

Word count (abstract)	297
Word count (main text)	5561
Word count (no methods)	3137
Figures	5
Tables	2
Supplementary Information	1
Supplementary Data	2

DNA methylome-wide association study of genetic risk for depression implicates antigen processing and immune responses

Xueyi Shen^{1*}, Doretta Caramaschi², Mark J Adams¹, Rosie M Walker³, Josine L Min², Alex Kwong^{1,2}, Gibran Hemani², Genetics of DNA Methylation Consortium², Miruna C Barbu¹, Heather C Whalley¹, Sarah E Harris⁵, Ian J Deary⁵, Stewart W Morris⁴, Simon R Cox⁵, Caroline L Relton², Riccardo E Marioni⁴, Kathryn L Evans⁴, Andrew M McIntosh^{1*}

¹ Division of Psychiatry, University of Edinburgh, Edinburgh, United Kingdom

² Medical Research Council Integrative Epidemiology Unit, Bristol Medical School, Population Health Sciences, University of Bristol, Bristol, United Kingdom

³ Centre for Clinical Brain Sciences, Chancellor's Building, 49 Little France Crescent, Edinburgh BioQuarter, Edinburgh, UK

⁴ Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, United Kingdom

⁵ Lothian Birth Cohorts group, Department of Psychology, The University of Edinburgh, UK

*Corresponding authors:

Dr Xueyi Shen

E-Mail: xueyi.shen@ed.ac.uk

Address:

University of Edinburgh, Royal Edinburgh Hospital, Morningside Park
Edinburgh EH10 5HF
United Kingdom

Prof Andrew M McIntosh

E-Mail: andrew.mcintosh@ed.ac.uk

Address:

University of Edinburgh, Royal Edinburgh Hospital, Morningside Park
Edinburgh EH10 5HF
United Kingdom

Abstract

Background Depression is a disabling and highly prevalent condition where genetic and epigenetic differences, such as DNA methylation (DNAm), contribute to prediction of disease liability.

Method We investigated the association between polygenic risk scores (PRS) for depression and DNAm by conducting a methylome-wide association study (MWAS) in Generation Scotland (N=8,898, mean age=49.8 years) with replication in the Lothian Birth Cohorts of 1921 and 1936 and adults in Avon Longitudinal Study of Parents and Children (ALSPAC) (N_{combined}=2,049, mean age=79.1, 69.6 and 47.2 years, respectively). We also conducted a replication MWAS in the ALSPAC children (N=423, mean age=17.1 years).

Result Wide-spread associations were found between PRS constructed using genetic risk variants for depression and DNAm in cytosine-guanine dinucleotide (CpG) probes that localised to genes involved in immune responses and neural development (N_{CpG}=599, $p_{\text{Bonferroni}} < 0.05$, $p < 6.5 \times 10^{-8}$). The effect sizes for the significant associations were highly correlated between the discovery and replication samples in adults ($r=0.83$) and in adolescents ($r=0.76$). Additional analysis on the methylome-wide associations was conducted for each lead genetic risk variant. Over 40% of the independent genetic risk variants showed associations with CpG probe DNAm located in both the same (*cis*) and distal probes (*trans*) to the genetic loci ($p_{\text{Bonferroni}} < 0.045$). Subsequent Mendelian randomisation analysis showed that DNAm and depression are mutually causal ($p_{\text{FDR}} < 0.039$), and there is a greater number of causal effects found from DNAm to depression (DNAm to depression: p_{FDR} ranged from 0.045 to 2.06×10^{-120} ; depression to DNAm: p_{FDR} ranged from 0.046 to 2.1×10^{-23}).

Conclusion Polygenic risk scores for depression, especially those constructed from genome-wide significant genetic risk variants, showed epigenome-wide methylation association differences in the methylome associated with immune responses and brain development. We also found evidence from Mendelian randomisation evidence that DNAm may be causal to depression, as well as a causal consequence of depression.

Introduction

Depression is a highly prevalent condition and a leading cause of global disability¹, for which the underlying biological mechanisms are unclear. Genetic factors account for a substantial proportion of differences in liability to depression, which has a twin-based heritability of approximately 37% and with common genetic variants capturing around 6-10% of phenotypic variance^{2,3}. Recent genome-wide association studies (GWAS) have identified specific genetic risk variants for depression that implicate regional brain alterations^{4,5}. Polygenic risk scores (PRS) derived from the results of GWAS studies, have been widely used to estimate additive genetic risk⁶. PRS provides a means to identify traits that share their genetic architecture with depression, which may help to prioritise factors of biological and mechanistic relevance for the disorder⁷.

DNA methylation (DNAm) at cytosine- guanine dinucleotides (CpG) sites is one of the most studied epigenetic markers and there is growing evidence of its role in understanding depression⁸. DNAm risk scores have been developed from the results of DNA methylome-wide association studies (MWAS)⁸. These can be used to predict prevalent depression in independent samples, and chronic depression that requires long-term treatment⁹. DNAm is influenced by both genetic and environmental factors^{8,10} and, in blood tissue, it has a mean heritability of 19% across the epigenome¹¹ with ~7% of its variance captured by common genetic variants¹². For the highly heritable DNA methylation probes, genetic effects are consistent across tissues¹³ and developmental stages¹². Genetic risk variants for diseases, such as schizophrenia, have been found enriched in DNA methylation variation¹⁴⁻¹⁶. Associations between genetic risk and epigenetic changes can enrich our understanding of the functional composition of genetic risk loci, and thus inform the mechanisms that lead to the onset of depression^{17,18}. However, systematic examination of the molecular genetic associations between genetic risk of depression and DNAm has not, to the best of our knowledge, been conducted.

In the present study, we aim to investigate the association between PRS for depression and genome-wide DNA methylation. Methylome-wide association studies (MWAS) were conducted on four cohorts: Generation Scotland: Scottish Family Health Study (GS, discovery sample, N=8,898), the Lothian Birth Cohort (LBC1921), the Lothian Birth Cohort 1936 (LBC1936), Avon Longitudinal Study of Parents and Children (ALSPAC) adults (adult replication sample, combined N=2,049) and ALSPAC children for replication (adolescent replication sample, N=423). Mendelian randomisation was used to test for causal associations between DNAm and depression using data from the Genetics of DNA Methylation Consortium (GoDMC) (N=25,561) and GS.

Methods

Sample descriptions

GS

Generation Scotland (GS) is a family-based population cohort with over 24,000 participants^{19,20} set-up to identify the causes of common complex disorders, such as depression. DNA methylation data and genetic data were both collected, processed and quality-checked for 8,898 people (mean age=49.8 years, SD of age=13.7 years, 40.90% were men) in two sets. Sample sizes for set 1 and set 2 were 4757 (mean age=48.5 years, SD of age=14.0 years, 38.5% were men) and 4141 (mean age=51.4 years, SD of age=13.2 years, 43.66% were men), respectively. Written informed consent was obtained for all participants. The study was approved by the NHS Tayside Research Ethics committee (05/s1401/89).

Lothian Birth Cohorts (LBC) 1921 and 1936

Participants from LBC 1921 and LBC 1936^{21,22} were born in 1921 and 1936. Almost all lived in the Edinburgh and surrounding Lothian area when recruited. They are a mostly healthy, community-dwelling sample of men and women. The sample used in the current analysis included 1,330 participants from both cohorts combined with genetic and DNA methylation data (LBC 1921: mean age=79.1 years, SD of age=0.6, 39.7% were men; LBC 1936: mean age=69.6 years, SD of age=0.8, 50.6% were men; all participants were unrelated). Written informed consent was obtained from all participants. Ethics permission for LBC1921 was obtained from the Lothian Research Ethics Committee (LREC/1998/4/183). Ethics permission for LBC1936 was obtained from the Multi-Centre Research Ethics Committee for Scotland (MREC/01/0/56) and the Lothian Research Ethics Committee (LREC/2003/2/29)^{23,24}.

Avon Longitudinal Study of Parents and Children (ALSPAC)

The Avon Longitudinal Study of Parents and Children (ALSPAC) is an ongoing longitudinal population-based study that recruited pregnant women residing in Avon (South-West of England) with expected delivery dates between 1st April 1991 and 31st December 1992^{25,26}. The cohort consists of 13,761 mothers and their partners, and their 14,901 children, now young adults²⁷. The study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool (<http://www.bristol.ac.uk/alspac/researchers/our-data/>). Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. A subsample of 719 unrelated mothers with DNA methylation (DNAm) data (mean age=47.2 years, SD of age=4.6) were included in the replication study²⁸. Supplementary analyses were also conducted on a younger subsample with DNAm consisting of 423 young people (mean age=17.1 years, SD of age=1.1 and 41% were boys). Details of the selection of participants for these subsamples are in the study by Relton et al²⁸. Consent for biological samples has been collected in accordance with the Human Tissue Act (2004).

Genotyping and imputation

Detailed information on the quality control and genotyping methods for GS¹⁹, LBC1921,

LBC1936²⁹ and ALSPAC³⁰ have been previously published and is described briefly below. Analyses were conducted on European participants.

GS

Each sample was genotyped using the IlluminaHumanOmniExpressExome-8v1.0 BeadChip (48.8%) or Illumina HumanOmniExpressExome-8 v1.2 BeadChip (51.2%) with Infinum chemistry³¹. Quality control included removing participants with genotyping call rate<98%, SNP removal of those with a minor allele frequency (MAF)<1%, call rate<98%, Hardy-Weinberg equilibrium (HWE) p-value<5×10⁻⁶. Imputation was performed using the Sanger Imputation server with the Haplotype Reference Consortium reference panel (HRC.r1-1). SNPs with INFO score<0.8 were removed from analysis.

LBC1921 and LBC1936

Genotyping was performed using the Illumina610-Quadv1 chip (Illumina, Inc., San Diego, CA, USA). Participants were excluded with a call rate<95%. SNPs were removed if MAP<5%, call rate<98%, HWE p-value<0.001. Imputation and quality control based on INFO score were the same as GS.

ALSPAC

Genotyping arrays used were the Illumina Human660W-quad chip for mothers and Illumina HumanHap550-quad chip for children. SNPs with missingness>0.05, HWE p-value<1×10⁻⁶, MAF<0.01 were excluded. The above quality control steps were conducted on the entire genotyped sample. Imputation and quality control based on INFO score were consistent with similar procedures used in GS.

Polygenic profiling

Polygenic risk scores (PRS) of depression were calculated using PRSice-2³² for GS, LBC1921, LBC1936 and ALSPAC separately, using the summary statistics of a genome-wide meta-analysis of depression by Howard *et al.*³³ excluding individuals from GS previously included in that GWAS meta-analysis. Nine p-value cut-offs were used for thresholding SNPs in the summary statistics (pT): 1, 0.5, 0.1, 0.05, 0.01, 1×10⁻³, 1×10⁻⁴, 1×10⁻⁵ and 5×10⁻⁸ for clumping and thresholding. Each set of SNPs was used to generate a depression-PRS in GS. A separate PRS was generated using the lead genetic risk variants or their closest proxies (in LD r²>0.1) reported in the GWAS by Howard *et al.*³³ for supplementary analysis. Details of the PRS profiling procedures can be found elsewhere³³ (also see Supplementary Information and Supplementary Table 1-2).

Subsequently, using the lead risk variants reported by Howard *et al.*³³, we tested for individual SNP-CpG associations in GS. Lead risk variants were selected by extracting the most significant proxy SNPs (p<5×10⁻⁸) in linkage disequilibrium (LD R²>0.01) with the lead variants reported in the Howard *et al.* study⁴. A total of 96 SNPs were available and thus selected as leading risk variants for further analysis.

DNA methylation data

GS

Genome-wide DNA methylation data was obtained from whole blood samples using the Illumina Infinium Methylation EPIC array (https://emea.support.illumina.com/array/array_kits/infinium-methylationepic-beadchip-kit.html). Data processing was performed separately for each set. Quality control (QC) and normalisation were conducted using R packages 'ShinyMethyl' (version 1.28.0)³⁴, 'watermelon' (version 1.36.0)³⁵ and 'meffil' (version 1.1.1)³⁶. Details of the protocol are described elsewhere³⁷. In summary, quality control procedures removed probes if there was an outlying log median methylated signal intensity against unmethylated signal for each array, or a bead count <3 in ≥5 % of the total probe sample, or a detection p-value >0.05 for set 1 and p-value >0.01 for set 2 in ≥0.5% of the total sample in each respective set. Cross-hybridising probes that map to genetic variants at MAF >0.05 and polymorphic probes were removed³⁸. Samples were excluded if sex prediction from methylation data was inconsistent with self-reported data, or a detection p-value >0.05 for set 1 and p-value >0.01 for set 2 found in >1% of the overall probes for each set respectively. The data was then normalised using the 'dasen' method from the 'waterRmelon' R package (version 1.36.0).

The raw intensities were then transformed into M-values by log-transforming the proportional methylation intensity³⁹. The M-values were corrected using a linear-mixed model, controlling for relatedness using the GCTA-estimated genetic relationship matrix⁴⁰ for set 1. This step was omitted for set 2 as all participants were unrelated within the set and to set 1. The residualised M-values for over 750,000 autosomal CpG probes were then used for further analysis.

LBC1921 and LBC1936

Genome-wide DNA methylation data was obtained from blood sample using the

HumanMethylation450K array (https://emea.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_humanmethylation450.pdf)^{41,42}.

Quality control and normalisation were performed using the 'minfi' R package (version 1.38.0)⁴¹. Probes with low call rate (<95%), outlying M-values (>3 SD from mean) or identified as cross-hybridising and polymorphic were removed⁴³. Participants with insufficient cell count information were excluded from analysis.

All participants in LBC1921 and LBC1936 with methylation data were unrelated. M-value transformation were conducted consistently with the GS sample. Data for over 450,000 CpG probes were retained for further analysis.

ALSPAC

Illumina Infinium HumanMethylation450 Beadchip arrays were used for measuring genome-wide DNA methylation data from peripheral blood sample. The R package 'meffil' (version 1.1.1) was used for pre-processing, quality control and normalisation as previously reported⁴⁴. Further removal of probes was conducted based on background detection (p > 0.05) and if they reached beyond the 3 times inter-quantile range from 25% and 75% or identified as cross-hybridising or polymorphic⁴³. Related (IBD > 0.1) participants were not included in the analyses³⁰.

M-value transformation was conducted and over 468,000 CpG probes remained for analysis.

Statistical models for methylome-wide association analysis (MWAS)

A discovery MWAS was initially conducted in GS. Two separate analyses were conducted on sets 1 and 2, and the final summary statistics were obtained by meta-analysing the two sets of results using METAL (version released in 2011)⁴⁵. We used the default analysis scheme without genomic control correction (genomic inflation factors reported in the Supplementary Information). P-values for the meta-analysis were obtained from a fixed-effect inverse-variance model. A replication analysis on adults was then conducted on the total sample from LBC1921, LBC1936 and ALSPAC adults. Replication MWAS was first conducted separately for each cohort and then meta-analysed using the same parameters for the discovery analysis. Finally, an additional replication MWAS was conducted on ALSPAC children. All analyses were conducted using R (version 3.5.1) under Linux environment.

Linear regression was used to test the associations between depression-PRS and for each CpG using the R package 'limma'⁴⁶ (version 3.48.0) for GS, LBC1921 and LBC1936. The 'lmFit' function was first implemented to test the association for each CpG. The inference statistics of each linear model was then adjusted using the 'eBayes' function, by which an empirical Bayes method was used to adjust for gene-wise variance using a shrinkage factor. Moderated t-statistics and p-values were produced by this step. In ALSPAC, the analyses were conducted using the 'meffil' (version 1.1.1) R package, using the 'sva' option⁴⁴.

Self-reported smoking status, smoking pack years, DNAm-estimated white-blood cell proportions (CD8+T, CD4+T, natural killer cells, B cells and granulocytes)³⁷, batch, the first 20 principal component derived from the M-values, age and sex were included as covariates for the discovery methylome-wide association analysis (see Supplementary Information and Supplementary Table 3 for details). Where possible, the same covariates were used in the replication analyses, although only smoking status (and not pack years) were available in LBC 1921, LBC 1936 and ALSPAC. Details for all the covariates included in the replication analysis can be found in the Supplementary Information. MWAS were conducted for the nine depression-PRS scores separately. P-values were Bonferroni-corrected (p-value threshold = 6.5×10^{-8} for EPIC array used in GS, 1.1×10^{-7} for 450k array used for replication analysis in LBC1921, LBC1936 and ALSPAC adults, and 1.1×10^{-7} for replication in ALSPAC children). Standardised regression coefficients are reported as effect sizes. For the significant CpG probes, gene symbol annotation and UCSC classification of CpG Island positions were acquired from the 'UCSC_RefGene_Name' and 'Relation_to_Island' columns, respectively, from the annotation object generated by the 'IlluminaHumanMethylationEPICanno.ilm10b4.hg19' R package (version 3.13)⁴⁷.

Individual SNP-CpG DNAm association tests were also performed, using the same covariates and p-value corrections as used in the PRS association analyses.

Gene ontology analysis

Gene ontology analysis was conducted on the MWAS results from GS using the 'gometh' function in R package missMethyl⁴⁸. Default settings were used, which include correction for the number of probes per gene. CpGs that showed significant association with depression-PRS at $pT < 5 \times 10^{-8}$ in the discovery analysis were selected as CpGs of interest, 'EPIC' was

chosen for array type and all CpGs included in the analysis were used as the background list. FDR-correction was applied for all analyses.

Colocalisation analysis

We used Howard et al.'s (2019) MDD GWAS⁴ for MDD-associated SNPs and GoDMC summary statistics for mQTL analysis with PGC studies and GS study removed, which resulted in 32 remaining studies imputed to the 1000 Genome reference panel¹³. We used the package “gwasglue” (version 0.0.0.9000, <https://mrcieu.github.io/gwasglue/>) to extract SNPs that were +/- 1Mb of each of the 102 genome-wide significant, lead SNPs identified in Howard et al. (2019) and then extracted the same SNPs within those regions from the GoDMC mQTL analysis. We used the `coloc.abf` function with default parameters in the “coloc” package in R (version 5.1.0)⁴⁹ to perform colocalisation analysis for each SNP association. The method tests for five mutually exclusive scenarios in a genetic region: H_0 : there exist no causal variants for either trait; H_1 : there exists a causal variant for trait one only; H_2 : there exists a causal variant for trait two only; H_3 : there exist two distinct causal variants, one for each trait; and H_4 : there exists a single causal variant common to both traits.

Mendelian randomisation (MR)

Three MR methods, inverse-weighted median (IVW), weighted median (WM) and MR Egger, were used to test for causal effects between DNA methylation and depression using the ‘TwoSampleMR’ R package (version 0.5.6)^{50,51}.

GWAS summary statistics for DNAm were from GoDMC and GS. mQTL summary statistics from GoDMC included 32 cohorts with 25,561 participants from European ancestry¹³. The summary statistics were computed using a two-phased design. First, every study performs a full analysis of all candidate mQTL associations, returning only associations at a threshold of $p < 1 \times 10^{-5}$. All candidate mQTL associations at $p < 1 \times 10^{-5}$ are combined to create a unique ‘candidate list’ of mQTL associations. The candidate list ($n=120,212,413$) is then sent back to all cohorts, and the association estimates are obtained for every mQTL association on the candidate list. Candidate mQTL associations were meta-analysed using fixed-effect inverse-variance method. Details of the database can be found elsewhere¹³. mQTL summary statistics from GS ($N \sim 10,000$) included a full set of all SNPs with no p-value thresholding. Summary mQTL statistics from GS were generated using the OmicS-data-based Complex Trait Analysis package (<https://cnsngenomics.com/software/osca/#eQTL/mQTLAnalysis>)⁵². Covariates were consistent with the MWAS for depression-PRS discovery analysis. Further details of the mQTL analysis can be found in the Supplementary Information.

Summary statistics for depression GWAS by Howard *et al.*³³ were used. A total of 807,553 unrelated, European participants were included in the analysis. Details for the study can be found elsewhere³³.

GoDMC, GS and depression GWAS samples were mutually exclusive. Individual cohorts that overlapped with the Howard *et al.* depression GWAS and GS were removed from the GoDMC mQTL meta-analysis. Depression GWAS summary statistics from Howard et al (2019) were calculated excluding GS participants³³. See Supplementary Information for details.

First, we used mQTL summary statistics from GoDMC to identify causal effects from DNA methylation to depression. Second, we used full mQTL summary statistics from GS to replicate the findings and to test the causal effect to DNA methylation and depression bi-directionally. Finally, in contrast to the univariable MR analyses (that is, each risk-outcome pair tested separately), a multi-variable MR analysis was conducted to test for direct causal associations from DNAm at multiple CpGs to MDD, using the mQTL data from GS. Using this method, all CpG probes where there was evidence of a potential causal effect on depression were entered into the two-sample MR analysis simultaneously, in order to prioritise SNPs that showed the strongest independent casual associations with MDD.

Exposures were selected from CpG probes found significantly associated with depression-PRS generated using the p-threshold = 5×10^{-8} . As the *MHC* region contains a large number of highly correlated probes, we selected the four independent probes after pruning ($r < 0.1$ for at least two nearest probes, pruning window = 3M base pairs). The probes were further removed from analyses if they did not present in the GoDMC/GS mQTL data or had < 5 independent genetic instruments overlapping with the outcome summary statistics. In result, 14 probes were selected for final analyses.

Results

Discovery MWAS of depression-PRS in GS

Methylome-wide association (MWAS) with depression-PRS at all p thresholds

There were 599 CpG probes significantly associated with depression-PRS with p-threshold (pT) at 5×10^{-8} ($p < 6.5 \times 10^{-8}$ to reach significance after Bonferroni correction). In contrast to many other studies that use polygenic risk profiling at different thresholds to predict depression⁵³, both the number of significant associations and the effect sizes decreased as PRSs were calculated at increasingly lenient thresholds (Figure 1). For pT of 1×10^{-6} , 1×10^{-4} and 0.001, 414, 57 and 17 CpGs were associated with depression-PRS, respectively (Supplementary Figures 1). No significant associations were found for PRS using p-value thresholds greater than or equal to 0.01. Quantile-quantile plot and statistics for genomic inflation factors (ranged from 0.954 to 1.011) can be found in Supplementary Figure 2 and Supplementary Table 4. Results using the depression-PRS calculated using only genome-wide significant variants are presented below.

DNAm association with depression-PRS at a GWAS p-value association threshold of 5×10^{-8}

The most significant associations of DNAm with depression-PRS were found in the major histocompatibility complex (*MHC*) region (25-35 Mb on Chromosome 6, Figure 2), with 575/599 (96.0%) of significant associations within this region ($p_{\text{Bonferroni}}$ ranged from 0.05 to 1.37×10^{-111}). The top ten probes that showed the greatest associations are listed in Supplementary Table 5 (all $p_{\text{Bonferroni}} < 2.78 \times 10^{-83}$). The majority of probes were in Open Sea (47.9%) and CpG shores (28.2%), and they were particularly enriched in the CpG shores compared to non-significant CpGs (18% for non-significant CpGs, $\chi^2=41.0$, $df=1$, $p=1.53 \times 10^{-10}$, see Supplementary Table 6 and Supplementary Information). After pruning ($r < 0.1$ for at least two nearest probes, window = 3Mb), four independent CpG probes were identified within the *MHC* region: cg03270340, cg14345882, cg10046620, cg08116408 (all $p_{\text{Bonferroni}} < 2.40 \times 10^{-58}$). UCSC gene database annotation shows genes that are nearest to the probes are *TRIM27*, *HIST1H2AI*, *TRIM40* and *BTN3A2*. See Table 1-2, Supplementary Table 5 and Supplementary Data 1.

Supplementary MWAS were conducted on two additional depression-PRSs to investigate the associations found within the *MHC* region. Two additional PRSs were calculated using (1) the independent genetic risk variants reported in the depression GWAS by Howard *et al.* with a wider pruning window of 3M base pairs and retaining only one variant in the *MHC* region and (2) SNPs located outside of *MHC* region, respectively. Analysis (1) was conducted to identify if the associations found in the *MHC* region was due to the additive effect from a large number of genetic variants included in the *MHC* region. Analysis (2) was conducted to find out if the associations found within the *MHC* region was contributed by genetic variants located *trans* to this region. The number of significant associations found within the *MHC* region for the PRS calculated using independent genetic risk variants reduced from 599 to 25, at the MDD-GWAS PRS p-value threshold of 5×10^{-8} . No CpGs within the *MHC* region were found to be significantly associated with the PRS generated from variants mapping without the *MHC* region. See Supplementary Figure 3.

Outside of the *MHC* region, 24 probes showed significant associations with depression-PRS

estimated across the genome at pT of 5×10^{-8} ($p_{\text{Bonferroni}}$ ranged from 0.048 to 4.74×10^{-70}). The top ten probes are listed in Supplementary Table 5. Genes mapping near to the top probes were associated with histone deacetylase, DNA binding and transcriptional processes (such as *MAD1L1*, *TCF4*, *RERE* and *ZSCAN31*), and neuronal plasticity and growth (for example, *NEGR1*).

The effect sizes for the significant CpG probes showed high correlations between the two data sets ($r=0.90$), and direction for all significant associations was consistent between sets. For these significant probes, the distance to the nearest depression risk locus was significantly lower than those that were not significant (significant versus not significant: standardised Cohen's $d=0.920$, $p < 1 \times 10^{-32}$). There were 21.2% of all significant CpGs located outside of 1Mb boundaries of genetic risk loci for depression and outside of the region consisted of SNPs in LD ($R^2 > 0.1$) with the genetic risk loci (see Supplementary Data 1).

Replication depression-PRS MWAS in LBC1921, LBC1936 and ALSPAC

MWAS of depression-PRS on pT of 5×10^{-8} on adult and adolescent samples (LBC1921, LBC1936 and ALSPAC)

We looked at a subset of CpG probes that were significant in the discovery MWAS analysis and found that the standardised effect sizes were correlated between the discovery and replication meta-MWAS of LBC1921, LBC1936 and ALSPAC adults, with ($N_{\text{probe}}=507$, $r=0.83$) or without the probes located in the MHC region ($N_{\text{probe}}=14$, $r=0.78$). There were 98.1% associations found in the discovery MWAS remained in the same direction and 71.8% remained significant after Bonferroni-correction within the replication analysis. See Figure 3.

Standardised effect sizes for the significant CpG probes found in the discovery MWAS were highly correlated with those in the MWAS on adolescents from ALSPAC (all CpG probes: $N_{\text{probe}}=506$, $r=0.76$; no MHC region: $N_{\text{probe}}=14$, $r=0.79$).

MWAS for depression-PRS on all p thresholds on adult samples

Meta-analysis of the methylome-wide association analysis of depression-PRS for replication cohorts (LBC1921, LBC1936 and ALSPAC adults) showed that, for depression-PRS of pT at 5×10^{-8} , 1×10^{-6} , 1×10^{-4} and 0.001, the number of significant CpG probes were 265, 203, 59 and 18, respectively. Similar to the discovery analysis, no significant associations were found for PRS of $pT \geq 0.01$.

A full list of results for replication analysis can be found in Supplementary Data 1 and Supplementary Figures 4-7.

Pathway enrichment analysis

Gene ontology (GO) terms and KEGG pathways were assessed for the genes associated with depression-PRS of pT at 5×10^{-8} . There were 119 enriched GO terms and 29 KEGG pathways significant after FDR correction (all $p_{\text{FDR}} < 0.044$). The majority of the significant GO terms

were associated with antigen processing and the immune response. Enriched KEGG pathways include antigen processing and immune disease, for example, allograft rejection, autoimmune thyroid disease and human T-cell leukemia virus 1 infection. The top ten GO terms and KEGG pathways are listed in Tables 1-2.

A supplementary analysis without probes in the MHC region showed no significant GO terms or KEGG pathways after FDR correction. Top GO terms that were nominally significant include neural development and metabolic process, for example, phosphagen and phosphocreatine metabolic process, cerebellum and cerebellar cortex maturation and cerebellar Purkinje cell layer maturation ($p_{\text{uncorrected}}$ ranged from 0.049 to 1.54×10^{-3}). No nominally significant KEGG pathway was found ($p_{\text{uncorrected}} > 0.063$).

SNP–CpG mapping for the depression risk loci

SNP–CpG probe associations were investigated by conducting MWAS for each of the independent genetic risk loci for depression. There were 6,009 CpG probes that showed significant associations with at least one leading genetic risk variant after Bonferroni correction.

There were 94 of the 96 genetic risk variants tested showed significant *cis* association with CpGs within 1 Mb distance (see Figure 4). There were 88 genetic risk variants (91.6% of all variants tested) that showed *trans* associations outside of their 1Mb window and 82 variants (85.4% of all variants tested) that had *trans* associations with CpGs located on at least one different chromosome.

Five genetic loci showed associations with methylation levels at CpGs located in more than half of the distal autosomal chromosomes (see Figure 4). Genes close to these genetic risk variants were involved in, for example, nucleic acid transcription activities, which includes nucleic acid binding (*ZNF179* and *ESR2*), mitotic assembly (*MAD1L1*) and encoding proteins that colocalise with transcription factors (*RERE*). Regional association plots showing genes within 1Mb distance from the five genetic variants can be found in Supplementary Figure 8.

Colocalisation analysis

We hypothesised that SNPs influencing MDD risk and those influencing DNAm would be shared. Colocalisation analysis, however, indicated that there was no strong evidence ($PP_4 > 0.8$) for shared genetic factors between MDD loci and DNAm. The posterior probability for one region was supportive of a convincing co-localised association signal for both MDD and DNAm in that region ($PP_4 = 0.710$). Within this region, three SNPs were indicated to have a posterior probability of being shared SNPs for the two traits > 0.1 : rs73163796=66%, rs73163779=17%, and rs11710605=11%. All three loci colocalised with genetic variation influencing a smoking-associated CpG probe, cg15099418⁵⁴. Supplementary Data 2 contains results for all 102 regions investigated in colocalisation analysis.

Mendelian randomisation (MR)

Discovery MR: causal effect of DNA methylation on depression using GoDMC data

Three probes showed significant causal effect on depression for all three MR methods with consistent direction of effects between methods: cg08344181 on chromosome3, cg17862947 on chromosome 1 and cg17841099 on chromosome 12 (absolute β_{IVW} ranged from 0.042 to 0.103, p_{FDR} ranged from 2.99×10^{-47} to 3.60×10^{-120} ; absolute $\beta_{MR-Egger}$ ranged from 0.052 to 0.122, p_{FDR} ranged from 0.045 to 1.23×10^{-6} , p_{FDR} for MR-Egger intercept ranged from 0.797 to 0.228; absolute β_{WM} ranged from 0.047 to 0.107, p_{FDR} ranged from 1.23×10^{-6} to 2.54×10^{-76} , p_{FDR} for Q-statistics ranged from 0.917 to 0.038).

Seven other probes: cg07519229, cg13813247, cg14159747, cg17925084, cg19624444, cg19800032 and cg23275840, showed significant causal effect using the IVW and WM methods (absolute β_{IVW} ranged from 0.019 to 0.046, p_{FDR} ranged from 4.2×10^{-3} to 1.10×10^{-8} ; β_{WM} ranged from 0.018 to 0.048, p_{FDR} ranged from 1.70×10^{-5} to 4.40×10^{-11} , p_{FDR} for Q-statistics ranged from 0.917 to 3.7×10^{-28}). Effect sizes were consistent for the above probes were consistent between the IVW and WM methods. No significant causal effect on depression was found using the MR-Egger method for these probes (absolute $\beta_{MR-Egger}$ ranged from 4.7×10^{-7} to 0.003, p_{FDR} ranged from 0.739 to 0.127). However, the direction of effects remained the same with the IVW and weighted-median methods and the MR-Egger intercept tests showed no evidence of horizontal pleiotropy p_{FDR} ranged from 0.797 to 0.163).

Results are also shown in Figure 5 and Supplementary Figure 9 and Supplementary Table 7.

Replication MR: causal effect of DNA methylation on depression using GS data

All of the causal effects of DNA methylation to depression found in the discovery MR analysis showed consistent direction in the replication analysis and across all three MR methods. For all three MR methods, the effect sizes were highly correlated between discovery and replication analyses (r ranged from 0.855 to 0.934). Eight out of ten significant effects found in the discovery MR analysis were significant for all three MR methods in the replication analyses (absolute β_{IVW} ranged from 0.081 to 1.022, $p_{IVW-FDR}$ ranged from 1.20×10^{-7} to 4.21×10^{-119} ; absolute β_{WM} ranged from 0.102 to 1.059, p_{WM-FDR} ranged from 1.82×10^{-7} to 2.38×10^{-58}). The two other probes showed significant causal effect at two MR methods (cg14159747 significant for MR-Egger and WM and cg19800032 significant for IVW and WM, see statistics in Supplementary Table 3). MR-Egger intercepts were not significantly deviated from 0 for all replication MR ($p_{FDR} > 0.68$), and thus showed no evidence of horizontal pleiotropy. See Supplementary Table 8.

Multi-variable MR: independent causal effect of DNA methylation on depression using GS data

We next tested for causal associations between DNAm at multiple CpGs from the discovery analysis to MDD. The significant probes were entered into the two-sample MR analysis simultaneously, to identify the set of independent SNPs that showed the strongest and independent casual associations with MDD using the IVW method. Four probes showed causal effects when all CpGs were considered simultaneously. They included cg23275840 on chromosome 4, cg19624444 on chromosome 7, cg17925084 on chromosome 14 and

cg17862947 on chromosome 12 (absolute β_{IVW} ranged from 0.077 to 1.023, p_{FDR} ranged from 0.040 to 1.97×10^{-6} , see Supplementary Figure 10 and Supplementary Table 9). Genes annotated with these CpG probes are *SLC39A1*, *CORIN*, *KLC1* and *MAD1L1*. These genes are involved in signalling receptor binding in the brain and hormonal regulation.

MR: causal effects of depression liability on DNA methylation

MR provided evidence of a causal effect of depression liability on DNA methylation for seven CpG probes. Other than the effect to cg09256413 was significant for all three MR methods, effects on other probes were significant for two MR methods (for significant causal effects: absolute β_{IVW} ranged from 0.057 to 0.172, $p_{IVW-FDR}$ ranged from 6.95×10^{-3} to 1.48×10^{-5} ; absolute β_{WMM} ranged from 0.0237 to 0.058, $p_{WMM-FDR}$ ranged from 0.046 to 8.24×10^{-3} , and absolute $\beta_{MR-Egger}$ ranged from 0.251 to 6.376, $p_{MR-Egger-FDR}$ ranged from 0.024 to 2.05×10^{-23}).

Among all the CpG probes that showed significant causal effects from depression to DNA methylation, only cg17862947 on chromosome 12 showed significant bidirectional effects between depression and DNA methylation ($\beta_{IVW}=-0.093$, $p_{FDR}=1.15 \times 10^{-4}$; $\beta_{WMM}=-0.008$, $p_{FDR}=0.506$; $\beta_{MR-Egger}=-0.813$, $p_{FDR}=0.045$). However, reversed-MR for this probe showed significant heterogeneity of instrumental effects (p_{FDR} for Q statistics $< 1 \times 10^{-16}$). None of the other CpG probes showed consistently significant bi-directional effect with depression in the discovery and replication MR. See Figure 6, Supplementary Figure 11 and Supplementary Table 10.

Discussion

Polygenic risk scores (PRS) for genetic risk variants of depression showed wide-spread associations with peripheral blood DNAm across the epigenome in Generation Scotland: Scottish Family Health Study Cohort (GS, N=8,898). The strongest DNAm associations showed highly consistent results in the replication analysis in adults (N=2,049, $r_{\beta} = 0.83$) and in adolescents (N=423, $r_{\beta} = 0.76$). Significant CpG probes are enriched in immunological processes in the human leukocyte antigen (HLA) system and neuronal maturation and development. Influence from the genetic risk of depression was both local (*cis*) and distal (*trans*). Five genetic risk loci showed wide-spread *trans* effect across half of the autosomal chromosomes. Finally, using Mendelian randomisation, we found evidence of a mutually causal effect of DNAm on liability to MDD at CpG probes associated with polygenic risk scores for depression.

The probes associated with genetic risk variants for depression map to genes including *TRIM27*, *BTN3A2* and *HIST1H2AI*. These HLA-related genes have been widely found associated with psychiatric conditions such as schizophrenia and bipolar disorder⁶. Other genes that located outside of the MHC region, such as *BDNF*, *RERE* and *ZSCAN31*, are associated with neuronal development and guidance of neuronal growth⁵⁵, transcriptional processes^{56,57}, as well as other risk factors for depression, for example, obesity, smoking and abnormal physical development⁵⁸.

A large proportion of identified CpG probes were localised in open sea areas and CpG island shores. Previous studies found that probes showed greater individual variations within the CpG island shores⁵⁹, along with higher heritability and greater number of *trans* effect from SNPs¹³. Methylation at these classes of CpGs may be subject to genetic effects to a greater extent than those that map outside such regions. The genes related to the depression genetic risk variants in CpG island shores have previously been linked to DNA, RNA and chromatin binding⁶⁰. Differential methylation in these regions may reflect a regulatory role for genetic risk variants through influencing DNA methylation in regions adjacent to CpG islands⁵⁹.

Mendelian randomisation provided evidence for a causal effect of methylation levels at CpG probes associated with lead SNPs on depression. After controlling for functional pleiotropy shared between CpG probes, four probes showed an independent causal effect on depression. Genes annotated with these independent probes are associated with phenotypes such as lower total brain volume⁶¹, excessive C-reactive protein⁶², obesity⁶³ and adverse lifestyle factors such as smoking⁶⁴, which are implicated in both patients with depression and those who have been exposed to early environmental risks, such as childhood trauma⁶⁵. These phenotypes have also been shown to have causal effects on depression in previous MR studies^{4,5}.

Although evidence was found for bidirectional causal effects between DNA methylation and depression, statistical evidence was stronger for the causal effect from DNA methylation to depression compared to the opposite direction, despite that more genetic variants were used for the reversed causal effect, and thus statistical power was greater (N_{SNP} for DNAm to depression ranged from 5 to 18 and N_{SNP} for depression to DNAm was 112). The highly consistent methylome-wide associations found across adults and adolescents may indicate that early genetic influence on DNAm result in a predominantly directional effect from DNAm to depression⁶⁶.

The present paper utilises large samples with replication analyses yielding highly consistent results. One limitation for interpreting the current findings is that DNA methylation data was collected from blood samples, that may not reflect the most relevant cell types in depression. Nevertheless, studies have shown that the genetic drivers of DNAm have similar effects across multiple cell types^{13,67}. The greater accessibility of DNAm from whole blood also has clear sample size and other methodological advantages compared to measures obtained from neural tissue post-mortem, and it is more likely that these measures could be used in future clinical applications. Future studies could further expand the scope by including other cell and tissue types.

In the current study we demonstrate that genome-wide genetic risk variants for depression show wide-spread epigenome-wide DNAm associations both individually and when combined in a risk score. These changes implicate antigen processing and immune system responses and may provide clues to the underlying mechanisms of depression.

Acknowledgments

Generation Scotland received is currently supported by the Wellcome Trust Investigator Award in Science 01/06/2021 to 31/5/26 'Exploiting genomic approaches to identify the environmental basis of depression'. (Reference: 220857/Z/20/Z) to McIntosh AM (PI). The DNA methylation profiling and data preparation was supported by Wellcome Investigator Award 220857/Z/20/Z and Grant 104036/Z/14/Z (PI for both grants is Prof McIntosh AM) and through funding from NARSAD (Ref 27404, Dr Howard DM) and the Royal College of Physicians of Edinburgh (SIM Fellowship, Dr Whalley HC). Genotyping of the GS:SFHS samples was funded by the MRC and Wellcome Trust [104036/Z/14/Z]. Generation Scotland also receives support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006].

We thank the LBC1936 participants and team members who contributed to the LBC1921 and LBC1936 studies. Further study information can be found at <https://www.ed.ac.uk/lothian-birth-cohorts>. The LBC1921 was supported by the United Kingdom's Biotechnology and Biological Sciences Research Council (BBSRC), The Royal Society and The Chief Scientist Office of the Scottish Government. The LBC1936 is supported by Age UK (Disconnected Mind project, which supports Dr Harris SE), the Medical Research Council (G0701120, G1001245, MR/M013111/1, MR/R024065/1), and the University of Edinburgh. Methylation typing of LBC1921 and LBC1936 was supported by Centre for Cognitive Ageing and Cognitive Epidemiology (Pilot Fund award), Age UK, The Wellcome Trust Institutional Strategic Support Fund, The University of Edinburgh, and The University of Queensland. Dr Cox SR and Prof Deary IJ were supported by a National Institutes of Health (NIH) research grant R01AG054628, and Dr Cox SR is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (Grant Number 221890/Z/20/Z).

The UK Medical Research Council and Wellcome (Grant Ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. A comprehensive list of grants funding is available on the ALSPAC website (<http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf>). GWAS data was generated by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. We are grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. Part of this data was collected using REDCap, see the REDCap website for details <https://projectredcap.org/resources/citations/>.

References

1. Marcus, M., Yasamy, M. T., Ommeren, M. van, Chisholm, D. & Saxena, S. *Depression: A Global Public Health Concern*. (2012).
2. Sullivan, P. F., Neale, M. C. & Kendler, K. S. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry* **157**, 1552–1562 (2000).
3. Sullivan, P. F., Daly, M. J. & O'Donovan, M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat. Rev. Genet.* **13**, 537–551 (2012).
4. Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).
5. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
6. International, T. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **10**, 748–752 (2009).
7. Shen, X. *et al.* A phenome-wide association and Mendelian Randomisation study of polygenic risk for depression in UK Biobank. *Nat. Commun.* **11**, 1–16 (2020).
8. Barbu, M. C. *et al.* Epigenetic prediction of major depressive disorder. *Mol. Psychiatry* 1–12 (2020). doi:10.1038/s41380-020-0808-3
9. Clark, S. L. *et al.* A methylation study of long-term depression risk. *Mol. Psychiatry* **25**, 1334–1343 (2020).
10. Zeng, Y. *et al.* Parent of origin genetic effects on methylation in humans are common and influence complex trait variation. *Nat. Commun.* **10**, 1–13 (2019).
11. Van Dongen, J. *et al.* Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat. Commun.* **7**, (2016).
12. Van Dongen, J. *et al.* Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat. Commun.* **7**, 1–13 (2016).
13. Min, J. L. *et al.* Genomic and phenomic insights from an atlas of genetic effects on DNA methylation. *medRxiv* **25**, 81 (2020).
14. Hannon, E. *et al.* Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nat. Neurosci.* **19**, 48–54 (2015).
15. Hannon, E. *et al.* An integrated genetic-epigenetic analysis of schizophrenia: Evidence for co-localization of genetic associations and differential DNA methylation. *Genome Biol.* **17**, 176 (2016).
16. Hannon, E. *et al.* Elevated polygenic burden for autism is associated with differential DNA methylation at birth. *Genome Med.* **10**, 1–13 (2018).
17. Sullivan, P. F. & Posthuma, D. Biological pathways and networks implicated in psychiatric disorders. *Curr. Opin. Behav. Sci.* **2**, 58–68 (2015).
18. Sullivan, P. F. & Geschwind, D. H. Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. *Cell* **177**, 162–183 (2019).

19. Smith, B. H. *et al.* Generation Scotland: The Scottish Family Health Study; a new resource for researching genes and heritability. *BMC Med. Genet.* **7**, 74 (2006).
20. Smith, B. H. *et al.* Cohort profile: Generation scotland: Scottish family health study (GS: SFHS). The study, its participants and their potential for genetic research on health and illness. *Int. J. Epidemiol.* **42**, 689–700 (2013).
21. Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J. & Fox, H. C. The Impact of Childhood Intelligence on Later Life: Following Up the Scottish Mental Surveys of 1932 and 1947. *Journal of Personality and Social Psychology* **86**, 130–147 (2004).
22. Deary, I. J., Whalley, L. J. & Starr, J. M. *A lifetime of intelligence: Follow-up studies of the Scottish mental surveys of 1932 and 1947. A lifetime of intelligence: Follow-up studies of the Scottish mental surveys of 1932 and 1947.* (2009). doi:10.1037/11857-000
23. Cox, S. R. *et al.* Brain white matter integrity and cortisol in older men: The Lothian Birth Cohort 1936. *Neurobiol. Aging* **36**, 257–264 (2015).
24. Hillary, R. *et al.* Integrative omics approach to identify the molecular architecture of inflammatory protein levels in healthy older adults. 1–15 (2020). doi:10.1101/2020.02.17.952135
25. Fraser, A. *et al.* Cohort profile: The avon longitudinal study of parents and children: ALSPAC mothers cohort. *Int. J. Epidemiol.* **42**, 97–110 (2013).
26. Boyd, A. *et al.* Cohort profile: The 'Children of the 90s'-The index offspring of the avon longitudinal study of parents and children. *Int. J. Epidemiol.* **42**, 111–127 (2013).
27. Northstone, K. *et al.* The Avon Longitudinal Study of Parents and Children (ALSPAC): an update on the enrolled sample of index children in 2019 [version 1; peer review: 2 approved]. *Wellcome Open Res.* **4**, (2019).
28. Relton, C. L. *et al.* Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES). *Int. J. Epidemiol.* **44**, 1181–1190 (2015).
29. Davies, G. *et al.* Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Mol. Psychiatry* **16**, 996–1005 (2011).
30. Kwong, A. S. F. *et al.* Polygenic risk for depression, anxiety and neuroticism are associated with the severity and rate of change in depressive symptoms across adolescence. *J. Child Psychol. Psychiatry Allied Discip.* (2021). doi:10.1111/jcpp.13422
31. Gunderson, K. L. Whole-genome genotyping on bead arrays. *Methods Mol. Biol.* **529**, 197–213 (2009).
32. Choi, S. W., Mak, T. S. H. & O'Reilly, P. F. Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols* **15**, 2759–2772 (2020).
33. Howard, D. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).
34. Fortin, J. P., Fertig, E. & Hansen, K. shinyMethyl: Interactive quality control of Illumina 450k DNA methylation arrays in R. *F1000Research* **3**, 1–11 (2014).
35. Pidsley, R. *et al.* A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* **14**, 293 (2013).

36. Min, J. L., Hemani, G., Smith, G. D., Relton, C. & Suderman, M. Meffil: Efficient normalization and analysis of very large DNA methylation datasets. *Bioinformatics* **34**, 3983–3989 (2018).
37. Walker, R. M. *et al.* Assessment of dried blood spots for DNA methylation profiling [version 1; peer review: 2 approved]. *Wellcome Open Res.* **4**, 44 (2019).
38. Jaffe, A. E. *et al.* Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nat. Neurosci.* **19**, 40–47 (2015).
39. Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11**, 587 (2010).
40. Yang, J. *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* **43**, 519–525 (2011).
41. Zhang, Q. *et al.* Genotype effects contribute to variation in longitudinal methylome patterns in older people. *Genome Med.* **10**, 75 (2018).
42. Shah, S. *et al.* Genetic and environmental exposures constrain epigenetic drift over the human life course. *Genome Res.* **24**, 1725–1733 (2014).
43. McCartney, D. L. *et al.* Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genomics Data* **9**, 22–24 (2016).
44. Min, J. L., Hemani, G., Davey Smith, G., Relton, C. & Suderman, M. Meffil: efficient normalization and analysis of very large DNA methylation datasets. *Bioinformatics* **34**, 3983–3989 (2018).
45. Willer, C. J., Li, Y., Abecasis, G. R. & Overall, P. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
46. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
47. Hansen, K.D. IlluminaHumanMethylationEPICanno.ilm10b4.hg19: Annotation for Illumina’s EPIC methylation arrays. *R package version 0.6.0* (2017). Available at: https://bitbucket.com/kasperdanielhansen/Illumina_EPIC. (Accessed: 18th June 2021)
48. Phipson, B., Maksimovic, J. & Oshlack, A. MissMethyl: An R package for analyzing data from Illumina’s HumanMethylation450 platform. *Bioinformatics* **32**, 286–288 (2016).
49. Wallace, C. *et al.* Statistical colocalization of monocyte gene expression and genetic risk variants for type 1 diabetes. *Hum. Mol. Genet.* **21**, 2815–2824 (2012).
50. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* **13**, e1007081 (2017).
51. Hemani, G. *et al.* The MR-base platform supports systematic causal inference across the human phenome. *Elife* **7**, (2018).
52. Zhang, F. *et al.* OSCA: A tool for omic-data-based complex trait analysis. *Genome Biol.* **20**, 107 (2019).
53. Bigdeli, T. B. *et al.* Genetic effects influencing risk for major depressive disorder in

- China and Europe. *Transl. Psychiatry* **7**, 61 (2017).
54. Joehanes, R. *et al.* Epigenetic Signatures of Cigarette Smoking. *Circ. Cardiovasc. Genet.* **9**, 436–447 (2016).
 55. Sandrini, L. *et al.* Association between obesity and circulating brain-derived neurotrophic factor (BDNF) levels: Systematic review of literature and meta-analysis. *International Journal of Molecular Sciences* **19**, (2018).
 56. Fagerberg, L. *et al.* Analysis of the Human Tissue-specific Expression by Genome-wide Integration of Transcriptomics and Antibody-based Proteomics. *Mol. Cell. Proteomics* **13**, 397–406 (2014).
 57. Plaster, N., Sonntag, C., Schilling, T. F. & Hammerschmidt, M. REREa/atrophin-2 interacts with histone deacetylase and Fgf8 signaling to regulate multiple processes of zebrafish development. *Dev. Dyn.* **236**, 1891–1904 (2007).
 58. Mansur, R. B. *et al.* BDNF and BMI effects on brain structures of bipolar offspring: results from the global mood and brain science initiative. *Acta Psychiatr. Scand.* **136**, 607–614 (2017).
 59. Visone, R. *et al.* DNA methylation of shelf, shore and open sea CpG positions distinguish high microsatellite instability from low or stable microsatellite status colon cancer stem cells. *Epigenomics* **11**, 587–604 (2019).
 60. Hendrickson, D., Kelley, D. R., Tenen, D., Bernstein, B. & Rinn, J. L. Widespread RNA binding by chromatin-associated proteins. *Genome Biol.* **17**, 28 (2016).
 61. Xu, Y., Xiao, G., Liu, L. & Lang, M. Zinc transporters in Alzheimer's disease. *Molecular Brain* **12**, (2019).
 62. Noh, H., Paik, H. Y., Kim, J. & Chung, J. The alteration of zinc transporter gene expression is associated with inflammatory markers in obese women. *Biol. Trace Elem. Res.* **158**, 1–8 (2014).
 63. Starnawska, A. *et al.* Epigenome-wide association study of depression symptomatology in elderly monozygotic twins. *Transl. Psychiatry* **9**, 214–214 (2019).
 64. Zhong, R. *et al.* MAD1L1 Arg558His and MAD2L1 Leu84Met interaction with smoking increase the risk of colorectal cancer. *Sci. Rep.* **5**, 1–9 (2015).
 65. Pariante, C. M. & Lightman, S. L. The HPA axis in major depression: classical theories and new developments. *Trends Neurosci.* **31**, 464–468 (2008).
 66. Mooney, M. A. *et al.* Large epigenome-wide association study of childhood ADHD identifies peripheral DNA methylation associated with disease and polygenic risk burden. *Transl. Psychiatry* **10**, 1–12 (2020).
 67. Braun, P. R. *et al.* Genome-wide DNA methylation comparison between live human brain and peripheral tissues within individuals. *Transl. Psychiatry* **9**, 1–10 (2019).

Figure 1. Number of CpG probes associated with polygenic risk scores (PRS) at nine different p-thresholds (pTs) for discovery analysis. X-axis represents the pTs used for generating PRS. Y-axis shows the number of probes significant associated with the given PRS. The four different lines represent four types of methods to define significance.

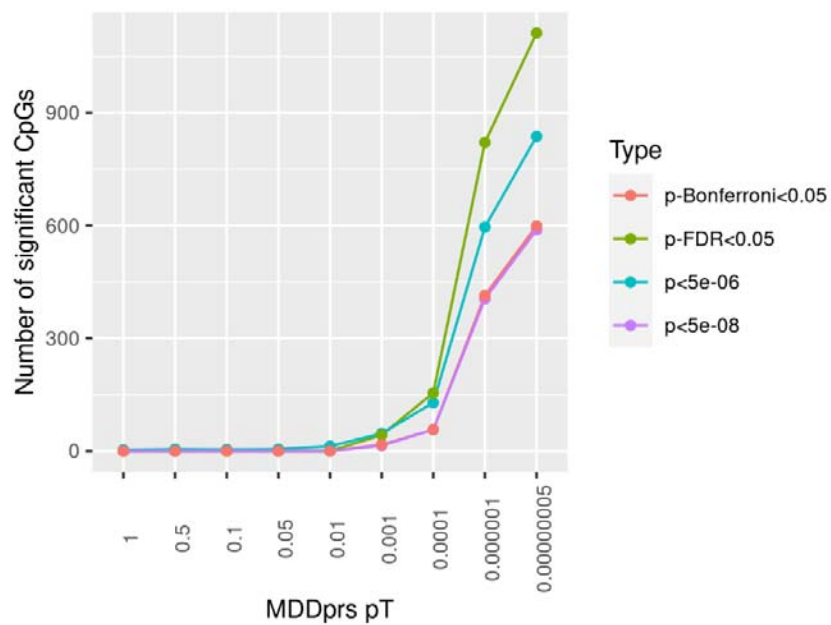


Figure 2. Manhattan plot for the discovery methylome-wide association study (MWAS) for PRS of pT at 5×10^{-8} in Generation Scotland (GS). Each dot represents a CpG probe. X-axis represents the relative position of the probes in the genome. Y-axis represents $-\log_{10}$ -transformed p values. The red dashed line represents the significance threshold for Bonferroni correction.

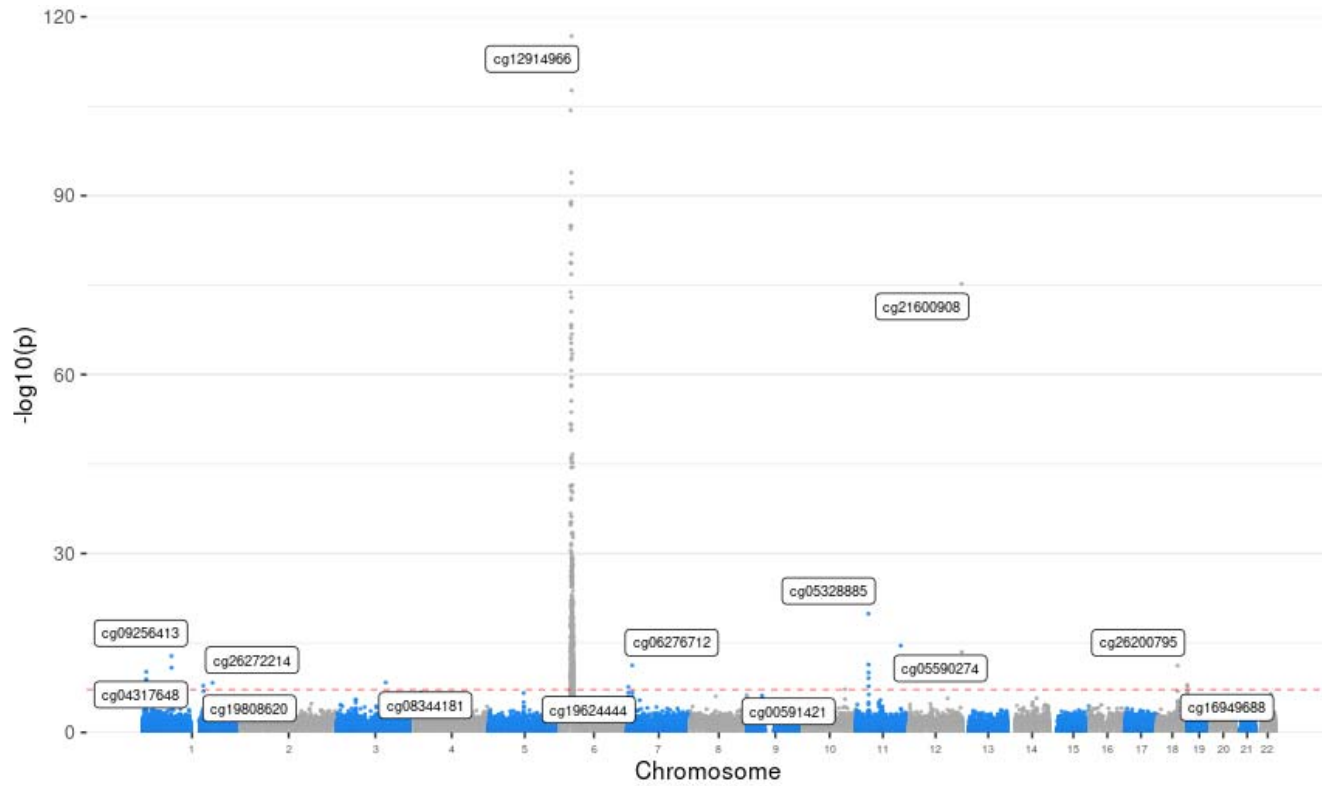


Figure 3. Replication MWAS in Lothian Birth Cohort (LBC) 1921, LBC1936 and Avon Longitudinal Study of Parents and Children (ALSPAC) adults. (a) Manhattan plot for the replication MWAS for PRS of pT at 5×10^{-8} . Each dot represents a CpG probe. X-axis represents the relative position of the probes in the genome. Y-axis represents $-\log_{10}$ -transformed p values. The red dashed line represents the significance threshold for Bonferroni correction. (b) Number of CpG probes significant associated with polygenic risk scores (PRS) at nine different pTs for replication analysis. X-axis represents the pTs used for generating PRS. Y-axis shows the number of probes significant associated with the given PRS. The four different lines represent four types of methods to define significance. (c) Scatter plot showing the correlation of standardised regression coefficients between the discovery (GS) and replication (LCB1921+LBC1936+ALSPAC adults) analysis. Each dot represents a CpG probe. Probes shown in the figure are those associated with depression-PRS of pT at 5×10^{-8} in the discovery MWAS (in GS). Dots in green represent probes locate in the major histocompatibility complex (MHC) regions and those in red represent other probes that locate outside of the MHC region.

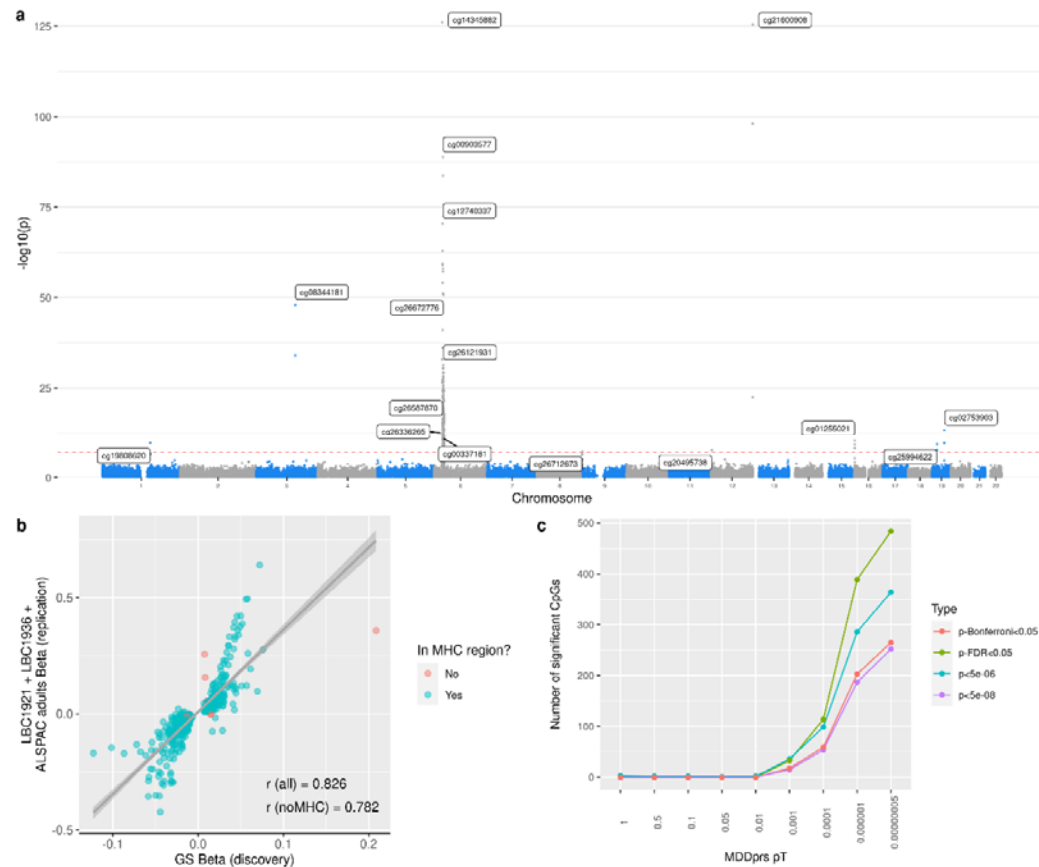


Figure 4. Heatmap showing the SNP-CpG mapping. Each row and column represent a CpG probe and depression risk locus, respectively. Those tests that are not significant after Bonferroni-correction are left blank. For those significant associations, a darker cell represents a higher $-\log_{10}$ -transformed p-value. All CpG probes and depression risk loci were categorised based on which chromosome (CHR) they locate. Within each chromosome, probes and SNPs are aligned from left to right or from bottom to top based on their genomic position.

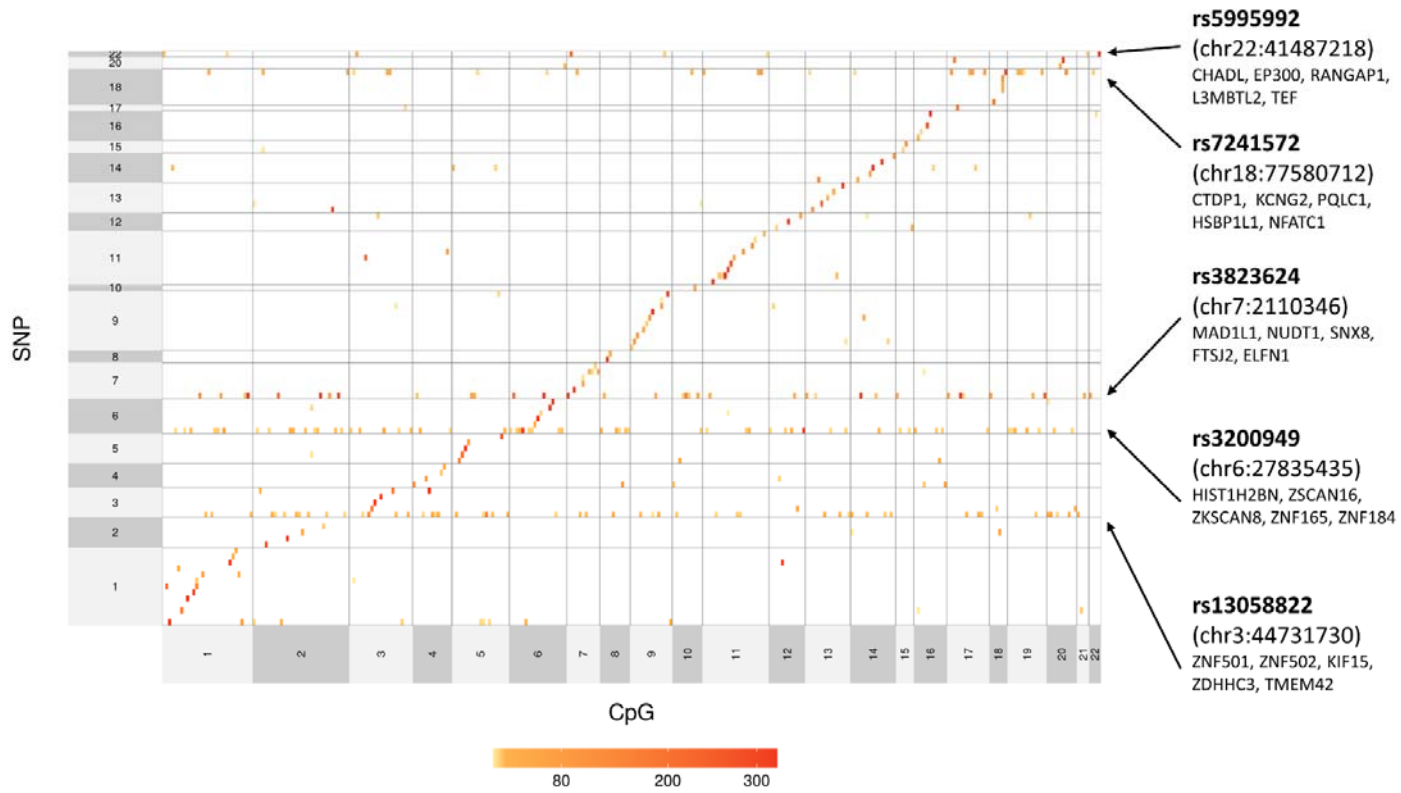


Figure 5. Mendelian randomisation (MR) analysis on DNA methylation and depression using data from the Genetics of DNA Methylation Consortium (GoDMC). (a) Discovery MR testing causal effect of DNA methylation to depression using GoDMC data. (b) Replication MR testing effect of DNA methylation using GS data to depression. (c) MR of reversed directionality testing the causal effect from depression to DNA methylation. X-axes represent p-values for MR analyses. Y-axes represent the individual tests conducted.

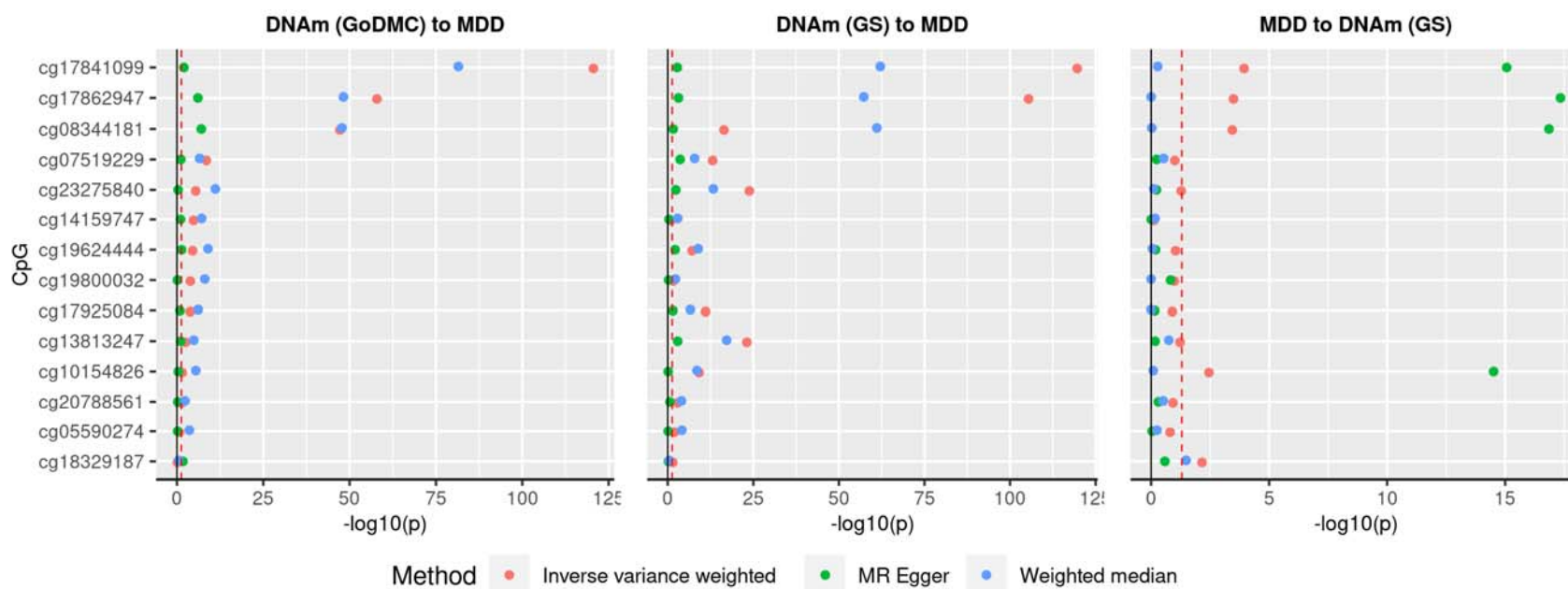


Table 1. Results for gene ontology (GO) analysis for the MWAS on PRS at $pT 5 \times 10^{-8}$. Analyses were conducted separately for including and excluding the MHC region. Top ten GO terms are listed in the table. BP=biological process, CC=cellular component and MF=molecular function.

	GO ID	Ontology	Term	N	DE	P _{DE}	P _{FDR}
With MHC region	GO:0002428	BP	antigen processing and presentation of peptide antigen via MHC class Ib	9	7	<1E-322	<1E-322
	GO:0002476	BP	antigen processing and presentation of endogenous peptide antigen via MHC class Ib	8	7	<1E-322	<1E-322
	GO:0002483	BP	antigen processing and presentation of endogenous peptide antigen	15	8	<1E-322	<1E-322
	GO:0019883	BP	antigen processing and presentation of endogenous antigen	22	8	<1E-322	<1E-322
	GO:0019885	BP	antigen processing and presentation of endogenous peptide antigen via MHC class I	15	8	<1E-322	<1E-322
	GO:0042613	CC	MHC class II protein complex	14	8	<1E-322	<1E-322
	GO:0042611	CC	MHC protein complex	23	14	1.66E-24	5.38E-21
	GO:0042605	MF	peptide antigen binding	23	11	1.97E-17	5.61E-14
	GO:0071556	CC	integral component of luminal side of endoplasmic reticulum membrane	26	11	3.92E-17	8.91E-14
	GO:0098553	CC	luminal side of endoplasmic reticulum membrane	26	11	3.92E-17	8.91E-14
Without MHC region	GO:0072686	CC	mitotic spindle	101	2	0.002	1
	GO:1902412	BP	regulation of mitotic cytokinesis	6	1	0.002	1
	GO:0021691	BP	cerebellar Purkinje cell layer maturation	2	1	0.002	1
	GO:0021590	BP	cerebellum maturation	3	1	0.002	1
	GO:0021699	BP	cerebellar cortex maturation	3	1	0.002	1
	GO:0045666	BP	positive regulation of neuron differentiation	353	3	0.003	1
	GO:0021578	BP	hindbrain maturation	6	1	0.004	1
	GO:0021626	BP	central nervous system maturation	7	1	0.004	1
	GO:0001093	MF	TFIIIB-class transcription factor binding	3	1	0.004	1
	GO:0070369	CC	beta-catenin-TCF7L2 complex	3	1	0.004	1

Table 2. Results for pathway analysis for the MWAS on PRS at $pT 5 \times 10^{-8}$. Analyses were conducted separately for including and excluding the MHC region. Top ten KEGG pathways are listed in the table.

	KEGG pathway ID	Description	N	DE	P	P _{FDR}
With MHC region	path:hsa04940	Type I diabetes mellitus	41	13	6.03E-19	2.01E-16
	path:hsa05330	Allograft rejection	34	12	1.17E-18	2.01E-16
	path:hsa05332	Graft-versus-host disease	37	12	2.39E-18	2.74E-16
	path:hsa04612	Antigen processing and presentation	66	14	3.82E-18	3.29E-16
	path:hsa05320	Autoimmune thyroid disease	46	12	2.61E-17	1.79E-15
	path:hsa05416	Viral myocarditis	55	12	8.53E-15	4.38E-13
	path:hsa04145	Phagosome	140	15	8.91E-15	4.38E-13
	path:hsa05150	Staphylococcus aureus infection	82	10	6.89E-12	2.96E-10
	path:hsa05322	Systemic lupus erythematosus	112	11.5	8.87E-12	3.39E-10
	path:hsa04514	Cell adhesion molecules	136	13	2.21E-11	7.59E-10
Without MHC region	path:hsa04110	Cell cycle	121	1	0.067	1
	path:hsa04114	Oocyte meiosis	114	1	0.070	1
	path:hsa04514	Cell adhesion molecules	118	1	0.083	1
	path:hsa04914	Progesterone-mediated oocyte maturation	86	1	0.063	1
	path:hsa05166	Human T-cell leukemia virus 1 infection	190	1	0.130	1
	path:hsa05203	Viral carcinogenesis	152	1	0.106	1
	path:hsa00010	Glycolysis / Gluconeogenesis	64	0	1	1
	path:hsa00020	Citrate cycle (TCA cycle)	28	0	1	1
	path:hsa00030	Pentose phosphate pathway	25	0	1	1
	path:hsa00040	Pentose and glucuronate interconversions	32	0	1	1