 1.3 equivalent to FDR$_{SCZ|MS}$ < 0.05), whereas the small points represent nonsignificant SNPs. All SNPs are shown without 'pruning' (that is, without removing all SNPs with $r^2$ > 0.2 based on 1000 Genome Project (1KGP) linkage disequilibrium (LD) structure). The strongest signal in each LD block is illustrated with a black line around the circles. This was identified by ranking all SNPs in increasing order, based on the FDR$_{SCZ|MS}$ value and then removing SNPs in LD $r^2$>0.2 with any higher-ranked SNP. Thus, the selected locus was the most significantly associated with SCZ in each LD block.

**Annotation of novel loci.** On the basis of 1KGP linkage disequilibrium (LD) structure, significant SNPs identified by conditional FDR were clustered into LD blocks at the LD $r^2$<0.2 level. These blocks are numbered (focus #) in Tables 1 and 2. Any block may contain more than one SNP. Genes close to each SNP were obtained from the NCBI gene database. Only blocks that did not contain previously reported SNPs or genes related to previously reported SNPs were deemed as novel loci in the current study (Tables 1 and 2). Loci that contained either SNPs or genes known to be associated with SCZ were considered as replication findings.
Figure 1. Genetic pleiotropy enrichment of schizophrenia (SCZ) conditional on MS. (a) Conditional Q-Q plot of nominal versus empirical $-\log_{10} P$-values (corrected for inflation) in SCZ below the standard GWAS threshold of $P < 5 \times 10^{-8}$ as a function of significance of association with multiple sclerosis (MS) at the level of $-\log_{10}(P) \geq 0$, $-\log_{10}(P) \geq 1$, $-\log_{10}(P) \geq 2$, $-\log_{10}(P) \geq 3$ corresponding to $P \leq 1$, $P \leq 0.1$, $P \leq 0.01$, $P \leq 0.001$, respectively. Dotted lines indicate the null hypothesis. (b) Conditional true discovery rate (TDR) plots illustrating the increase in TDR associated with increased pleiotropic enrichment in SCZ conditioned on MS (SCZ|MS). (c) Cumulative replication plot showing the average rate of replication ($P < 0.05$) within SCZ substudies for a given $P$-value threshold shows that pleiotropic enriched SNP categories replicate at a higher rate in independent SCZ samples, for SCZ conditioned on MS (SCZ|MS). The vertical intercept is the overall replication rate per category. (d) Z-score-z-score plot demonstrates that the empirical replication $z$-scores closely match the expected \(a posteriori\) effect sizes of SCZ and are strongly dependent upon pleiotropy with MS. Analysis is based on split half method of the 17 PGC SCZ substudies.

Figure 2. Genetic pleiotropy enrichment of BD conditional on MS. (a) Conditional Q-Q plot of nominal versus empirical $-\log_{10} P$-values (corrected for inflation) in bipolar disorder (BD) below the standard GWAS threshold of $P < 5 \times 10^{-8}$ as a function of significance of association with multiple sclerosis (MS) at the level of $-\log_{10}(P) \geq 0$, $-\log_{10}(P) \geq 1$, $-\log_{10}(P) \geq 2$, $-\log_{10}(P) \geq 3$ corresponding to $P \leq 1$, $P \leq 0.1$, $P \leq 0.01$, $P \leq 0.001$, respectively. Dotted lines indicate the null hypothesis. (b) Conditional true discovery rate (TDR) plots illustrating the increase in TDR associated with increased pleiotropic enrichment in BD conditioned on MS (BD|MS).
HLA allele analysis. The PGC1 genotype data from the 17 substudies were used for HLA imputation (a detailed description of the data sets, quality control procedures, imputation methods, and principal components estimation, are given in Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. First, genotypes of SNPs in the extended MHC (Major Histocompatibility Complex) (chr6: 25652429–33368333) of each individual in all the samples were extracted. Then, the program HIBAG was used to impute genotypes of classical HLA alleles for each individual.

Figure 3. ‘Conditional FDR Manhattan plot’. Conditional FDR Manhattan plot of conditional $-\log_{10}(FDR)$ values for schizophrenia (SCZ) alone (gray) and SCZ conditioned on multiple sclerosis (MS; SCZMS, red). SNPs with conditional $-\log_{10}(FDR) > 1.3$ (that is, FDR $< 0.05$) are shown with large points. A black line around the large points indicates the most significant SNP in each LD block. This SNP is annotated with the closest gene. Genes are listed in increasing order in terms of SNPs genomic position within each chromosome from left to right and novel ones are marked by stars (*). The figure shows the localization of the 21 independent loci on a total of 13 chromosomes.

<table>
<thead>
<tr>
<th>Locus #</th>
<th>SNP</th>
<th>Location</th>
<th>Gene</th>
<th>SCZ P</th>
<th>FDR SCZ</th>
<th>FDR SCZMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rs1625579</td>
<td>1p21.3</td>
<td>AK094607$^{ab}$ (MIR137HG)</td>
<td>5.52E-06</td>
<td>4.92E-02</td>
<td>3.69E-02</td>
</tr>
<tr>
<td>2</td>
<td>rs17180327</td>
<td>2q31.1</td>
<td>CWC22$^{ac}$</td>
<td>6.37E-06</td>
<td>5.19E-02</td>
<td>3.95E-03</td>
</tr>
<tr>
<td>3</td>
<td>rs7646226</td>
<td>3p21-p14</td>
<td>PTPRG$^{bc}$</td>
<td>5.51E-06</td>
<td>4.92E-02</td>
<td>2.43E-03</td>
</tr>
<tr>
<td>4</td>
<td>rs9462875</td>
<td>6p21.1</td>
<td>CUL9$^{bc}$</td>
<td>1.20E-05</td>
<td>6.59E-02</td>
<td>4.14E-02</td>
</tr>
<tr>
<td>5</td>
<td>rs10257990</td>
<td>7p22</td>
<td>MAD1L1$^{ab}$</td>
<td>5.53E-06</td>
<td>4.92E-02</td>
<td>1.63E-02</td>
</tr>
<tr>
<td>6</td>
<td>rs10503253</td>
<td>8p23.2</td>
<td>CSMD1$^{ab}$</td>
<td>3.96E-06</td>
<td>4.70E-02</td>
<td>4.04E-02</td>
</tr>
<tr>
<td>7</td>
<td>rs9990941</td>
<td>8q21.3</td>
<td>MMP16$^{ab}$</td>
<td>2.28E-06</td>
<td>4.32E-02</td>
<td>1.29E-02</td>
</tr>
<tr>
<td>8</td>
<td>rs396861</td>
<td>9p24</td>
<td>AK3</td>
<td>6.89E-06</td>
<td>5.19E-02</td>
<td>4.53E-02</td>
</tr>
<tr>
<td>9</td>
<td>rs432960</td>
<td>10q24.32</td>
<td>AS3MT$^b$</td>
<td>2.65E-06</td>
<td>4.32E-02</td>
<td>1.29E-02</td>
</tr>
<tr>
<td>10</td>
<td>rs12411886</td>
<td>10q24.32</td>
<td>CNNM2$^b$</td>
<td>1.79E-06</td>
<td>4.10E-02</td>
<td>1.86E-02</td>
</tr>
<tr>
<td>11</td>
<td>rs11191732</td>
<td>10q25.1</td>
<td>NEURL$^b$</td>
<td>2.55E-06</td>
<td>4.32E-02</td>
<td>2.69E-02</td>
</tr>
<tr>
<td>12</td>
<td>rs1025641</td>
<td>10q26.2</td>
<td>C1orf90</td>
<td>7.51E-06</td>
<td>5.54E-02</td>
<td>4.87E-02</td>
</tr>
<tr>
<td>13</td>
<td>rs2852034</td>
<td>11q22.1</td>
<td>CNTN5</td>
<td>1.12E-05</td>
<td>6.00E-02</td>
<td>2.90E-02</td>
</tr>
<tr>
<td>14</td>
<td>rs540723</td>
<td>11q23.3</td>
<td>STT3A$^b$</td>
<td>1.82E-06</td>
<td>4.10E-02</td>
<td>2.56E-02</td>
</tr>
<tr>
<td>15</td>
<td>rs7972947</td>
<td>12p13.3</td>
<td>CANC1$^{ab}$</td>
<td>7.12E-06</td>
<td>5.54E-02</td>
<td>4.87E-02</td>
</tr>
<tr>
<td>16</td>
<td>rs2007044</td>
<td>12p13.3</td>
<td>CANC1$^{ab}$</td>
<td>2.74E-05</td>
<td>9.43E-02</td>
<td>1.75E-02</td>
</tr>
<tr>
<td>17</td>
<td>rs12436216</td>
<td>14q13.2</td>
<td>KIAA0391$^b$</td>
<td>7.40E-06</td>
<td>5.54E-02</td>
<td>4.87E-02</td>
</tr>
<tr>
<td>18</td>
<td>rs1869901</td>
<td>15q15</td>
<td>PLCB2$^b$</td>
<td>3.66E-06</td>
<td>4.70E-02</td>
<td>4.04E-02</td>
</tr>
<tr>
<td>19</td>
<td>rs4887348</td>
<td>15q25</td>
<td>NTRK3</td>
<td>4.69E-05</td>
<td>1.39E-01</td>
<td>3.05E-02</td>
</tr>
<tr>
<td>20</td>
<td>rs4309482</td>
<td>18</td>
<td>AK093940</td>
<td>9.66E-06</td>
<td>6.00E-02</td>
<td>1.34E-02</td>
</tr>
</tbody>
</table>

Independent complex or single-gene loci with SNP(s) with a conditional FDR (SCZMS) $< 0.05$ in schizophrenia (SCZ) given association in multiple sclerosis (MS). All significant SNPs are listed and sorted in each LD block and independent loci are listed consecutively (Locus #). Chromosome location (Location), closest gene (Gene), P-value of SCZ (SCZ P-value) and false discovery rate of SCZ, FDR (SCZ) are also listed. All data were first corrected for genomic inflation. $^a$Loci identified by GWASs without leveraging genetic pleiotropy structure between phenotypes. $^b$Loci identified using conditional FDR method on SCZ with CVD. $^c$Loci identified using conditional FDR method on SCZ with BD.
SNPs located in the MHC region identified with a conditional FDR (SCZ|MS) < 0.05 in schizophrenia (SCZ) given in multiple sclerosis (MS). Chromosome location (Location), closest gene (Gene), P-value of SCZ (SCZ P-value) and false discovery rate of SCZ, FDR (SCZ) are also listed. All data were first corrected for genomic inflation. *p identified using conditional FDR method on SCZ with CVD. **p identified using conditional FDR method on SCZ with BD. *p identified by GWAS without leveraging genetic pleiotropy structure between phenotypes.

Table 2. Conditional FDR < 0.05, MHC SNPs associated with schizophrenia given multiple sclerosis

<table>
<thead>
<tr>
<th>SNP</th>
<th>Location</th>
<th>Gene</th>
<th>SCZ P</th>
<th>FDR SCZ</th>
<th>FDR SCZ</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs9379780</td>
<td>6p22.3</td>
<td>SCGN10b</td>
<td>3.25E-06</td>
<td>4.51E-02</td>
<td>1.59E-02</td>
<td></td>
</tr>
<tr>
<td>rs3585746</td>
<td>6p21.3</td>
<td>HIST1H1E2</td>
<td>3.87E-08</td>
<td>4.49E-03</td>
<td>1.47E-03</td>
<td></td>
</tr>
<tr>
<td>rs13218591</td>
<td>6p22.1</td>
<td>B1N3A2</td>
<td>4.24E-05</td>
<td>1.23E-01</td>
<td>4.85E-02</td>
<td></td>
</tr>
<tr>
<td>rs7746199</td>
<td>6p22.1</td>
<td>POM121L2</td>
<td>1.18E-08</td>
<td>2.69E-03</td>
<td>1.59E-03</td>
<td></td>
</tr>
<tr>
<td>rs853676</td>
<td>6p22.3-p22.1</td>
<td>ZNF232*</td>
<td>6.71E-08</td>
<td>2.69E-03</td>
<td>1.59E-03</td>
<td></td>
</tr>
<tr>
<td>rs2113203</td>
<td>6p22.1</td>
<td>ZSSCAN5*</td>
<td>3.64E-06</td>
<td>4.70E-02</td>
<td>1.15E-03</td>
<td></td>
</tr>
<tr>
<td>rs2844776</td>
<td>6p21.3</td>
<td>TRIM62cnc</td>
<td>2.34E-09</td>
<td>7.23E-04</td>
<td>8.15E-05</td>
<td></td>
</tr>
<tr>
<td>rs3094127</td>
<td>6p21.3</td>
<td>FLOT1*</td>
<td>6.66E-05</td>
<td>1.57E-01</td>
<td>3.68E-02</td>
<td></td>
</tr>
<tr>
<td>rs182532</td>
<td>6p21.3</td>
<td>VARS2*</td>
<td>8.61E-04</td>
<td>4.37E-01</td>
<td>4.69E-02</td>
<td></td>
</tr>
<tr>
<td>rs1650999</td>
<td>6p21.3</td>
<td>PSORS1C1*</td>
<td>3.04E-05</td>
<td>9.43E-02</td>
<td>3.38E-03</td>
<td></td>
</tr>
<tr>
<td>rs9264942</td>
<td>6p21.3</td>
<td>HLA-B8c</td>
<td>3.25E-04</td>
<td>3.26E-01</td>
<td>2.36E-02</td>
<td></td>
</tr>
<tr>
<td>rs2857595</td>
<td>6p21.3</td>
<td>NCR3</td>
<td>8.96E-05</td>
<td>1.98E-01</td>
<td>9.55E-03</td>
<td></td>
</tr>
<tr>
<td>rs805294</td>
<td>6p21.3</td>
<td>LYN66C</td>
<td>2.93E-05</td>
<td>1.08E-01</td>
<td>3.99E-03</td>
<td></td>
</tr>
<tr>
<td>rs3113442</td>
<td>6p21.3</td>
<td>NOTCH4</td>
<td>3.05E-06</td>
<td>4.51E-02</td>
<td>2.11E-03</td>
<td></td>
</tr>
<tr>
<td>rs2395173</td>
<td>6p21.3</td>
<td>HLA-DRA4b</td>
<td>8.07E-04</td>
<td>4.37E-01</td>
<td>4.69E-02</td>
<td></td>
</tr>
<tr>
<td>rs3129690</td>
<td>6p21.3</td>
<td>HLA-DRA4b</td>
<td>1.89E-06</td>
<td>4.10E-02</td>
<td>6.98E-04</td>
<td></td>
</tr>
<tr>
<td>rs7383287</td>
<td>6p21.3</td>
<td>HLA-DRA4b</td>
<td>3.44E-05</td>
<td>1.08E-01</td>
<td>3.99E-03</td>
<td></td>
</tr>
<tr>
<td>rs1480380</td>
<td>6p22.1</td>
<td>ZKSCAN3</td>
<td>3.05E-06</td>
<td>4.51E-02</td>
<td>2.11E-03</td>
<td></td>
</tr>
</tbody>
</table>

© 2015 Macmillan Publishers Limited

Molecular Psychiatry (2015), 207 – 214

intervals given in Supplementary Figure 1. The enrichment also remained significant after removing the SNPs with genome-wide significant P-values (censored QQ plots, Supplementary Figure 2). In contrast, we found no evidence for enrichment in BD conditional on MS (Figure 2).

Association with MS increases conditional true discovery rate (TDR) in SCZ
Variation in enrichment in pleiotropic SNPs is associated with corresponding variation in conditional TDR, equivalent to one minus the conditional FDR. A conservative estimate of the conditional TDR for each nominal P-value is equivalent to 1-\((P/q)\) as plotted on the conditional QQ plots (see Methods). This relationship is shown for SCZ conditioned on MS in a conditional TDR plot (Figure 1b; TDR\textsubscript{SCZ | MS} and for BD Figure 2b; TDR\textsubscript{BD|MS}). For a given conditional TDR, the corresponding estimated nominal P-value threshold varied by a factor of 100 from the most to the least enriched SNP category for SCZ conditioned by MS. Since the conditional TDR is strongly related to predicted replication rate, the replication rate is expected to increase for SNPs in categories with higher conditional TDR.

Replication rate in SCZ is increased by MS association
To address the possibility that the observed pattern of differential enrichment results from spurious (that is, nongeneralizable) associations due to category-specific stratification or statistical modeling errors, we examined the empirical replication rate across independent substudies for SCZ. Figure 1c shows the empirical cumulative replication rate plots as a function of nominal P-value, for the same categories as for the conditional Q and TDR plots in Figures 1a and b. Consistent with the conditional TDR pattern, we found that the nominal P-value corresponding to a wide range of replication rates was 100 times higher for \(-\log_{10}(P_{\text{MS}}) \geq 3\) relative to the \(-\log_{10}(P_{\text{MS}}) \geq 0\) category (Figure 1c). Similarly, SNPs from pleiotropic SNP categories showing the greatest enrichments (\(-\log_{10}(P_{\text{MS}}) \geq 3\)) replicated at highest rates, up to five times higher than all SNPs (\(-\log_{10}(P_{\text{MS}}) > 0\)), for a wide range of P-value thresholds. This strongly suggests that replication of SNP associations varies as a function of estimated conditional TDR.

Replication effect size depends upon MS association
Consistent with the pattern observed for replication rates in SCZ substudies (see above), we found that the effect sizes of SNPs in enriched categories (for example, \(-\log_{10}(P_{\text{MS}}) \geq 3\)) replicated better than effect sizes of SNPs in less-enriched categories (for example, \(-\log_{10}(P_{\text{MS}}) > 0\); Figure 1d). This indicates that the fidelity of replication effect sizes is closely related to the conditional TDR.
BD
14 (Tables 1 and 2). All SNPs with conditional FDR < 0.05 are listed in Supplementary Table 1.

Effect of the size of strata on enrichment

The observed enrichment was further confirmed by performing the same analysis on additional categories (−log10 (pMS) ≥ 4, −log10 (pMS) ≥ 5 and −log10 (pMS) ≥ 6, Supplementary Figure 3). Although the general enrichment pattern persisted, the number of valid SNPs, which exist in both SCZ and MS data set and also have valid P-values, in these extra categories was smaller. In total, 425 028 SNPs having valid P-values for both SCZ and MS were analyzed in this study. They contribute 425 028, 47410, 7077, 1781, 808, 525 and 391 to the six categories conditioned by the significance level of MS, respectively.

Distribution of allele frequencies in strata

We also investigated the distribution of the minor allele frequencies (MAF) of the corresponding SNPs of each stratum from the 1KGP. Supplementary Figure 4 shows the average MAF * (1 − MAF), namely, the genetic variance, in strata after pruning SNPs in LD (r² > 0.2). As the significance level of SNPs with MS increases, there is a noticeable increase in average genetic variance, which is expected as MAF confounds multiplicatively with the true effect size of the variants. However, the effect of MAF alone cannot explain the observed enrichment (see Supplementary Figure 4).

HLA imputation and association analysis

Among the loci identified by conditional FDR methods, eight are located in the MHC (Table 2). It is possible that these signals may be driven by common HLA alleles affecting both SCZ and MS. To test this hypothesis, we imputed HLA class I and class II alleles using the PGC1 genotype data (see Materials and Methods). We performed association analysis between imputed HLA alleles and SCZ. We replicated earlier findings
10 that the alleles HLA-B*08:01, HLA-C*07:01, HLA-DRB1*03:01, HLA-DQA1*05:01 and HLA-DQB1*02:01 are negatively associated with SCZ (P < 7.8 × 10⁻⁴). Among these, HLA-DRB1*03:01 and HLA-DQB1*02:01 have been reported to be positively associated with MS. However, no association was seen with SCZ for the strong MS predisposing HLA-DRB1*15:01 and HLA-DRB1*13:03 alleles, nor for the protective HLA-A*02:01 allele. We further tested whether SNPs in the MHC with conditional FDR < 0.05 were independent of the association signal with the classical HLA alleles (see Materials and Methods). SNPs rs9379780, rs3857546, rs7746199, rs8536767 and rs2844776 seem to be independent of the HLA allelic signal (Supplementary Figure 5 and Supplementary Table 2). We further tested if the associated HLA alleles were independent of each other by conditional analysis between them (see Materials and Methods). The results in Supplementary Table 3 suggest that the observed associations are driven by a single haplotype block (that is, ancestral haplotype 81), consisting of the five individual HLA alleles. However, because of the low number of successfully imputed HLA alleles, it is difficult to exclude independent signals.

The effect of MHC SNPs on enrichment

We also investigated the effect of MHC-related SNPs (SNPs located within the MHC or SNPs within 1 Mb and in LD (r² > 0.2) with such SNPs) on the observed enrichment for SCZ and BD conditional on MS (see Supplementary Information and Supplementary Figure 6). After removing the MHC-related SNPs, the enrichment of SCZ conditioned on MS was substantially attenuated (Supplementary Figure 6). In contrast, removing the MHC-related SNPs did not affect the enrichment of BD conditioned on MS (Supplementary Figure 6). Further, we also investigated the effect of removing the MHC-related SNPs on the previously reported enrichment of SCZ conditioned on BD. As illustrated in Supplementary Figure 8, the enrichment between BD and SCZ was not affected by removing the MHC-related SNPs.

Enrichment analysis of other psychiatry disorders

Using the analysis approach described above, we evaluated genetic enrichment in major depressive disorder,
26 autism spectrum disorder
27 and attention deficit/hyperactivity disorder
28 GWAS summary statistics from the PGC conditioned on MS. In contrast to SCZ, none of these phenotypes demonstrated significant enrichment (Supplementary Figure 7).

DISCUSSION

Our results demonstrate genetic overlap between SCZ and MS and identify 21 independent loci associated with SCZ, which are driven by the overlap in the MHC region. In contrast, despite known genetic overlap between SCZ and BD, we found no overlap between BD and MS. As most of the overlap between MS and SCZ was driven by the MHC, and given the previous evidence that the genetic overlap between BD and SCZ is located outside the MHC, our findings suggest that the MHC cluster could differentiate between BD and SCZ susceptibility.

These results implicate shared molecular pathways between SCZ and MS. Many of the common association signals between SCZ and MS are located within the MHC, strongly suggesting the MHC, and possibly HLA alleles, in SCZ development. This is in line with previous findings of common variants in the MHC associated with SCZ,
7,8,10,35 which seems to implicate immune factors. The current SNP and HLA allelic analyses indicate that there could be several risk loci in the MHC for SCZ, possibly including both classical antigen-presenting HLA alleles and other immune loci. The DRB1*03:01 and DQB1*02:01 alleles that increase risk of MS were found to decrease the risk for SCZ. Other MS-associated HLA alleles were not found to be associated with SCZ in our analysis. Thus, although MS and SCZ are both associated with several loci in the MHC, their genetic risk profile appears to be divergent. Because of the low imputation rate in our study, it is difficult to draw firm conclusions about precisely which loci and allelic variants in the MHC are involved in SCZ. Several of the non-MHC loci suggested to be involved in SCZ susceptibility by the current study are also related to immune pathways, which are established as key mechanisms in MS pathogenesis.
16 Further, support for immune involvement in SCZ comes from recent findings of abnormalities in cytokine levels and other immune markers in SCZ,
6,34 and from strong evidence suggesting a role for prenatal infections in SCZ development.
35 However, given the presence of neurobiologic loci, like NOTCH4 and TRIM26, within the MHC, and the role of MHC class I in neuronal plasticity,
17 we cannot rule out the possibility that the genetic associations found in the MHC may reflect shared pathobiology between SCZ and MS,
11 in addition to immunological factors. Since myelin is affected in MS, the current findings are in line with previous work suggesting a role of abnormal myelination also in SCZ.
36 Considered together, these findings strongly indicate the need for additional research to determine the MHC-related mechanisms underlying SCZ.

The three-way relationship between SCZ-MS-BD observed here and in our previous study
14 demonstrates that overlap between SCZ and MS may be due to shared genetic risk loci mainly in the MHC, while overlap between BD and SCZ is present across almost the entire genome with the exception of the MHC. Indeed, removing the MHC-related SNPs did not affect the enrichment between BD and SCZ. Together with our current findings of no overlap between BD and MS, these results indicate that the MHC involvement could differentiate SCZ from BD susceptibility. The lack of pleiotropy between MS and BD indicates that the MHC is
not associated with psychosis in general, but is more specifically related to SCZ. This is further supported by the lack of MS-conditional enrichment observed with autism spectrum disorder, major depressive disorder and ADHD (Supplementary Figure 7). Several lines of evidence support the notion of shared genes and phenotypes between SCZ and BD. Given that the SCZ and BD data sets contained partly overlapping controls, the lack of findings in BD argues against unknown population stratification factors driving results for SCZ. Some yet unidentified population stratification related to immune genes within European populations could still have a role, but this would have also affected the results of the original GWAS.

The current findings of overlapping genetic variants in SCZ and MS show the feasibility of using a genetic epidemiology framework that leverages overlap in genetic signal from independent GWAS of brain phenotypes to improve statistical power for gene discovery. In the original SCZ GWAS sample, three loci were significant before replication in additional samples. By combining the original SCZ sample with the independent GWAS of MS, we identified significant pleiotropic signals in a total of 21 loci, only eleven of which had been reported in previous SCZ GWAS (Tables 1 and 2). This illustrates the increased power of our combined analytical approach. Note that the conditional FDR method protects against the possibility that the observed pleiotropic effects are driven by a strong signal in only one phenotype. It is likely that the new loci uncovered here would have been discovered if the number of subjects in the SCZ GWAS had been adequately large. However, we here show how summary statistics from independent samples provide a large advantage for gene discovery without the extra cost and resources needed to obtain new case–control samples. This has significant implications for GWAS, which have been criticized for not identifying enough of the genetic component of known heritable traits, and are increasingly seen as just a first step in more sophisticated analyses.

Our findings also suggest that the etiology of complex brain disorders may involve a high number of pleiotropic variants, each with a small effect. SCZ appears to show an especially high degree of pleiotropy, with significant overlap with a number of different phenotypes. It is important to note, however, that this pleiotropy is not general, but is phenotype specific. The genetic pleiotropy observed between SCZ and MS is weaker than that between SCZ and BD. SCZ also shows varying degrees of pleiotropy with different CVF risk factors, with some showing strong effects (for example, triglycerides), others showing weaker effects (for example, blood pressure, waist-hip ratio) and still others showing no significant enrichment (for example, type 2 diabetes). With regard to SCZ pleiotropy with CVD risk, we showed that several of the SCZ loci detected here based on pleiotropy with MS were previously identified, leveraging pleiotropy between SCZ and CVD. This suggests the likely presence of genetic pleiotropy between MS and CVD risk factors that is worth further exploration.

To conclude, the conditional FDR approach identified 21 loci associated with SCZ conditioned on MS, and this enrichment was driven by the overlap in the MHC region. Imputation of the HLA alleles showed that the HLA alleles found to increase MS risk were associated with decreased SCZ risk. We found no overlap in the MHC between BD and MS, suggesting that MHC could be a key difference between SCZ and BD genetics. Our findings also indicate that leveraging existing GWASs for pleiotropic genes may represent an important avenue by which currently available resources can be optimized to identify more of the missing heritability of complex phenotypes.

CONFLICT OF INTEREST
The authors declare no conflict of interest.


