

Genetic variation in the Major Histocompatibility Complex and association with depression

Kylie P Glanville¹, Jonathan R I Coleman^{1,2}, Ken B Hanscombe³, Jack Euesden¹, Shing Wan Choi¹, Kirstin L Purves¹, Gerome Breen^{1,2}, Tracy M Air⁴, Till F M Andlauer^{5,6}, Bernhard T Baune⁷, Elisabeth B Binder^{5,8}, Douglas H R Blackwood⁹, Dorret I Boomsma¹⁰, Henriette N Buttenschøn^{11,12}, Lucía Colodro-Conde¹³, Udo Dannlowski¹⁴, Nese Direk^{15,16}, Erin C Dunn^{17,18,19}, Andreas J Forstner^{20,21,22,23}, Eco JC de Geus^{10,24}, Hans J Grabe²⁵, Steven P Hamilton²⁶, Ian Jones²⁷, Lisa A Jones²⁸, James A Knowles²⁹, Zoltán Kutalik^{30,31}, Douglas F Levinson³², Glyn Lewis³³, Penelope A Lind¹³, Susanne Lucae³⁴, Patrik K Magnusson³⁵, Peter McGuffin¹, Andrew M McIntosh^{9,36}, Yuri Milaneschi³⁷, Ole Mors^{12,38}, Sara Mostafavi^{39,40}, Bertram Müller-Myhsok^{5,6,41}, Nancy L Pedersen³⁵, Brenda WJH Penninx³⁷, James B Potash⁴², Martin Preisig⁴³, Stephan Ripke^{44,45,46}, Jianxin Shi⁴⁷, Stanley I Shyn⁴⁸, Jordan W Smoller^{17,18,19}, Fabian Streit⁴⁹, Patrick F Sullivan^{35,50,51}, Henning Tiemeier^{15,52,53}, Rudolf Uher⁵⁴, Sandra Van der Auwera²⁵, Myrna M Weissman^{55,56}, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium⁵⁷, Paul F O'Reilly¹, Cathryn M Lewis^{1,2}

1. Social Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, GB

2. NIHR Biomedical Research Centre, South London and Maudsley NHS Trust, King's College London, London, GB

3. Department of Medical & Molecular Genetics, King's College London, London, GB

4. Discipline of Psychiatry, Adelaide Medical School, University of Adelaide, Adelaide, SA, AU

5. Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, DE

6. Munich Cluster for Systems Neurology (SyNergy), Munich, DE
7. Department of Psychiatry, Melbourne Medical School, University of Melbourne, Melbourne, AU
8. Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, US
9. Division of Psychiatry, University of Edinburgh, Edinburgh, GB
10. Dept of Biological Psychology & EMGO+ Institute for Health and Care Research, Vrije Universiteit Amsterdam, Amsterdam, NL
11. NIDO | Danmark, Regional Hospital West Jutland, DK
12. iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, DK
13. Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, AU
14. Department of Psychiatry, University of Münster, Münster, Nordrhein-Westfalen, DE
15. Epidemiology, Erasmus MC, Rotterdam, Zuid-Holland, NL
16. Psychiatry, Dokuz Eylul University School Of Medicine, Izmir, TR
17. Department of Psychiatry, Massachusetts General Hospital, Boston, MA, US
18. Psychiatric and Neurodevelopmental Genetics Unit (PNGU), Massachusetts General Hospital, Boston, MA, US
19. Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, US
20. Department of Psychiatry (UPK), University of Basel, Basel, CH
21. Human Genomics Research Group, Department of Biomedicine, University of Basel, Basel, CH
22. Institute of Human Genetics, University of Bonn, Bonn, DE
23. Life&Brain Center, Department of Genomics, University of Bonn, Bonn, DE
24. Amsterdam Public Health Institute, Vrije Universiteit Medical Center, Amsterdam, NL
25. Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Mecklenburg-Vorpommern, DE
26. Psychiatry, Kaiser Permanente Northern California, San Francisco, CA, US

27. MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, GB
28. Department of Psychological Medicine, University of Worcester, Worcester, GB
29. Psychiatry & The Behavioral Sciences, University of Southern California, Los Angeles, CA, US
30. Institute of Social and Preventive Medicine (IUMSP), University Hospital of Lausanne, Lausanne, VD, CH
31. Swiss Institute of Bioinformatics, Lausanne, VD, CH
32. Psychiatry & Behavioral Sciences, Stanford University, Stanford, CA, US
33. Division of Psychiatry, University College London, London, GB
34. Max Planck Institute of Psychiatry, Munich, DE
35. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, SE
36. Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, GB
37. Department of Psychiatry, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, NL
38. Psychosis Research Unit, Aarhus University Hospital, Risskov, Aarhus, DK
39. Medical Genetics, University of British Columbia, Vancouver, BC, CA
40. Statistics, University of British Columbia, Vancouver, BC, CA
41. University of Liverpool, Liverpool, GB
42. Psychiatry, University of Iowa, Iowa City, IA, US
43. Department of Psychiatry, University Hospital of Lausanne, Prilly, Vaud, CH
44. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, US
45. Department of Psychiatry and Psychotherapy, Universitätsmedizin Berlin Campus Charité Mitte, Berlin, DE
46. Medical and Population Genetics, Broad Institute, Cambridge, MA, US
47. Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, US

48. Behavioral Health Services, Kaiser Permanente Washington, Seattle, WA, US
49. Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Baden-Württemberg, DE
50. Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, US
51. Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, US
52. Child and Adolescent Psychiatry, Erasmus MC, Rotterdam, Zuid-Holland, NL
53. Psychiatry, Erasmus MC, Rotterdam, Zuid-Holland, NL
54. Psychiatry, Dalhousie University, Halifax, NS, CA
55. Division of Epidemiology, New York State Psychiatric Institute, New York, NY, US
56. Psychiatry, Columbia University College of Physicians and Surgeons, New York, NY, US
57. Consortium members listed in supplementary material

Address correspondence to Kylie P. Glanville, kylie.glanville@kcl.ac.uk; Social, Genetic and Developmental Psychiatry Centre; Institute of Psychiatry, Psychology & Neuroscience; King's College London; London SE5 8AF; UK.

Abstract

Background: The prevalence of depression is higher in individuals suffering from autoimmune diseases, but the mechanisms underlying the observed comorbidities are unknown. Epidemiological findings point to a bi-directional relationship - that depression increases the risk of developing an autoimmune disease, and vice-versa. Shared genetic etiology is a plausible explanation for the overlap between depression and autoimmune diseases. In this study we tested whether genetic variation in the Major Histocompatibility Complex (MHC), which is associated with risk for autoimmune diseases, is also associated with risk for depression.

Method: We fine-mapped the classical MHC (chr6: 29.6-33.1 Mb), imputing 216 Human Leukocyte Antigen (HLA) alleles and four Complement Component 4 (C4) haplotypes in studies from the Psychiatric Genomics Consortium (PGC) Major Depressive Disorder (MDD) working group and the UK Biobank (UKB). In the 26 PGC-MDD studies, cases met a lifetime diagnosis of MDD, determined by a structured diagnostic interview. In the UKB, cases and controls were identified from an online mental health questionnaire. The total sample size was 45,149 depression cases and 86,698 controls. We tested for association between depression status and imputed MHC variants in each study and performed an inverse-variance weighted meta-analysis across the PGC-MDD and UKB samples, applying both a conservative region-wide significance threshold (3.9×10^{-6}) and a candidate threshold (1.6×10^{-4}).

Results: No HLA alleles or C4 haplotypes were associated with depression at the conservative threshold in the PGC, UKB or meta-analysis. HLA-B*08:01 was

associated with modest protection for depression at the candidate threshold in the meta-analysis. Under the conservative threshold, 70 SNPs were detected in the UKB and 143 SNPs were detected in the meta-analysis, mirroring previous findings from highly powered GWAS of depression.

Discussion: We found no evidence that HLA alleles, which play a major role in the genetic susceptibility to autoimmune diseases, or C4 haplotypes, which are strongly associated with schizophrenia, confer risk for depression. These results indicate that autoimmune diseases and depression do not share common risk loci of moderate or large effect in the MHC.

Introduction

Depression is a debilitating psychiatric disorder with an estimated lifetime prevalence of 15%¹, making it the leading cause of global disability². The disorder is characterised by heterogeneous symptoms profiles³ and variable treatment outcomes⁴. Developing effective pharmaceutical treatments relies on uncovering the etiology of a disorder⁵, and the field of psychiatric genetics has made great progress toward this objective in the last decade by devoting substantial resources to the study of depression^{6,7}. Despite this progress, the underlying biology of depression is still not fully understood. Comorbid psychiatric and physical traits may indicate shared biological pathways and provide a path to uncovering the etiology of idiopathic psychiatric disorders⁸. Here, we focus on comorbid autoimmune diseases and depression, consider the mechanisms that could drive the overlap, and test for evidence of shared genetic influences in the Major Histocompatibility Complex (MHC).

Epidemiological studies indicate that individuals with a history of autoimmune disease are at greater risk of developing mood disorders compared to individuals without a history of autoimmune disease⁹⁻¹². For example, a Danish Registry study⁹ identified individuals who had hospital contact for autoimmune diseases or mood disorders over a thirty year period and showed that the risk of developing a mood disorder increased following onset of any autoimmune disease (Incident Rate Ratio (IRR) = 1.45; 95% CI = 1.39-1.52).

One interpretation is that the distress arising from autoimmune afflictions is causal to the onset of a mood disorder. However, other evidence indicates that the relationship

is bi-directional^{13,14}. For example, another Danish Registry study¹³ showed that individuals with a history of depression were at increased risk for developing any autoimmune disease, compared to those without a history of depression (IRR = 1.25, 95% CI = 1.19-1.31), and that this increase remained relatively stable across the first decade after diagnosis of depression.

There are a number of plausible explanations for the observed overlap between depression and autoimmunity. Shared environmental influences may increase risk for both disorders, e.g. stress is a risk factor for autoimmune disease¹⁵, and there is a phenotypic and genetic correlation between anxiety and depression¹⁶. Another view is that shared genetic influences act on autoimmune disease and depression through common immune pathways. Inflammation is a hallmark characteristic of autoimmune disease¹⁷ and elevated levels of pro-inflammatory cytokines have been observed in some individuals with depression¹⁸. Despite increasing interest in the role of inflammatory pathways in the pathogenesis of depression¹⁹, the role of genetic variation in the MHC, which plays a crucial role in human immunity²⁰, has not been thoroughly interrogated in the context of depression.

The MHC is divided into three functionally distinct regions: class I and II regions contain highly polymorphic human leukocyte antigen (HLA) genes that are strongly associated with risk for autoimmune disease^{17,21,22}, and the class III region contains complement component 4 genes (C4), which are strongly associated with risk for schizophrenia²³. Three recent genome-wide association studies (GWAS) indicate that genetic variation within the MHC is involved in risk for depression. The largest GWAS of depression from the Psychiatric Genomics Consortium²⁴ identified 44

independent loci associated with depression including a SNP in the classical class I region. A GWAS of a broadly defined depression phenotype in the UK Biobank²⁵, revealed a peak of association in the MHC, with the strongest evidence for association also coming from the classical class I region. A recent meta-analysis²⁶ combined summary statistics from the latter two studies^{24,25} with estimates from a GWAS of depression including 23andMe data²⁷, and identified a SNP within the extended class I region with strong evidence for association with depression.

Highly polymorphic loci and long-range linkage-disequilibrium in the MHC complicate the interpretation of SNP associations²⁴. However, with the advent of methods for imputing HLA alleles²⁸ and C4 haplotypes²³, there is an opportunity to dissect SNP signal in the region with fine-mapping techniques. We used this approach to test whether genetic variation associated with risk for autoimmune disease and schizophrenia is also associated with risk for depression. To our knowledge, this is the first study to leverage imputation to interrogate the involvement of HLA alleles and C4 haplotypes in depression. Our efforts should lead to a better understanding of the role of these loci in depression, and may provide insights into the mechanisms driving comorbid autoimmunity and depression.

Methods

Participants

Participant data came from a subset of the Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium (PGC-MDD)²⁴ and from the UK Biobank (UKB)²⁹ to give a total of 131,847 individuals of European ancestry (55% female, 45,149 depression cases and 86,698 controls). We selected a subset of participants from the PGC-MDD study because individual-level genotype data, required for imputation of HLA alleles and C4 haplotypes, were not available for the entire PGC-MDD cohort. Individual-level genotype and phenotype data were available for 26 studies shared with the PGC, totaling 39,145 individuals (54% female, 15,805 cases, 23,340 controls). Across the PGC-MDD studies, structured diagnostic interviews were conducted to identify cases with a lifetime diagnosis of MDD according to the Diagnostic and Statistical Manual of Mental Health Disorders, 4th Edition³⁰, the International Classification of Diseases, 9th Edition (ICD-9)³¹, the International Classification of Diseases, 10th Edition (ICD-10)³² or the Composite International Diagnostic Interview Short Form (CIDI-SF)³³. In most of the PGC-MDD studies, bipolar disorder, non-affective psychosis and substance use disorder were exclusion criteria in the cases, and controls were screened for absence of MDD and other psychiatric disorders. Ethical approvals were obtained by Principal Investigators of each study, with all participants giving full informed consent.

The UKB is a prospective cohort study that has collected genotype and phenotype data for over 500,000 individuals across the UK, aged between 40 and 69 at the time of recruitment²⁹. 157,366 UKB participants completed an online Mental Health

Questionnaire (MHQ) which assesses lifetime depressive disorder³⁴. Using the recommended MHQ scoring protocol³⁴, we identified 29,344 individuals with lifetime depressive disorder and 63,358 controls. Cases were excluded if they endorsed diagnosis of psychosis or bipolar disorder. Controls were excluded if they endorsed diagnosis of any psychiatric disorder in the MHQ, or self-reported depression or use of anti-depressant medication at baseline and follow-up interviews, or suffered from a mood disorder according to Hospital Episode Statistics, or met the criteria for a mood disorder according to Smith, et al.³⁵. Further details of the PGC and UKB samples are in Supplementary Tables 1a and 1b. Access to data from the UKB was granted under application number 18177.

Genotyping and quality control

Genotyping arrays used in the PGC and UKB samples are shown in Supplementary Table 1a. Quality control (QC) of genotype data in the 26 PGC studies was performed by the PGC Statistical Analysis Group using the ricopili pipeline²⁴ such that variants and individuals below the following thresholds were retained: SNP missingness (before individual QC) < 0.05, individual missingness < 0.02, SNP missingness (after individual QC) < 0.02, deviation from heterozygosity $|F_{\text{het}}| < 0.20$, HWE p-value > 10^{-10} (cases) and p-value > 10^{-6} (controls). After imputation with the 1000 Genomes reference panel²⁴, SNPs with INFO > 0.8 and minor allele frequency (MAF) > 0.05 were retained for relatedness testing and principal component analysis. One individual from each pair with relatedness > 0.2 was removed and only individuals of European ancestry were retained for analysis.

Using genotype data that had undergone preliminary QC by the UKB²⁹, we created an inclusion list of individuals of European ancestry using 4-means clustering on the first two principal components provided by the UKB. We created an exclusion list of individuals using relatedness kinship (KING) estimates provided by the UKB, and removed one individual from each pair up to 3rd-degree relationships (KING $r^2 > 0.044$)³⁶. In the remaining data we applied QC such that variants and individuals below the following thresholds were retained: SNP missingness (before individual QC) < 0.02 , individual missingness < 0.02 , SNP missingness (after individual QC) < 0.02 , MAF > 0.01 , HWE p-value $> 10^{-8}$. The UKB²⁹ imputed SNPs using the IMPUTE4 software²⁹ with the Haplotype Reference Consortium (HRC) reference panel³⁷ and the UK10K Consortium reference panel³⁸ to produce dosage data in BGEN file format (version 1.2)³⁹. We extracted imputed SNPs from the classical class I MHC (chr6: 29,640,000-33,120,000) and converted to PLINK 2 binary format (pgen, pvar, psam) for association analyses in PLINK 2.0⁴⁰.

HLA allele and C4 haplotype imputation

HLA alleles were imputed using genotype data from the PGC studies using the SNP2HLA software²⁸ with the Type 1 Diabetes Genetics Consortium (T1DGC) reference panel⁴¹ to produce dosage data in Beagle format⁴². The T1DGC reference panel contains MHC haplotype information to enable imputation of HLA alleles at two-digit and four-digit resolution in eight HLA genes: *HLA-A*, *HLA-B* and *HLA-C* in the classical class I MHC and *HLA-DRB1*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DPA1* and *HLA-DPB1* in the classical class II MHC.

HLA alleles were imputed in the UKB by the core analytical team using the HLA*IMP:02 software²⁹ with multi-population reference panels⁴³. Collectively, the reference panels contained MHC haplotype information to enable imputation of HLA alleles in 11 HLA genes: *HLA-A*, *HLA-B* and *HLA-C* in the classical class I MHC and *HLA-DRB5*, *HLA-DRB4*, *HLA-DRB3*, *HLA-DRB1*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DPA1* and *HLA-DPB1* in the classical class II MHC. HLA alleles were encoded as bi-allelic in the PGC and UKB such that imputed dosages referred to the presence of 0, 1 or 2 copies of each HLA allele.

C4 haplotypes were imputed using genotype data in the PGC and UKB studies using the SNP2HLA software²⁸ with the C4 reference panel developed by the McCarroll lab⁴¹ (http://mccarrolllab.com/wp-content/uploads/2014/12/MHC_haplotypes_CEU_HapMap3_ref_panel.bgl) to produce dosage data in Beagle format⁴². The reference panel consists of SNP and C4 haplotypes within the extended MHC (25 to 34 Mb on chromosome 6) for 110 individuals from the HapMap CEU population. The reference panel contains 17 C4 haplotypes, defined by copy number variation of *C4A* and *C4B* genes in short and long form (defined by a HERV insertion). Four C4 haplotypes with frequency > 0.01 were retained for analysis: AL-AL, AL-BL, AL-BS and BS (where A and B correspond to two isotypes of the C4 gene and L and S correspond to the long and short forms). Three of the common C4 haplotypes (AL-AL, AL-BL and AL-BS) segregate on two, three and five different SNP haplotypes, respectively. Association results for these C4 haplotypes were calculated by meta-analysing across SNP haplotypes corresponding to each C4 structure.

Statistical analyses

In the PGC we tested each HLA allele and C4 haplotype for association with MDD case/control status using an additive logistic regression model applied to dosage data. We included 6 principal components (calculated by the PGC) to control for population structure. We extracted association results for SNPs in the classical class I MHC from PGC analyses in each study, applying a second layer of QC such that only variants with a minor allele frequency >0.01 and an info score ≥ 0.6 were retained. Post-QC variants were meta-analysed across the 26 PGC studies using an inverse-variance weighted approach.

In the UKB we tested each HLA allele, C4 haplotype and imputed SNP for association with depression case/control status using an additive linear regression model applied to dosage data. We regressed 6 principal components (calculated by the UKB), batch and centre on the depression phenotype using logistic regression in R 3.4.1⁴⁴ and used the residuals as the outcome variable in subsequent linear regression. We then filtered on variants with a minor allele frequency above 0.01 and an info score ≥ 0.6 before meta-analysing across the PGC and UKB results. Analyses were performed using PLINK (version 1.9 and version 2.0)⁴⁰.

To calculate the MHC region-wide significance threshold, we used the Genetic Type 1 Error Calculator⁴⁵, an online resource that calculates the number of effective tests by estimating linkage disequilibrium (LD) between variants and applying a Bonferroni correction. We calculated a conservative region-wide significance threshold ($3.9e-06$), controlling for all imputed SNPs in the classical MHC, and a candidate

significance threshold ($1.6e-4$), controlling only for HLA intragenic SNPs in the classical MHC.

The power calculation using the Genetic Power Calculator³⁹ showed that with a sample size of 45,149 cases and 86,698 controls, for an HLA allele of frequency 0.05 (the median in our study), we had 80% power to detect an odds ratio above 1.09, at a region-wide significance level of $\alpha = 3.9e-06$.

We compared the imputation accuracy and frequency of HLA alleles and C4 haplotypes in the PGC and UKB for variants present in both samples. The average imputation info score and frequency was calculated by weighting variant info scores and frequencies by the effective sample size in each PGC study calculated as $4 / (1/[\text{number of cases}] + 1/[\text{number of controls}])$ ⁴⁰.

Drawing on evidence from epidemiological studies^{9,13}, we identified autoimmune diseases with evidence for a bi-directional relationship with depression. We conducted a literature review on PubMed using the search terms "HLA" and the relevant autoimmune disease to identify HLA risk alleles. We filtered on HLA alleles with evidence for independent association ($p < 3.9e-06$) in European populations. We evaluated evidence for involvement of these HLA alleles in depression, selecting those with $MAF > 0.05$ in our study and applying a Bonferroni correction to adjust for the number of HLA alleles investigated ($0.05/15 = 0.003$).

A variable for genetically predicted C4A brain expression was calculated for each individual in the PGC and UKB samples. We leveraged work from Sekar, et al., who fit a linear model between C4 structures and C4A brain expression measured in post-mortem brain tissue²³. We used the betas from this model (provided in the

supplementary materials by Sekar, et al.) to estimate C4A brain expression corresponding to each C4 haplotype present in the reference panel (Supplementary Table 2). Individual scores for C4A brain expression were calculated by multiplying the dosage for each C4 haplotype by the corresponding value for C4A brain expression. We tested genetically predicted C4A brain expression for association with depression case/control status in the PGC and UKB samples, controlling for six principal components. Genetically predicted C4A brain expression was also used as a covariate in conditional analyses on HLA alleles.

Results

In total, 207 HLA alleles were imputed in at least two PGC studies, and 102 HLA alleles were imputed in the UKB, of which 93 were shared across data sets (Table 1). Variants imputed in either data set were included in the final meta-analysis (minimum effective sample size was 669 for HLA-B-3906 in the PGC). Four C4 haplotypes (AL-AL, AL-BL, AL-BS, BS) were imputed in all data sets.

Table 1: Variants imputed in at least two of the 26 PGC studies and the UKB.

Gene	PGC	UKB	Variants in both PGC and UKB	Variants in either PGC or UKB
<i>HLA-A</i>	31	13	13	31
<i>HLA-B</i>	48	18	18	48
<i>HLA-C</i>	30	14	14	30
<i>HLA-DPA</i>	5	3	3	5
<i>HLA-DPB</i>	25	11	11	25
<i>HLA-DQA</i>	12	7	7	12
<i>HLA-DQB</i>	19	12	12	19
<i>HLA-DRB</i>	37	24	15	46
Total HLA alleles	207	102	93	216
C4 haplotypes	4	4	4	4
SNPs	49,611	47,799	40,561	56,779

There was strong consistency between the frequency and info scores of HLA alleles and C4 haplotypes imputed in both the PGC and UKB (correlation $r=0.99$ for frequency and $r=0.86$ for info score; Supplementary Figures 1 and 2). For the 93 HLA alleles imputed in both data sets, the info score was consistently higher in the UKB compared to the PGC studies (UKB mean = 0.98, PGC mean = 0.96), possibly

due to the larger HLA reference panel, and/or greater efficiency of the imputation algorithm, used by the UKB.

Testing for association with depression in the PGC sample, no HLA allele, C4 haplotype or SNP surpassed region-wide significance (Figure 1a). In the UKB, no HLA allele or C4 haplotype surpassed region-wide significance (Figure 1b). The strongest evidence for association with an HLA allele was HLA-B*08:01 ($p = 4e-4$, OR = 0.98, 95% CI = 0.97-0.99). Among SNPs, 70 met region-wide significance (Supplementary Table 3). The variant with the lowest p-value ("UKB index variant") was a SNP in the classical class I region; rs1264373 ($p = 3.21e-7$, OR = 0.97, 95% CI = 0.96-0.98). All variants surpassing region-wide significance in the UKB sample were in LD with the UKB index variant ($0.66 < R^2 < 1.00$). The UKB index variant was also in LD with the most significant SNP within the MHC reported in the PGC MDD GWAS²⁴ (rs115507122, $R^2 = 0.63$).

In the UKB – PGC-MDD meta-analysis, no HLA allele or C4 haplotype met region-wide significance, however, HLA-B*0801 met the candidate threshold ($p = 1.26e-4$, OR = 0.98, 95% CI = 0.97-0.99) (Figure 1c). 143 SNPs reached region-wide significance (Supplementary Table 4). The variant with the lowest p-value ("meta-analysis index variant") was a SNP in the classical class I region; rs9262120 ($p = 8.74e-8$, OR = 1.03, 95% CI = 1.02-1.05). The meta-analysis index SNP was in LD with the other 142 significant variants ($0.44 < R^2 < 1.00$), and with the most significant SNP within the MHC in the PGC MDD GWAS²⁴ (rs115507122, named rs3095337 in this study, $R^2 = 0.66$), which did not reach region-wide significance in the meta-analysis ($p = 7.14e-6$, OR = 1.03, 95% CI = 1.01-1.04).

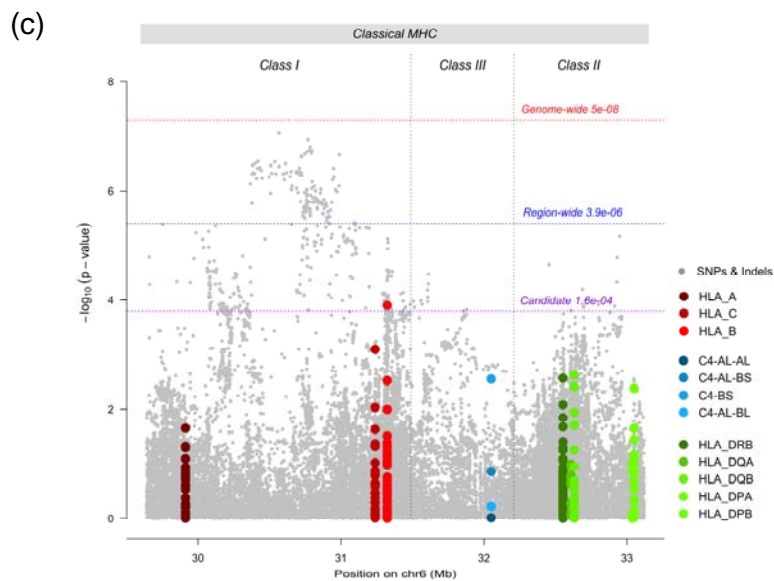
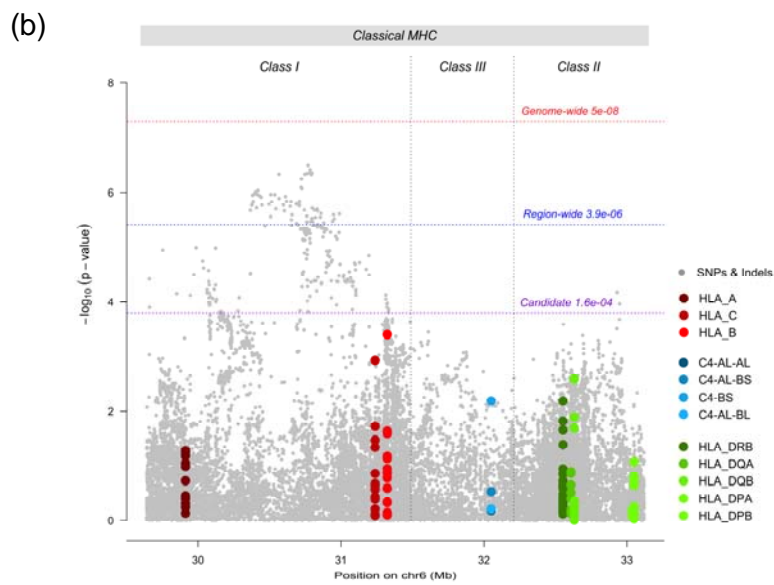
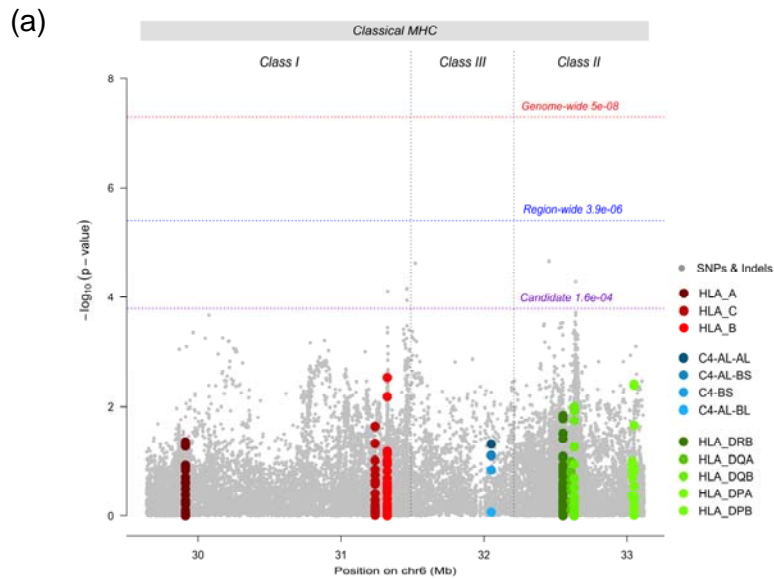


Figure 1: Region-wide Manhattan plots for SNPs (grey), HLA alleles (red/green) and C4 haplotypes (blue) in (a) PGC MDD studies, (b) UK Biobank, and (c) meta-analysis of PGC MDD studies and UK Biobank.

We identified six autoimmune diseases with evidence for a bi-directional relationship with depression: Crohn's disease, multiple sclerosis, primary adrenocortical insufficiency, psoriasis vulgaris, systemic lupus erythematosus (SLE) and type 1 diabetes mellitus^{9,13}. We identified 15 HLA alleles associated with risk for these autoimmune diseases ($p < 3.9e-6$) in European populations⁴⁶⁻⁵⁷, with a MAF > 0.05 in our study. Three HLA alleles had evidence for association with depression after correcting for multiple testing ($p < 0.003$): HLA-B*08:01 (for psoriasis and SLE), HLA-DQB1*02:01 (SLE) and HLA-DRB1*03:01 (multiple sclerosis, primary adrenocortical insufficiency, SLE) (Table 2).

Table 2: HLA alleles associated with risk for six autoimmune diseases and their association with depression in the PGC, UKB and meta-analysis. Bolded p-values met correction for multiple testing. Frq: allele frequency. OR: odds ratio.

Trait	HLA allele	PGC		UKB		META-ANALYSIS		
		Frq	OR	Frq	OR	OR	95% CI	p-value
Crohn's disease ^{46,47}	HLA-A*03:01	0.15	0.96	0.14	0.99	0.99	0.98-1.00	0.176
	HLA-C*06:02	0.09	1.00	0.09	1.02	1.02	1.00-1.04	0.043
	HLA-DRB1*07:01	0.13	1.01	0.14	1.01	1.01	1.00-1.02	0.179
	HLA-DRB1*13:02	0.05	0.97	0.04	0.99	0.99	0.97-1.01	0.431
Multiple sclerosis ^{48,49}	HLA-DQB1*03:02	0.11	1.00	0.10	1.00	1.00	0.99-1.01	0.537
	HLA-DRB1*03:01	0.13	0.95	0.15	0.98	0.98	0.97-0.99	0.003
	HLA-DRB1*15:01	0.14	0.98	0.14	1.00	0.99	0.98-1.00	0.355
Primary adrenocortical insufficiency (Addison's disease) ^{50,51}	HLA-DRB1*03:01	0.13	0.95	0.15	0.98	0.98	0.97-0.99	0.003
Psoriasis vulgaris ^{52,53}	HLA-A*02:01	0.28	1.01	0.27	1.01	1.01	1.00-1.02	0.005
	HLA-B*08:01	0.12	0.96	0.14	0.98	0.98	0.97-0.99	1.26e-4
	HLA-C*06:02	0.09	1.00	0.09	1.02	1.02	1.00-1.04	0.043
	HLA-DQA1*02:01	0.13	1.01	0.14	1.01	1.01	1.00-1.02	0.212
Systemic lupus erythematosus ^{54,55}	HLA-B*08:01	0.12	0.96	0.14	0.98	0.98	0.97-0.99	1.26e-4
	HLA-DQA1*01:02	0.20	0.97	0.19	0.99	0.99	0.98-1.00	0.163
	HLA-DQB1*02:01	0.13	0.95	0.15	0.98	0.98	0.97-0.99	0.002
	HLA-DRB1*03:01	0.13	0.95	0.15	0.98	0.98	0.97-0.99	0.003
Type 1 diabetes	HLA-A*24:02	0.08	0.97	0.07	1.01	1.00	0.98-1.02	0.578

mellitus ^{56,57}	HLA-DPB1*01:01	0.05	0.92	0.06	0.99	0.98	0.96-1.00	0.067
	HLA-DPB1*04:02	0.12	0.99	0.11	1.00	1.00	0.99-1.01	0.857

Genetically predicted C4A brain expression was not significantly associated with depression status in the PGC ($p = 0.066$, OR = 1.06, 95% CI = 1.00-1.13), the UKB ($p = 0.333$, OR = 1.01, 95% CI = 0.99-1.03), or the meta-analysis ($p = 0.150$, OR = 1.01, 95% CI = 0.99-1.03). Evidence for association with HLA-B*08:01 remained after conditioning on genetically predicted C4A brain expression with the p-value decreasing from $p = 1.26e-4$ to $p = 3.39e-5$ in the conditional analyses. The meta-analysis results for genetically predicted C4A brain expression and four common C4 haplotypes are summarised in Figure 2. The full results for HLA alleles, C4 haplotypes and conditional analyses in the PGC, UKB and meta-analysis are provided in Supplementary Table 5.

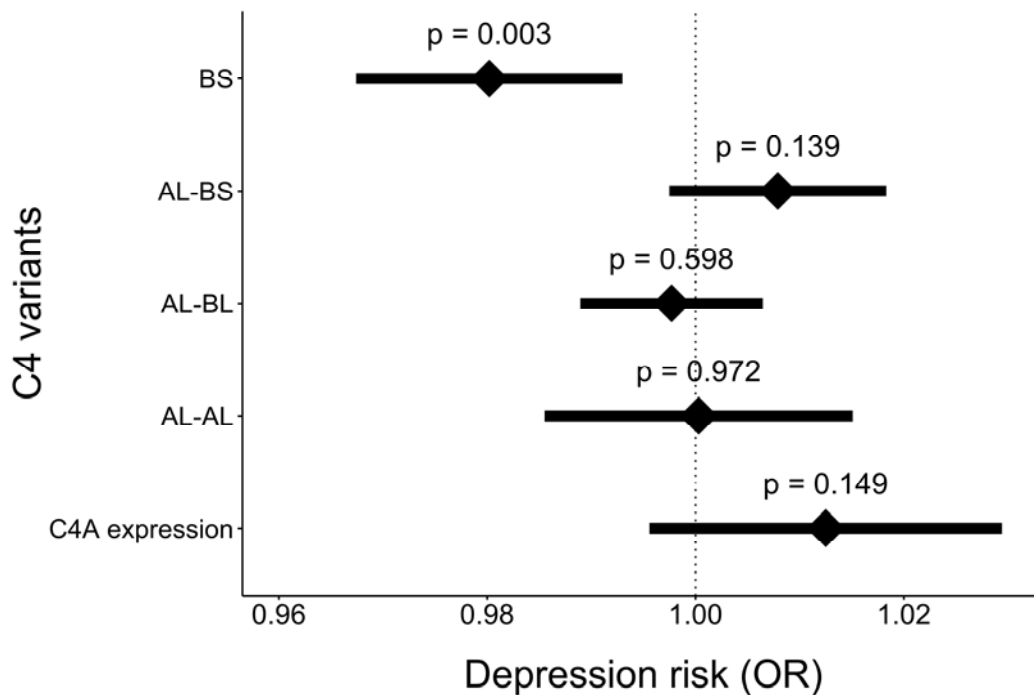


Figure 2: Association of genetically predicted C4A brain expression and four C4 haplotypes in the meta-analysis of PGC-MDD studies and UK Biobank. Error bars show 95% confidence intervals.

Discussion

To further understand the mechanisms driving comorbid autoimmunity and depression, we investigated evidence for shared genetic influences in the MHC, a region harboring genetic risk for autoimmune diseases and psychiatric disorders. Our primary aim was to test HLA alleles and C4 haplotypes for association with depression. Under a conservative region-wide significance threshold, we found no evidence that HLA alleles, which play a major role in susceptibility to autoimmune diseases, or C4 haplotypes, which associate strongly with risk for schizophrenia, also confer risk for depression. However, under a candidate threshold, HLA-B*0801 had significant evidence for association with depression status. Conditioning on genetically predicted C4A brain expression did not diminish signal from HLA-B*0801, indicating independence from C4 haplotypes.

We further explored common HLA alleles associated with autoimmune diseases that have evidence of a bi-directional relationship with depression. The strongest evidence for association with depression was HLA-B*08:01, followed by HLA-DQB1*02:01 and HLA-DRB1*03:01. Previous studies have shown that all three HLA alleles are risk increasing for SLE^{54,55}, HLA-DRB1*03:01 is also risk increasing for multiple sclerosis^{48,49} and primary adrenocortical insufficiency^{50,51}, and HLA-B*08:01 is also risk increasing for psoriasis^{52,53}. In contrast, our findings indicate that HLA-B*08:01, HLA-DQB1*02:01 and HLA-DRB1*03:01 have modest protective effects in depression, indicating these alleles do not harbor shared risk for autoimmune disease and depression.

Imputation of C4 haplotypes identified four common haplotypes, none of which were associated with risk for depression in the PGC studies, in the UKB or the meta-analysis. These results are in stark contrast to schizophrenia where association with C4 haplotypes

accounts for most of the observed SNP association in the HLA region. Our results suggest the C4 association does not contribute to the common genetic susceptibility between depression and schizophrenia, observed in a genetic correlation of $r_G=0.34$ ($p = 7.7e-40$) between these disorders²⁴.

At the level of region-wide significance, we detected 70 SNPs associated with depression in the UKB sample, and 143 in the meta-analysis. In each case, the top SNP was in moderate to strong LD with other significant variants, indicating a single peak of independent association. We found consistency in SNP signal between our study and the PGC-MDD GWAS of depression²⁴, with the top SNPs in each study showing moderate to strong LD. This was not unexpected given that our study is a subset of the studies included in the PGC-MDD meta-analysis²⁴.

The true identity of causal variants within the MHC remains unresolved, and fine-mapping within the MHC is challenging due to the high density of genetic variation and strong LD. Our results strongly suggest that the association signal observed in the MHC in depression^{24,25} does not arise from HLA alleles or C4 haplotypes. These results suggest that any associations with these variants are either rare or have very modest effect sizes. We note that Howard, et al.²⁵ increased power by leveraging a broader phenotyping approach. It is interesting to speculate that the broader depression phenotype captures individuals distressed by physical disease. This would go some way to explaining signal in the MHC, which has evidence for association with more diseases than another other region of the genome²⁰. However, a more parsimonious explanation could be that MHC signal in depression maps to SNPs or to other genetic loci not imputed in this study. This is highly plausible in light of the fact that the MHC contains more genes than any other region in the

human genome²⁰. Under this scenario, large sample sizes and sequencing may be required to dissect SNP signal within the MHC.

Although our findings do not support a role for HLA alleles within the MHC in risk for depression, it is possible that shared genetic risk for depression and autoimmune diseases is situated outside the MHC. Efforts to identify genome-wide pleiotropy were undertaken in the recent PGC-MDD GWAS using LD Score to estimate genetic correlations between depression and several autoimmune diseases²⁴. There was no evidence for significant cross-trait correlations; the strongest observed was between depression and inflammatory bowel disease ($r_G = 0.07$, $p = 0.06$). In other efforts to detect genome-wide pleiotropy, Euesden et al.¹⁴ found no evidence that polygenic risk scores (PRS) for rheumatoid arthritis predicted depression status in an independent sample, nor did PRS for depression predict autoimmune disease status.

One possibility is that there is a sub-group of individuals enriched for depression and autoimmune risk alleles. Under this scenario, there may be insufficient power to detect the relationship. Identifying, for example, a sub-group of individuals with depression, who are also enriched for autoimmune risk alleles, would go some way to explaining the observed comorbidity between these traits. Furthermore, identifying a sub-type of depression, e.g. an 'immune related' depression group, would help to dissect heterogeneity in the depression phenotype.

In summary, this study is the first to interrogate the involvement of HLA alleles and C4 haplotypes in depression risk, and we find no evidence that either type of genetic variant plays a major role in susceptibility for depression. In contrast, the three HLA alleles that showed nominal significance in our study conferred modest protective effects for

depression. Furthermore, the strong association for C4 alleles seen in schizophrenia is absent in depression. Large sample sizes and regional sequence data may be required to dissect SNP signal within the MHC.

Acknowledgements

This work was supported by a PhD studentship awarded to K.P.G. from the UK Medical Research Council. PFO receives funding from the UK Medical Research Council (MR/N015746/1).

This paper represents independent research part-funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

We thank participants and scientists involved in making the UK Biobank resource available (<http://www.ukbiobank.ac.uk/>). UK Biobank data used in this study were obtained under approved application 18177.

We are deeply indebted to the investigators who comprise the PGC, and to the hundreds of thousands of subjects who have shared their life experiences with PGC investigators. The PGC has received major funding from the US National Institute of Mental Health and the US National Institute of Drug Abuse (U01 MH109528 and U01 MH1095320). Acknowledgements and funding sources for each of the primary studies in the PGC are contained in supplementary materials.

Statistical analyses were carried out on the NL Genetic Cluster Computer (<http://www.geneticcluster.org/>) hosted by SURFsara, and the King's Health Partners High Performance Compute Cluster funded with capital equipment grants from the GSTT Charity (TR130505) and Maudsley Charity (980).

Competing interests

J.E. is an employee of GlaxoSmithKline Pharmaceuticals. P.F.S. is on the scientific advisory board for Pfizer, Inc., and the advisory committee for Lundbeck. No other authors have anything to declare.

Previous presentation of this data

K.P.G presented a poster entitled "Genetic variation in the Major Histocompatibility Complex and association with depression" at the World Congress of Psychiatric Genetics in Glasgow, UK on 14th October 2018.

References

1. Verbeek, E. C., Bevova, M. R., Hoogendijk, W. J. G. & Heutink, P. The Genetics of MDD – A Review of Challenges and Opportunities. *Journal of Depression and Anxiety* **03**, 1–6 (2014).
2. World Health Organization. Depression and other common mental disorders: global health estimates. *World Health Organisation* (2017).
3. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (V)*. (American Psychiatric Association, 2013).
4. Garcia-Gonzalez, J. *et al.* Pharmacogenetics of antidepressant response: A polygenic approach. *Progress in neuro-psychopharmacology & biological psychiatry* **75**, 128–134 (2017).
5. Breen, G. *et al.* Translating genome-wide association findings into new therapeutics for psychiatry. *Nat. Neurosci.* **19**, 1392–1396 (2016).
6. Niamh Mullins, C. M. L. Genetics of Depression: Progress at Last. *Current Psychiatry Reports* **19**, 1–7 (2017).
7. Sullivan, P. F. *et al.* Psychiatric Genomics: An Update and an Agenda. *American Journal of Psychiatry* **175**, 15–27 (2017).
8. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nature Genetics* **47**, 1236–1241 (2015).
9. Benros, M. E. *et al.* Autoimmune Diseases and Severe Infections as Risk Factors for Mood Disorders A Nationwide Study. *JAMA Psychiatry* **70**, 812–820 (2013).
10. Anderson, R. J., Freedberg, K. A., Clouse, R. E. & Lustman, P. J. The Prevalence of Comorbid Depression in Adults With Diabetes. *Diabetes Care* **24**, 1069–1078 (2001).
11. Patten, S. B., Beck, C. A., Williams, J. V. A., Barbui, C. & Metz, L. M. Major

- depression in multiple sclerosis: A population-based perspective. *Neurology* 1524–1527 (2003).
12. Kurina, L., Goldacre, M., Yeates, D. & Gill, L. Depression and anxiety in people with inflammatory bowel disease. *Journal of Epidemiology and Community Health* **55**, 716–720 (2001).
 13. Andersson, N. W. *et al.* Depression and the risk of autoimmune disease: a nationally representative, prospective longitudinal study. *Psychol Med* **45**, 3559–3569 (2015).
 14. Euesden, J., Danese, A., Lewis, C. M. & Maughan, B. A bidirectional relationship between depression and the autoimmune disorders - New perspectives from the National Child Development Study. *PLOS ONE* **12**, 1–14 (2017).
 15. Dooley, M. A. & Hogan, S. L. Environmental epidemiology and risk factors for autoimmune disease. *Current Opinion in Rheumatology* **15**, 99–103 (2003).
 16. Waszczuk, M. A., Zavos, H. M. S., Gregory, A. M. & Eley, T. C. The Phenotypic and Genetic Structure of Depression and Anxiety Disorder Symptoms in Childhood, Adolescence, and Young Adulthood. *JAMA Psychiatry* **71**, 905–916 (2014).
 17. Cruz-Tapias, P., Castiblanco, J. & Anaya, J. in *Autoimmunity: From Bench to Bedside* (El Rosario University Press, 2013).
 18. Dowlati, Y. *et al.* A meta-analysis of cytokines in major depression. *Biological Psychiatry* **67**, 446–457 (2010).
 19. Morgan, J. If other ways don't work, try the immune system? *The Lancet Neurology* **16**, 109 (2017).
 20. Horton, R. *et al.* Gene map of the extended human MHC. *Nat. Rev. Genet.* **5**, 889–899 (2004).
 21. Dendrou, C. A., Petersen, J., Rossjohn, J. & Fugger, L. HLA variation and disease.

- Nature Reviews Immunology* **18**, 325–339 (2018).
22. Gough, S. C. L. & Simmonds, M. J. The HLA region and autoimmune disease: Associations and mechanisms of action. *Current Genomics* **8**, 453–465 (2007).
 23. Sekar, A. *et al.* Schizophrenia risk from complex variation of complement component 4. *Nature* **530**, 177–183 (2016).
 24. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics* **50**, 668–681 (2018).
 25. Howard, D. M. *et al.* Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nature Communications* **9**, 1–10 (2018).
 26. Howard, D. M. *et al.* Genome-wide meta-analysis of depression in 807,553 individuals identifies 102 independent variants with replication in a further 1,507,153 individuals. *bioRxiv* 433367 (2018). doi:10.1101/433367
 27. Hyde, C. L. *et al.* Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nature Genetics* **48**, 1031–1036 (2016).
 28. Jia, X. M. *et al.* Imputing Amino Acid Polymorphisms in Human Leukocyte Antigens. *PLOS ONE* **8**, 1–10 (2013).
 29. Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv* 166298 (2017). doi:10.1101/166298
 30. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (IV)*. (American Psychiatric Association, 1994).
 31. World Health Organization. *International Classification of Diseases (9th Edition)*.

- (World Health Organization, 1978).
32. World Health Organization. *International Classification of Diseases (10th Edition)*. (World Health Organisation, 1992).
 33. Kessler, R. C., Andrews, G., Mroczek, D., Ustun, B. & Wittchen, H. U. The World Health Organization Composite International Diagnostic Interview short-form (CIDI-SF). *International Journal of Methods in Psychiatric Research* **7**, 171–185 (2006).
 34. Davis, K. A. S. *et al.* Mental health in UK Biobank: development, implementation and results from an online questionnaire completed by 157 366 participants. *BJPsych Open* **4**, 83–90 (2018).
 35. Smith, D. J. *et al.* Prevalence and Characteristics of Probable Major Depression and Bipolar Disorder within UK Biobank: Cross-Sectional Study of 172,751 Participants. *PLOS ONE* **8**, e75362 (2013).
 36. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
 37. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics* **48**, 1279–1283 (2016).
 38. UK10K Consortium *et al.* The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90 (2015).
 39. Band, G. & Marchini, J. BGEN: a binary file format for imputed genotype and haplotype data. *bioRxiv* 308296 (2018). doi:10.1101/308296
 40. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 1–16 (2015).
 41. The Type 1 Diabetes Genetics Consortium *et al.* Genome-wide association study and

- meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nature Genetics* **41**, 703–707 (2009).
42. Browning, B. L. & Browning, S. R. A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. *The American Journal of Human Genetics* **84**, 210–223 (2009).
 43. Diltthey, A. *et al.* Multi-Population Classical HLA Type Imputation. *PLoS Comput Biol* **9**, e1002877 (2013).
 44. R Development Core Team. R: A language and environment for statistical computing. (2008).
 45. Li, M. X., Yeung, J. M. Y., Cherny, S. S. Sham, P. C. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Human Genetics* **131**, 747–756 (2012).
 46. Goyette, P. *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. *Nature Genetics* **47**, 172–179 (2015).
 47. Ahmad, T. Genetics of inflammatory bowel disease: The role of the HLA complex. *WJG* **12**, 3628 (2006).
 48. Patsopoulos, N. A. *et al.* Fine-Mapping the Genetic Association of the Major Histocompatibility Complex in Multiple Sclerosis: HLA and Non-HLA Effects. *PLoS Genet.* **9**, e1003926 (2013).
 49. the International Multiple Sclerosis Genetics Consortium *et al.* Class II HLA interactions modulate genetic risk for multiple sclerosis. *Nature Genetics* **47**, 1107–1113 (2015).

50. Skinningsrud, B. *et al.* Multiple Loci in the HLA Complex Are Associated with Addison's Disease. *J Clin Endocrinol Metab* **96**, E1703–E1708 (2011).
51. Myhre, A. G. *et al.* Autoimmune Adrenocortical Failure in Norway Autoantibodies and Human Leukocyte Antigen Class II Associations Related to Clinical Features. *J Clin Endocrinol Metab* **87**, 618–623 (2002).
52. Okada, Y. *et al.* Fine Mapping Major Histocompatibility Complex Associations in Psoriasis and Its Clinical Subtypes. *The American Journal of Human Genetics* **95**, 162–172 (2014).
53. Chen, H. *et al.* Psoriasis Patients Are Enriched for Genetic Variants That Protect against HIV-1 Disease. *PLoS Genet.* **8**, e1002514 (2012).
54. Morris, D. L. *et al.* Unraveling Multiple MHC Gene Associations with Systemic Lupus Erythematosus: Model Choice Indicates a Role for HLA Alleles and Non-HLA Genes in Europeans. *The American Journal of Human Genetics* **91**, 778–793 (2012).
55. Hanscombe, K. B. *et al.* Genetic fine mapping of Systemic Lupus Erythematosus MHC associations in Europeans and African Americans. *Human Molecular Genetics* (2018). doi:10.1093/hmg/ddy280
56. Hu, X. *et al.* Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. *Nature Genetics* **47**, 898–905 (2015).
57. Howson, J. M. M., Walker, N. M., Clayton, D., Todd, J. A. Type 1 Diabetes Genetics Consortium. Confirmation of HLA class II independent type 1 diabetes associations in the major histocompatibility complex including HLA-B and HLA-A. *Diabetes, obesity & metabolism* **11**, 31–45 (2009).