

SUPPLEMENTARY INFORMATION

Participant Samples

The schizophrenia (SCZ) GWAS summary statistics results were obtained from the Psychiatric Genomics Consortium (PGC) Schizophrenia Work Group¹, which consisted of 9,394 cases with schizophrenia or schizoaffective disorder and 12 462 controls (52% screened) from a total of 17 samples from 11 countries. Semi-structured interviews were used by trained interviewers to collect clinical information, and operational criteria were used to establish the diagnosis. The quality of the phenotypic data was verified by a systematic review of data collection methods and procedures at each site, and only studies that fulfilled these criteria were included. Controls were selected from the same geographical and ethnic populations as cases. For further details on sample characteristics and quality control procedures applied, please see Ripke et al¹.

The bipolar disorder (BD) GWAS summary statistics results were obtained from the PGC Bipolar Disorder Working Group², which consisted of n=16 731 participants, including 7481 cases and 9 250 controls, from 11 studies from 7 countries. Standardized semi-structured interviews were used by trained interviewers to collect clinical information about lifetime history of psychiatric illness and operational criteria applied to make lifetime diagnosis according to recognized classifications. All cases have experienced pathologically relevant episodes of elevated mood (mania or hypomania) and meet operational criteria for a BD diagnosis. The sample consisted of BD I (84%), BD II (11%), schizoaffective disorder bipolar type (4%), and BD NOS (1%). Controls were selected from the same geographical and ethnic populations as cases. For further details on sample characteristics and quality control procedures applied, please see Sklar et al².

The multiple sclerosis (MS) GWAS summary statistics results were obtained from the International Multiple Sclerosis Genetics Consortium (IMSGC)³, n=27 148, consisting of 10 299 cases and 16 849 controls from 15 countries. The diagnosis was obtained using established and well-validated criteria that combine clinical and para-clinical laboratory-based information⁴. For

further details on sample characteristics and quality control procedures applied, please see Sawcer et al³.

There were 2 974 controls in the SCZ UK case control sample⁵ and the BD UK case control sample² from the Wellcome Trust Case Control Consortium that were also included in the MS GWAS. These constitute 24% of the total number of controls (n=12 462) in the SCZ PGC sample¹ and 32% of the total number of controls (n=9 250) in the BD PGC sample². Approximately 50% of the controls in the BD GWAS were also included in the SCZ GWAS. The relevant institutional review boards or ethics committees approved the research protocol of the individual GWASs used in the current analysis and all participants gave written informed consent.

Conditional Q-Q plots

Q-Q plots compare a nominal probability distribution against an empirical distribution. In the presence of all null relationships, nominal p-values form a straight line on a Q-Q plot when plotted against the empirical distribution. For each phenotype, for all SNPs and for each categorical subset (strata), $-\log_{10}$ nominal p-values were plotted against $-\log_{10}$ empirical p-values (conditional Q-Q plots). Leftward deflections of the observed distribution from the projected null line reflect increased tail probabilities in the distribution of test statistics (z-scores) and consequently an over-abundance of low p-values compared to that expected by chance, also named ‘enrichment’.

Conditional True Discovery Rate (TDR)

The ‘enrichment’ seen in the conditional Q-Q plots can be directly interpreted in terms of true discovery rate ($TDR = 1 - FDR$)⁶. More specifically, for a given p-value cutoff, the FDR is defined as

$$FDR(p) = \pi_0 F_0(p) / F(p), \quad [1]$$

where π_0 is the proportion of null SNPs, F_0 is the null cumulative distribution function (cdf), and F is the cdf of all SNPs, both null and non-null⁷. Under the null hypothesis, F_0 is the cdf of the uniform distribution on the unit interval $[0,1]$, so that Eq. [1] reduces to

$$\text{FDR}(p) = \pi_0 p / F(p), \quad [2]$$

The cdf F can be estimated by the empirical cdf $q = N_p / N$, where N_p is the number of SNPs with p -values less than or equal to p , and N is the total number of SNPs. Replacing F by q in Eq. [2], we get

$$\text{Estimated FDR}(p) = \pi_0 p / q, \quad [3]$$

which is biased upwards as an estimate of the FDR³². Replacing π_0 in Equation [3] with unity gives an estimated FDR that is further biased upward;

$$q^* = p/q \quad [4]$$

If π_0 is close to one, as is likely true for most GWASs, the increase in bias from Eq. [3] is minimal. The quantity $1 - p/q$, is therefore biased downward, and hence a conservative estimate of the TDR. Referring to the Q-Q plots, we see that q^* is equivalent to the nominal p -value divided by the empirical quantile, as defined earlier. We can thus read the FDR estimate directly off the Q-Q plot as

$$-\log_{10}(q^*) = \log_{10}(q) - \log_{10}(p), \quad [5]$$

i.e. the horizontal shift of the curves in the Q-Q plots from the expected line $x = y$, with a larger shift corresponding to a smaller FDR. This is illustrated in Fig. 1a. For each range of p -values in the pleiotropic trait (indicated by differently colored curves), we calculated the TDR as a function of the p -value in SCZ and reported it in Figure 1b (Fig. 2 for BD).

Further analyses performed

Significance of conditional enrichment

After pruning the SNPs by removing SNPs in linkage disequilibrium ($r^2 \geq 0.2$), we computed 95% confidence intervals for the conditional Q-Q plots. From these confidence intervals we calculated

standard errors and used two sample t-tests to estimate the difference (degree of departure) of the empirical distribution of SNPs in SCZ or BD (phenotype 1) that are above a given association threshold ($-\log_{10}(p) \geq 1$, $-\log_{10}(p) \geq 2$, $-\log_{10}(p) \geq 3$, $-\log_{10}(p) \geq 4$; red lines) in MS (phenotype 2) compared to the $-\log_{10}(p) \geq 0$ in phenotype 1 category (blue line). The same procedure was used for the “censored data” of MS conditional on SCZ. Supplementary Figure 1 and 2 indicate the most significant difference, as assessed using a two samples t-test, between the red ($-\log_{10}(p) > 1, 2, 3$ or 4) and blue ($-\log_{10}(p) > 0$) lines along with p-values. This is reflected in the biggest difference between the 95% confidence intervals.

Conditional analysis of HLA alleles

We tested if the associated HLA signals were independent of each other by conditional analysis between them. Samples with imputed HLA allele genotypes were combined before the analysis. The logistic regression method implemented in PLINK⁸ was employed to test each significant HLA allele for associations with SCZ, including another significant HLA allele, the first 5 principal components and sample indicator variable as covariates. Supplementary Table 3 shows that it is more probable that the observed associations were driven by a single haplotype-block, consisting of the 5 individual HLA alleles.

The effect of HLA region on enrichment

Due to the complexity of HLA region and low call rate of the conditional analysis performed only on imputed HLA alleles, we cannot exclude the possibility that the other HLA region genes may also be associated with SCZ, or, the whole region may be the driver for the observed enrichment. We reapplied the enrichment method to same dataset with SNPs either located within the HLA region or in LD ($r^2 > 0.2$) with such SNPs (in total 9379 SNPs). These results indicate that the enrichment of SCZ conditional on MS is largely the consequence of the HLA region

(Supplementary Fig. 6a) whereas, the enrichment pattern of BD is unaffected by the absence of the HLA region. This further confirms the important role of HLA region in SCZ pathology.

To further evaluate the role of the HLA region in SCZ and BD, we removed only SNPs located within the 5 HLA genes, which were shown to associate with SCZ by above conditional analysis, and other SNPs that in LD ($r^2 > 0.2$) with such SNPs (in total 3480 SNPs). In this setting, genetic enrichment in both SCZ and BD was unaffected (Supplementary Fig. 6b) (we should include BD as well here just for completeness and also to show contrast with SCZ). This corroborates the result of the conditional analysis of HLA allele that the SNPs revealed by our pleiotropic enrichment methods are independent of the known alleles comprising the HLA region.

List of the principle investigators in the IMSGC:

International Multiple Sclerosis Genetic Consortium (IMSGC) membership

Lisa Barcellos¹, David Booth², Jacob L McCauley³, Manuel Comabella⁴, Alastair Compston⁵, Sandra DAlfonso⁶, Philip De Jager⁷, Bertrand Fontaine⁸, An Goris⁹, David Hafler¹⁰, Jonathan Haines¹¹, Hanne F. Harbo¹², Stephen L Hauser¹³, Clive Hawkins¹⁴, Bernhard Hemmer¹⁵, Jan Hillert¹⁶, Adrian Ivinson¹⁷, Ingrid Kockum¹⁸, Roland Martin¹⁹, Filippo Martinelli Boneschi²⁰, Jorge Oksenberg¹³, Tomas Olsson¹⁸, Annette Oturai²¹, Nikolaos Patsopoulos⁷, Margaret Pericak-Vance²², Janna Saarela²³, Stephen Sawcer⁵, Anne Spurkland²⁴, Graeme Stewart², Frauke Zipp²⁵

¹Genetic Epidemiology and Genomics Laboratory, Division of Epidemiology, School of Public Health, University of California, Berkeley, California, USA.

² Westmead Millennium Institute, University of Sydney, New South Wales 2145, Australia.

³ John P. Hussman Institute for Human Genomics and The Dr. John T Macdonald Foundation Department of Human Genetics, University of Miami, Miller School of Medicine, Miami, Florida, USA

⁴ Clinical Neuroimmunology Unit, Multiple Sclerosis Center of Catalonia (CEM-Cat), Vall d'Hebron University Hospital, Barcelona, Spain.

⁵ Department of Clinical Neurosciences, University of Cambridge, Addenbrooke's Hospital, Cambridge, CB2 0QQ, UK

⁶ Department of Medical Sciences and Interdisciplinary Research Center of Autoimmune Diseases (IRCAD), University of Eastern Piedmont, Novara, Italy

⁷ Department of Neurology, Brigham & Women's Hospital and Harvard Medical School, Boston, MS, USA

⁸ Laboratory for Neuroimmunology, Section for Experimental Neurology, Katholieke Universiteit Leuven, Leuven, Belgium

⁹ INSERM UMR S 975 CRICM, UPMC, Departement de neurologie Pitie-Salpetriere, AP-HP, Paris, France

¹⁰ Departments of Neurology and Immunobiology, Yale School of Medicine, New Haven, CT, USA.

¹¹ Center for Human Genetics Research, Vanderbilt University Medical Center, 519 Light Hall, Nashville, Tennessee 37232, USA

¹² Department of Neurology, Oslo University Hospital, Ullevål and University of Oslo, Oslo Norway

¹³ Department of Neurology, University of California San Francisco, San Francisco, USA

¹⁴ Keele University Medical School, Stoke-on-Trent ST4 7NY, UK

¹⁵ Klinik für Neurologie, Klinikum rechts der Isar, Technische Universität, München, Germany

¹⁶ Department of Clinical Neurosciences, Centre for Molecular Medicine. Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden.

¹⁷ Harvard NeuroDiscovery Center, Harvard Medical School, Boston, Massachusetts, USA.

¹⁸ Neuroimmunology Unit, Center for Molecular Medicine, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

¹⁹ Institute for Neuroimmunology and Clinical MS Research (inims), Centre for Molecular Neurobiology, Falkenried 94, D-20251 Hamburg, Germany.

²⁰ Department of Neurology, Institute of Experimental Neurology (INSPE), Division of Neuroscience, San Raffaele Scientific Institute, Milan, Italy

²¹ Danish Multiple Sclerosis Center, University of Copenhagen and Department of Neurology, Rigshospitalet, Copenhagen, Denmark

²² John P. Hussman Institute for Human Genomics and The Dr. John T Macdonald Foundation
Department of Human Genetics, University of Miami, Miller School of Medicine, Miami, Florida ,
USA

²³ Institute for Molecular Medicine Finland FIMM, University of Helsinki, Helsinki, Finland

²⁴ Institute of Basal Medical Sciences, University of Oslo, Norway

²⁵ Department of Neurology, University Medicine Mainz, Johannes Gutenberg University Mainz,
Mainz, Germany.

References for Supplementary Information

1. Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* 2011; **43**: 969–976.
2. 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.
3. International Multiple Sclerosis Genetics Consortium, Wellcome Trust Case Control Consortium 2, Sawcer S, Hellenthal G, Pirinen M, Spencer CCA *et al.* Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 2011; **476**: 214–219.
4. McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD *et al.* Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Annals of Neurology*. 2001; **50**: 121–127.
5. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460**: 748–752.
6. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Series Stat Methodol* 1995; **57**: 289–300.
7. Efron B. Size, power and false discovery rates. *The Annals of Statistics* 2007; **35**: 1351–1377.
8. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* 2007; **81**: 559–575.