

RESEARCH

Open Access



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

Xueyi Shen^{1*} , Doretta Caramaschi^{2,3}, Mark J. Adams¹, Rosie M. Walker⁴, Josine L. Min³, Alex Kwong^{1,3}, Gibran Hemani³, Genetics of DNA Methylation Consortium³, Miruna C. Barbu¹, Heather C. Whalley¹, Sarah E. Harris⁵, Ian J. Deary⁵, Stewart W. Morris⁶, Simon R. Cox⁵, Caroline L. Relton³, Riccardo E. Marioni⁶, Kathryn L. Evans⁶ and Andrew M. McIntosh^{1*}

Abstract

Background: Depression is a disabling and highly prevalent condition where genetic and epigenetic, such as DNA methylation (DNAm), differences contribute to disease risk. DNA methylation is influenced by genetic variation but the association between polygenic risk of depression and DNA methylation is unknown.

Methods: We investigated the association between polygenic risk scores (PRS) for depression and DNAm by conducting a methylome-wide association study (MWAS) in Generation Scotland ($N = 8898$, mean age = 49.8 years) with replication in the Lothian Birth Cohorts of 1921 and 1936 and adults in the Avon Longitudinal Study of Parents and Children (ALSPAC) ($N_{\text{combined}} = 2049$, 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). Gene ontology analysis was conducted for the cytosine-guanine dinucleotide (CpG) probes significantly associated with depression PRS, followed by Mendelian randomisation (MR) analysis to infer the causal relationship between depression and DNAm.

Results: Widespread associations ($N_{\text{CpG}} = 71$, $p_{\text{Bonferroni}} < 0.05$, $p < 6.3 \times 10^{-8}$) were found between PRS constructed using genetic risk variants for depression and DNAm in CpG probes that localised to genes involved in immune responses and neural development. The effect sizes for the significant associations were highly correlated between the discovery and replication samples in adults ($r = 0.79$) and in adolescents ($r = 0.82$). Gene Ontology analysis showed that significant CpG probes are enriched in immunological processes in the human leukocyte antigen system. Additional MWAS was conducted for each lead genetic risk variant. Over 47.9% of the independent genetic risk variants included in the PRS showed associations with DNAm in CpG probes located in both the same (*cis*) and distal (*trans*) locations to the genetic loci ($p_{\text{Bonferroni}} < 0.045$). Subsequent MR analysis showed that there are a greater number of causal effects found from DNAm to depression than vice versa (DNAm to depression: p_{FDR} ranged from 0.024 to 7.45×10^{-30} ; depression to DNAm: p_{FDR} ranged from 0.028 to 0.003).

Conclusions: PRS for depression, especially those constructed from genome-wide significant genetic risk variants, showed methylome-wide differences associated with immune responses. Findings from MR analysis provided evidence for causal effect of DNAm to depression.

*Correspondence: xueyi.shen@ed.ac.uk; andrew.mcintosh@ed.ac.uk

¹ Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Morningside Park, Edinburgh EH10 5HF, UK

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Keywords: DNA methylation, Polygenic risk score, Depression, Methylome-wide association study, Replication, Mendelian randomisation

Background

Depression is a highly prevalent condition and a leading cause of global disability [1], 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 risk [6]. PRS is the sum of risk alleles, weighted by the effect sizes of an independent GWAS [7]. It 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 disorder [8].

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 depression [9]. DNAm risk scores have been developed from the results of DNA methylome-wide association studies (MWAS; also widely referred to as epigenome-wide association studies or EWAS in contemporary literature) [9]. These can be used to predict prevalent depression in independent samples, and chronic depression that requires long-term treatment [10]. DNAm is influenced by both genetic and environmental factors [9, 11] and, in blood tissue, it has a mean heritability of 19% across the methylome [12] with ~7% of its variance captured by common genetic variants [12]. For the highly heritable DNAm probes, genetic effects are consistent across tissues [13] and developmental stages [12]. Genetic risk variants for diseases (e.g. schizophrenia) have been found enriched in DNAm variation [14–16]. 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. MWAS were conducted on four

cohorts: Generation Scotland: Scottish Family Health Study (GS, discovery sample, $N = 8898$) [19, 20], the Lothian Birth Cohort (LBC1921) [21, 22], the Lothian Birth Cohort 1936 (LBC1936) [21, 22], Avon Longitudinal Study of Parents and Children (ALSPAC) adults (adult replication sample, combined $N = 2049$) and ALSPAC children for replication (adolescent replication sample, $N = 423$) [23, 24]. Mendelian randomisation (MR) 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

Generation Scotland: Scottish Family Health Study (GS)

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. DNAm data and genetic data were both collected, processed and quality-checked for 8898 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 Cohort (LBC)1921 and LBC1936

Participants from LBC1921 and LBC1936 [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 1330 participants from both cohorts combined with genetic and DNAm data (LBC1921: mean age = 79.1 years, SD of age = 0.6, 39.7% were men; LBC1936: 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) [25, 26].

Avon Longitudinal Study of Parents and Children (ALSPAC)

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 [23, 24]. The cohort consists of 13,761 mothers and their partners, and their 14,901 children, now young adults [27]. 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 DNAm data (mean age = 47.2 years, SD of age = 4.6) were included in the replication study [28]. 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 al. [28]. 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 GS [19], LBC1921, LBC1936 [29] and ALSPAC [30] has been previously published and is described briefly below. Analyses were conducted on European participants.

GS

Each sample was genotyped using the Illumina HumanOmniExpressExome-8v1.0 BeadChip (48.8%) or Illumina HumanOmniExpressExome-8 v1.2 BeadChip (51.2%) with Infinium chemistry [31]. 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 an information metric [32] (INFO score) <0.8 were removed from the 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 MAF <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} and 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

PRS of depression were calculated using PRSice-2 [7] 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. The summary statistics are available at the URL: <https://datashare.ed.ac.uk/handle/10283/3203> [33]. 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 $r^2 > 0.1$) reported in the GWAS by Howard et al. [33] for supplementary analysis. Details of the PRS profiling procedures and validation in the GS can be found elsewhere [33] (also see Additional file 1: Supplementary methods and Additional file 1: Tables S1-S2).

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 ($p < 5 \times 10^{-8}$) in linkage disequilibrium (LD $R^2 > 0.01$) with the lead variants reported in the Howard et al. study [4]. A total of 96 SNPs were available and thus selected as leading risk variants for further analysis.

DNAm data

GS

Genome-wide DNAm 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 elsewhere [37]. 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 $\geq 5\%$ of the total probe sample, or a detection p -value >0.05 for set 1 and p -value >0.01 for set 2 in $\geq 0.5\%$ of the total sample in each respective set. Cross-hybridising probes that map to genetic variants at $MAF >0.05$ and polymorphic probes were removed [38]. 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 intensity [39]. The M -values were corrected using a linear-mixed model, controlling for relatedness using the GCTA-estimated genetic relationship matrix [40] 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 769,526 autosomal CpG probes were then used for further analysis.

LBC1921 and LBC1936

Genome-wide DNAm data was obtained from blood sample using the HumanMethylation450K array (https://emea.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_human_methylation450.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 removed [43]. 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 409,319 CpG probes were retained for further analysis.

ALSPAC

Illumina Infinium HumanMethylation450 Beadchip arrays were used for measuring genome-wide DNAm data from peripheral blood samples [28]. The R package ‘meffil’ (version 1.1.1) was used for pre-processing, quality control and normalisation as previously reported [36]. 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 to 75% or identified as cross-hybridising or polymorphic [43]. Related (IBD >0.1) participants were not included in the analyses [30].

M -value transformation was conducted. In total, 481,600 and 449,595 CpG probes remained for analysis on ALSPAC adults and children, respectively.

Statistical models for 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) [44]. We used the default analysis scheme without genomic control correction (genomic inflation factors reported in the Additional file 1: Supplementary methods). P -values for the meta-analysis were obtained from a fixed-effect inverse-variance model. A sensitivity analysis was conducted on the unrelated participants in GS (methods and results reported in Additional file 1: Supplementary methods). 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’ [45] (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’ option [36].

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, the first 10 principal components derived from the imputed genetic data, age and sex were included as covariates for the discovery methylome-wide association analysis (see Additional file 1: Supplementary Methods and Additional file 1: Table S3 for details). Where possible, the same covariates were used in the replication analyses, although only smoking status (and not pack years) was available in LBC 1921, LBC 1936 and ALSPAC. Details for all the covariates included in the replication analysis can be found in the Additional file 1: Supplementary methods. 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 'Illumina-HumanMethylationEPICanno.ilm10b4.hg19' R package (version 3.13) [46].

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 missMethyl [47]. Default settings were used, which include correction for the number of probes per gene. CpGs that showed significant association with depression-PRS at $p_T < 5 \times 10^{-8}$ 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. Analyses were conducted on the Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways separately, by specifying the 'collection' parameter as 'GO' and 'KEGG', respectively. FDR correction was applied for all analyses.

Colocalisation analysis

We used Howard et al.'s depression GWAS [4] for depression-associated SNPs and GoDMC summary statistics for methylation quantitative trait loci (mQTL) analysis with PGC studies and GS study removed, which resulted in 32 remaining studies imputed to the 1000 Genome reference panel [13]. We used the package 'gwasglue' (version 0.0.0.9000, <https://mrcieu.github.io/gwasglue/>) [48] to extract SNPs that were ± 1 Mb of each of the 102 genome-wide significant, lead SNPs identified in Howard et al. [33] and then extracted the same SNPs within those regions from the GoDMC mQTL analysis. We used the colocalisation analysis 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.

MR

Three MR methods, inverse-variance weighted (IVW), weighted median (WM) and MR-Egger, were used to test for causal effects between DNAm 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 ancestry [13]. 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 elsewhere [13]. mQTL summary statistics from GS ($N = 8898$) 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://cnsgenomics.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 Additional file 1: Supplementary methods.

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 elsewhere [33].

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 participants [33]. See Additional file 1: Supplementary methods for details.

First, we used mQTL summary statistics from GoDMC to identify causal effects from DNAm to depression. Second, we used full mQTL summary statistics from GS to replicate the findings and to test the causal effect to DNAm 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 depression, 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

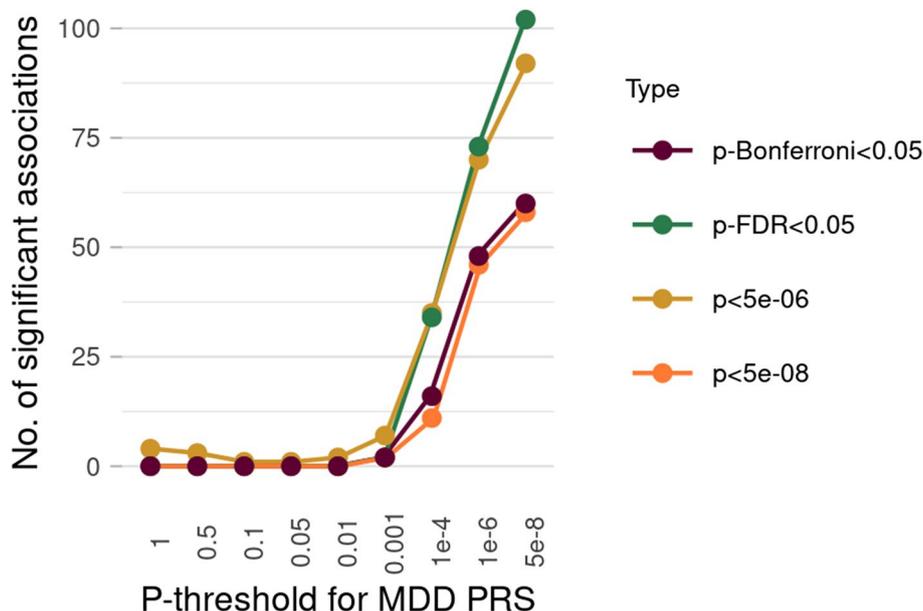


Fig. 1 Number of CpG probes associated with polygenic risk scores (PRS) at nine different p-thresholds (pTs) for discovery analysis. X-axis represents the pTs used for generating PRS. Y-axis shows the number of probes significantly associated with the given PRS. The four different lines represent four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

MR analysis simultaneously, in order to prioritise SNPs that showed the strongest independent casual associations with depression.

Exposures were selected from CpG probes found significantly associated with depression-PRS generated using the p-threshold = 5×10^{-8} . 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, 15 probes were selected for final analyses.

Scripts for all the above analyses can be found in the GitHub repository: https://github.com/xshen796/MDD_PRS_MWAS [53]. A detailed summary for the directories of each individual analysis can be found in the URL: https://github.com/xshen796/MDD_PRS_MWAS/wiki.

Results

Discovery MWAS of depression-PRS in GS

MWAS with depression-PRS at all p-thresholds

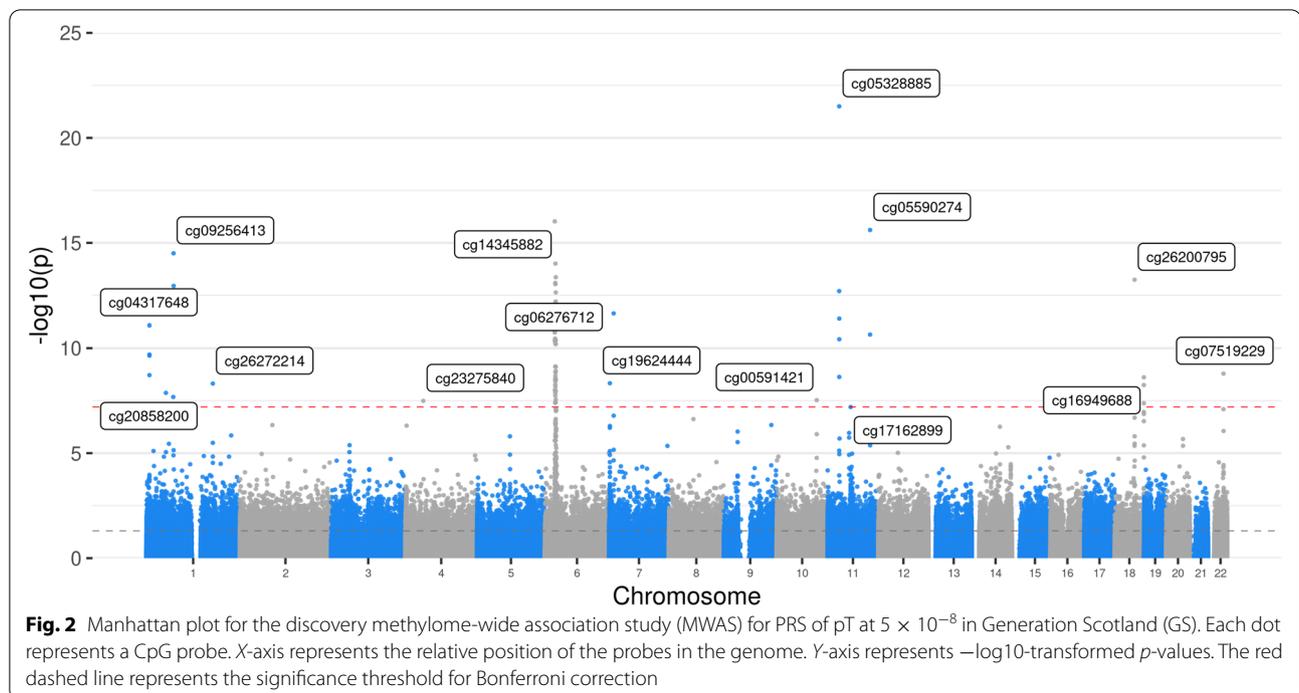
There were 71 CpG probes significantly associated with depression-PRS with p-threshold (pT) at 5×10^{-8} ($p < 6.3 \times 10^{-8}$ to reach significance after Bonferroni correction). In contrast to many other studies that use polygenic risk profiling at different thresholds to predict depression [54], both the number of significant associations and the effect sizes decreased as PRSs were calculated at increasingly lenient thresholds (Fig. 1). For pT of 1×10^{-6} , 29

CpGs were associated with depression-PRS (Additional file 1: Fig. S1). No significant associations were found for PRS using p-value thresholds greater than or equal to 1×10^{-4} . Quantile-quantile plot and statistics for genomic inflation factors (ranged from 0.960 to 0.970) can be found in Additional file 1: Fig. S2 and Table S4. 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, Fig. 2), with 45/71 (63.4%) of significant associations within this region ($p_{\text{Bonferroni}}$ ranged from 0.03 to 7.28×10^{-11}). The top ten probes that showed the greatest associations are listed in Additional file 1: Table S5 (all $p_{\text{Bonferroni}} < 8.62 \times 10^{-8}$). After pruning ($r < 0.1$ for at least two nearest probes, window = 3 Mb), the top CpG probe identified within the MHC region was cg14345882 (all $p_{\text{Bonferroni}} = 7.28 \times 10^{-11}$). UCSC gene database annotation shows genes that are nearest to the significant probes in the MHC region are, for example, *TRIM27*, *HIST1H2AI* and *BTN3A2*. See Tables 1 and 2, Additional file 1: Table S5 and Additional file 2: Table S10.

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 3 Mb 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 were due to the additive effect of many genetic variants included in the MHC region. Analysis (2) was conducted to find out if the CpG associations found within the MHC region were conferred 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 71 to 41, at the depression 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 Additional file 1: Fig. S3.

Outside of the MHC region, 26 probes showed significant associations with depression-PRS estimated across the genome at pT of 5×10^{-8} ($p_{\text{Bonferroni}}$ ranged from 0.049 to 2.41×10^{-16}). The top ten probes are listed in Additional file 1: Table S5. Genes mapping near to the top probes were associated with histone deacetylase, DNA binding and transcriptional processes (such as *MAD1L1*, *TCF4* and *RERE*), and neuronal plasticity and growth (for example, *NEGR1*).

The effect sizes for the significant CpG probes showed high correlations between Set 1 and 2 ($r = 0.84$), 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 12.7% of all significant CpGs located outside of 1 Mb 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 Additional file 2: Table S10).

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)

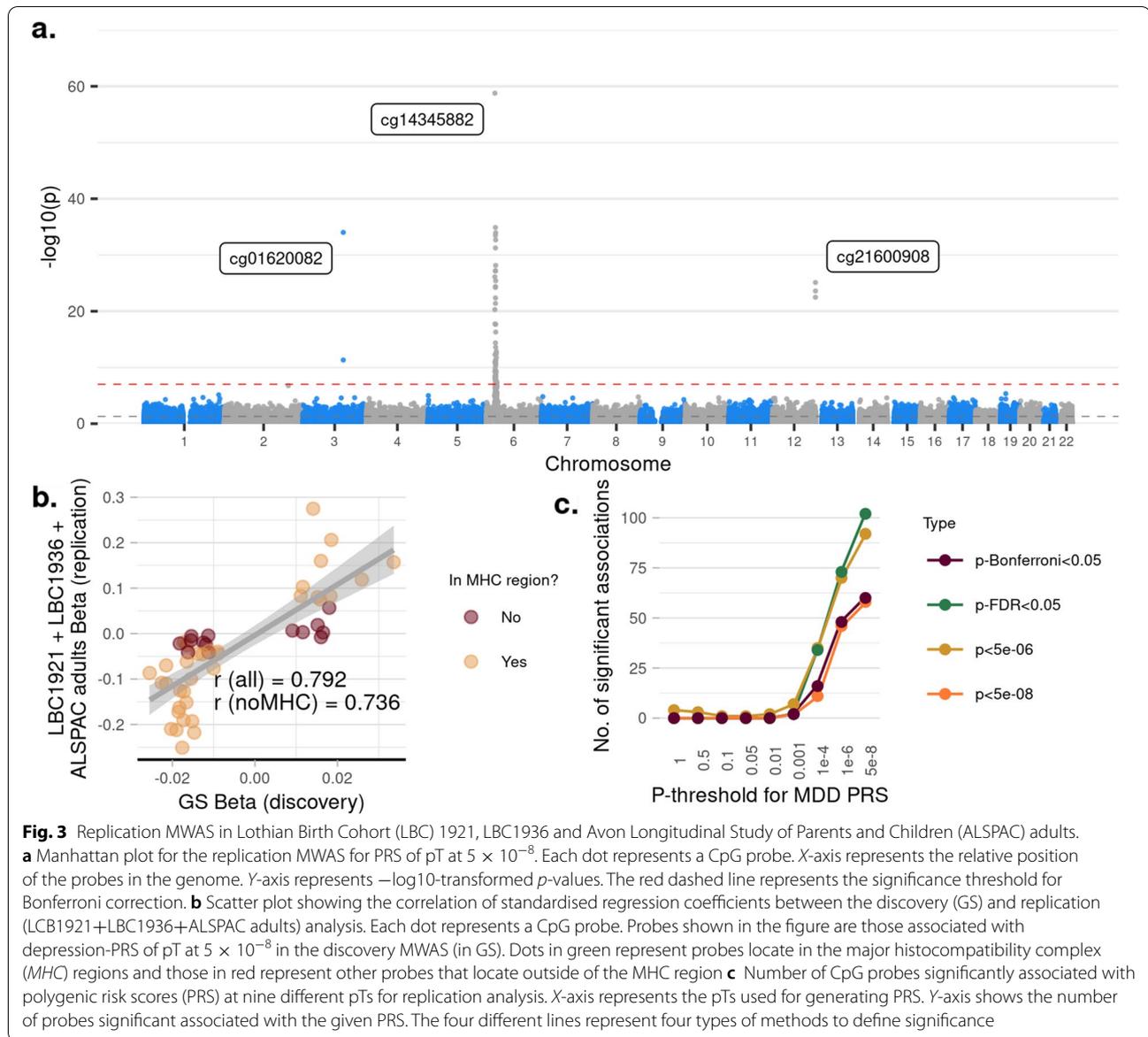
Replication in adults We looked at a subset of CpG probes that were significant in the discovery MWAS analysis and found that the standardised effect sizes were correlated between the discovery and replication meta-MWAS of LBC1921, LBC1936 and ALSPAC adults, with ($N_{\text{probe}} = 49$, $r = 0.79$) or without the probes located in the MHC region ($N_{\text{probe}} = 14$, $r = 0.74$). There were 98.0% associations found in the discovery MWAS which remained in the same direction and 77.6% and 67.3% remained significant before and after Bonferroni correction within the replication analysis, respectively. See Fig. 3.

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:0072643	BP	Interferon-gamma secretion	22	2	2.24E-04	1
	GO:0003050	BP	Regulation of systemic arterial blood pressure by atrial natriuretic peptide	2	1	2.64E-03	1
	GO:0021691	BP	Cerebellar Purkinje cell layer maturation	2	1	3.27E-03	1
	GO:0021590	BP	Cerebellum maturation	3	1	3.64E-03	1
	GO:0021699	BP	Cerebellar cortex maturation	3	1	3.64E-03	1
	GO:1902412	BP	Regulation of mitotic cytokinesis	6	1	3.88E-03	1
	GO:0032609	BP	Interferon-gamma production	105	2	4.06E-03	1
	GO:0072686	CC	Mitotic spindle	103	2	4.94E-03	1
	GO:0046340	BP	Diacylglycerol catabolic process	4	1	5.24E-03	1
	GO:0048408	MF	Epidermal growth factor binding	4	1	5.99E-03	1
Without MHC region	GO:0072686	CC	Mitotic spindle	101	2	2.14E-03	1
	GO:0003050	BP	Regulation of systemic arterial blood pressure by atrial natriuretic peptide	2	1	2.30E-03	1
	GO:1902412	BP	Regulation of mitotic cytokinesis	6	1	2.31E-03	1
	GO:0021691	BP	Cerebellar Purkinje cell layer maturation	2	1	2.45E-03	1
	GO:0021590	BP	Cerebellum maturation	3	1	2.56E-03	1
	GO:0021699	BP	Cerebellar cortex maturation	3	1	2.56E-03	1
	GO:0046340	BP	Diacylglycerol catabolic process	4	1	4.06E-03	1
	GO:0051315	BP	Attachment of mitotic spindle microtubules to kinetochore	13	1	4.23E-03	1
	GO:0021578	BP	Hindbrain maturation	6	1	4.40E-03	1
	GO:0021626	BP	Central nervous system maturation	7	1	4.51E-03	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 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways are listed in the table

	KEGG pathway ID	Description	N	DE	P	P_{FDR}
With MHC region	path:hsa05322	Systemic lupus erythematosus	112	1.5	0.093896415	1
	path:hsa04914	Progesterone-mediated oocyte maturation	88	1	0.10085395	1
	path:hsa04110	Cell cycle	123	1	0.110965189	1
	path:hsa04217	Necroptosis	147	1	0.116705111	1
	path:hsa04114	Oocyte meiosis	117	1	0.118584815	1
	path:hsa04925	Aldosterone synthesis and secretion	95	1	0.137824504	1
	path:hsa04613	Neutrophil extracellular trap formation	159	1.5	0.148531242	1
	path:hsa04514	Cell adhesion molecules	136	1	0.156764027	1
	path:hsa04723	Retrograde endocannabinoid signalling	134	1	0.157693357	1
	path:hsa05034	Alcoholism	163	1.5	0.162889755	1
Without MHC region	path:hsa04914	Progesterone-mediated oocyte maturation	88	1	0.077051	1
	path:hsa04110	Cell cycle	123	1	0.079099	1
	path:hsa04114	Oocyte meiosis	116	1	0.082834	1
	path:hsa04514	Cell adhesion molecules	118	1	0.098456	1
	path:hsa04925	Aldosterone synthesis and secretion	92	1	0.104019	1
	path:hsa04723	Retrograde endocannabinoid signalling	133	1	0.113819	1
	path:hsa05203	Viral carcinogenesis	152	1	0.127128	1
	path:hsa05166	Human T-cell leukaemia virus 1 infection	193	1	0.1555	1
	path:hsa04662	B cell receptor signalling pathway	76	0	1	1
	path:hsa05224	Breast cancer	143	0	1	1

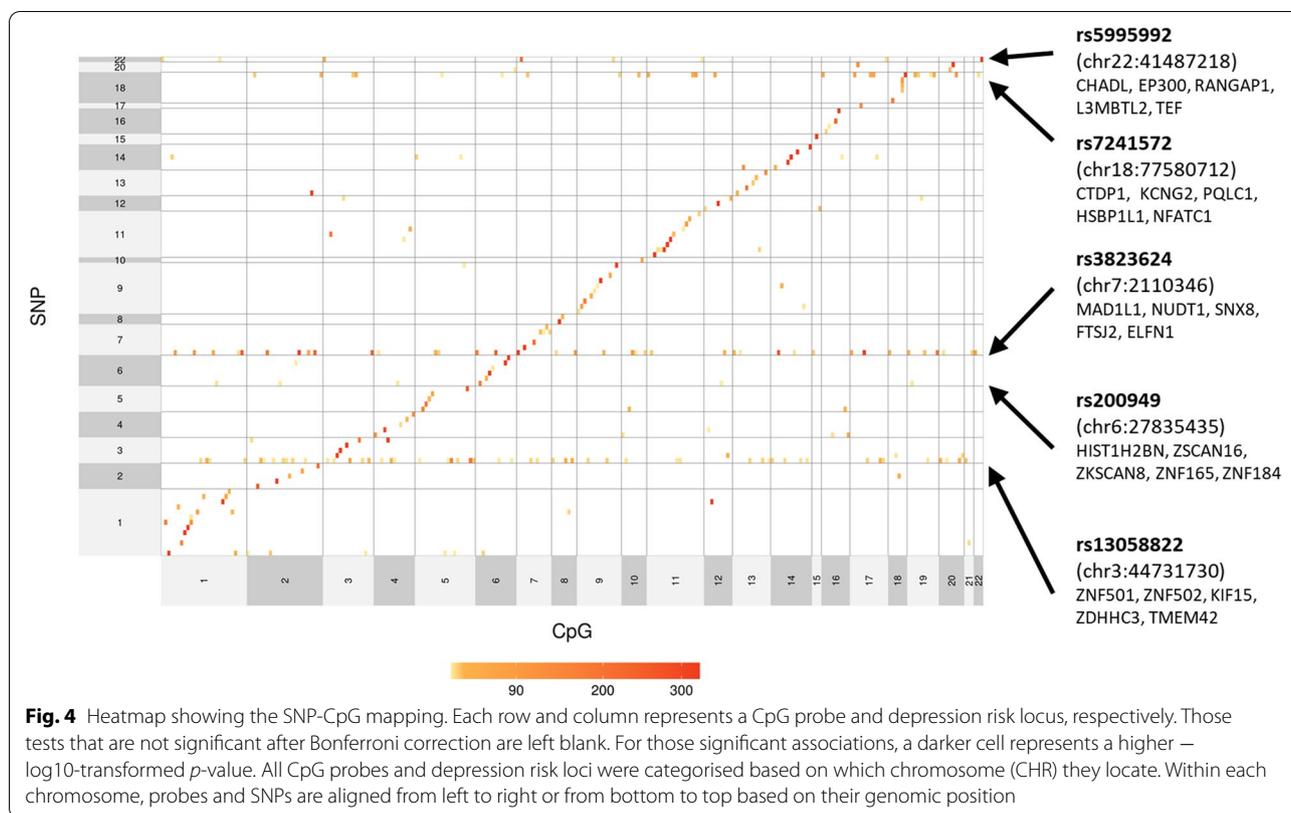


We then looked at the probes within and outside of the MHC region separately. Within the MHC region, effects for the top independent probe remained in the same direction and was significantly replicated ($p_{\text{Bonferroni}} = 8.24 \times 10^{-58}$). For the probes outside of the MHC region, 92.9% of the effects remained in the same direction, 35.7% were nominally significant and 7.1% were significant after Bonferroni correction.

Replication in adolescents 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: $N_{\text{probe}} =$

50, $r = 0.81$; no MHC region: $N_{\text{probe}} = 14$, $r = 0.64$. Effect for 89.8% of the probes remained in the same direction, 68.0% remained nominally significant and 46.0% were significant after Bonferroni correction.

Within the MHC region, effects for the top independent probe remained in the consistent direction and was significant ($p_{\text{Bonferroni}} = 1.71 \times 10^{-12}$). For the probes outside of the MHC region, effects for 85.7% of the probes remained in the same direction, 21.4% were nominally significant and 7.1% were significant after Bonferroni correction.



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

Meta-analysis of the MWAS of depression-PRS for replication cohorts (LBC1921, LBC1936 and ALSPAC adults) showed that, for depression-PRS of p_T at 5×10^{-8} , 1×10^{-6} , 1×10^{-4} and 0.001, the number of significant CpG probes were 60, 48, 16 and 2, respectively. Similar to the discovery analysis, no significant associations were found for PRS of $p_T \geq 0.01$.

A full list of results for replication analysis can be found in Additional file 1: Figs. S4-S7 and Additional file 2: Table S10.

Pathway enrichment analysis

GO terms and KEGG pathways were assessed for the genes associated with depression-PRS of p_T at 5×10^{-8} . There were 118 enriched GO terms nominally significant but none was significant after FDR correction ($p_{\min} = 2.02 \times 10^{-3}$). The majority of the nominally significant GO terms were associated with immune response and brain maturation. No enriched KEGG pathways reached significance ($p > 0.089$). The top ten GO terms and KEGG pathways are listed in Tables 1 and 2.

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. The analysis aimed to further inform individual associations between each genetic risk locus and DNA methylation. There were 3969 CpG probes that showed significant associations with at least one leading genetic risk variant after Bonferroni correction. Significant associations after Bonferroni correction are described below ($p < 1.31 \times 10^{-7}$).

There were 94 of the 96 genetic risk variants tested showed significant *cis* association with CpGs within 1 Mb distance (see Fig. 4). There were 46 genetic risk variants (47.9% of all variants tested) that showed *trans* associations outside of their 1-Mb window and 33 variants (34.4% 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 eight of the distal autosomal chromosomes (see Fig. 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 1 Mb distance from the five genetic variants can be found in Additional file 1: Fig. S8.

Colocalisation analysis

We hypothesised that SNPs influencing risk for depression and those influencing DNAm would be shared. Colocalisation analysis, however, indicated that there was no strong evidence ($PP_4 > 0.8$, $PP_4/PP_3 > 5$) [55] for shared genetic factors between loci for depression and DNAm. The posterior probability for one region was supportive of a suggestive colocalised association signal for both depression and DNAm in that region ($PP_4 = 0.71$) [56]. Within this region, the SNP with the highest posterior probability of being shared SNPs for the two traits (66%) was rs73163796, which colocalised with genetic variation influencing a smoking-associated CpG site, cg15099418 [57]. Additional file 3 contains results for all 102 regions investigated in colocalisation analysis.

MR

Discovery MR: causal effect of DNAm on depression using GoDMC data

Eight probes: cg06552810, cg07519229, cg14159747, cg14345882, cg14844989, cg19624444, cg23275840 and cg26647111, showed significant causal effect using the IVW and WM methods (absolute β_{IVW} ranged from 0.017 to 0.040, p_{FDR} ranged from 4.48×10^{-3} to 3.44×10^{-8} , absolute β_{WM} ranged from 0.012 to 0.038, p_{FDR} ranged from 1.75×10^{-3} to 2.07×10^{-16} , p_{FDR} for Q-statistics ranged from 0.071 to 8.74×10^{-7}). Effect sizes 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 ($\beta_{MR-Egger}$ ranged from 0.004 to 0.034, p_{FDR} ranged from 0.703 to 0.319). However, the direction of effects remained the same with the IVW and WM methods and the test for MR-Egger intercept showed no evidence of horizontal pleiotropy (p_{FDR} for MR-Egger intercept ranged from 0.797 to 0.267).

Results are also shown in Fig. 5 and Additional file 1: Fig. S9 and Table S6.

Replication MR: causal effect of DNAm on depression using GS data

A replication MR was conducted to look at the causal effect of DNAm on depression, using an independent set of mQTL data. All of the potentially causal MR effects of DNAm to depression found in the discovery 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 0.641 to 0.953). Four out of eight significant effects found in the discovery MR analysis were significant for all three MR methods in the replication analyses (absolute β_{IVW} ranged from 0.048 to 0.199, $p_{IVW-FDR}$ ranged from 1.20×10^{-7} to 1.15×10^{-26} ; absolute β_{WM} ranged from 0.032 to 0.192, p_{WM-FDR} ranged from 3.53×10^{-3} to 7.45×10^{-30}). Three other probes (cg06552810, cg14844989 and cg26647111) showed significant causal effect at IVW and WM MR methods (see statistics in Additional file 1: Table S3). MR-Egger intercepts were not significantly deviated from 0 for all replication MR ($p_{FDR} > 0.61$), and thus showed no evidence of horizontal pleiotropy. See Additional file 1: Table S7.

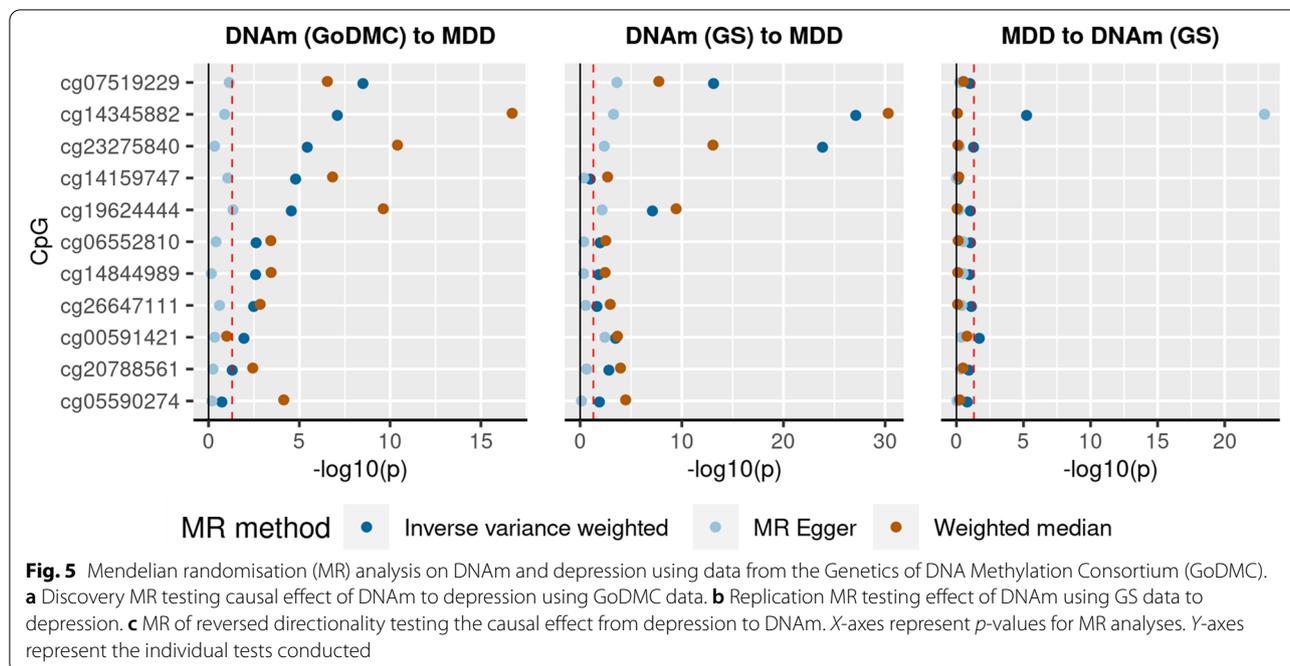
Multi-variable MR: independent causal effect of DNAm on depression using GS data

We next tested for causal associations between DNAm at multiple CpGs from the discovery analysis to depression. The significant probes were entered into the two-sample MR analysis simultaneously, to identify the set of independent CpGs that showed the strongest and independent causal associations with depression using the IVW method. Three probes showed putatively causal effects when all CpGs were considered simultaneously. They are as follows: cg23275840 on chromosome 4, cg14345882 on chromosome 6 and cg14844989 on chromosome 11 (absolute β_{IVW} ranged from 0.129 to 3.028, p_{FDR} ranged from 0.025 to 2.21×10^{-7} , see Additional file 1: Fig. S10 and Table S8). Genes annotated with these CpG probes are *BTN3A2* and *CORIN*. These genes are involved in signalling receptor binding in the brain and in hormonal regulation.

MR: causal effects of depression liability on DNA methylation

MR provided no consistent evidence of a causal effect of depression liability on DNA methylation. The effects to cg09256413 and cg16996682 were significant for the IVW method (absolute β_{IVW} ranged from 0.066 to 0.086, p_{FDR} ranged from 0.028 to 0.003), but the effects were not significant for neither WM nor MR-Egger methods (absolute $\beta_{WM/MR-Egger}$ ranged from 0.003 to 0.251, $p_{FDR} > 0.12$). The effect from depression to cg14345882 was significant for both IVW and MR-Egger ($\beta_{IVW/MR-Egger} = 0.004$ to -6.905 , $p_{FDR} < 8.85 \times 10^{-5}$), but the effect was not significant for the WM method ($\beta_{WM} = 0.004$, $p_{FDR} = 0.882$) with an opposite direction to the other two methods, and there was strong evidence for heterogeneity between genetic instruments (p for Q-statistics $< 1 \times 10^{-328}$). All other effects were not significant after FDR correction ($p_{FDR} > 0.074$).

See Fig. 5, Additional file 1: Fig. S11 and Table S9.



Discussion

PRS for depression showed widespread associations with peripheral blood DNAm across the methylome in Generation Scotland: Scottish Family Health Study Cohort (GS, $N = 8898$). DNAm associations showed highly consistent results in the replication analysis in adults ($N = 2049$, $r_{\beta} = 0.79$) and in adolescents ($N = 423$, $r_{\beta} = 0.81$). 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 widespread *trans* effect across more than eight of the autosomal chromosomes. Finally, using Mendelian randomisation (MR), we found evidence of a mutually causal effect of DNAm on liability to depression at CpG probes associated with PRS 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 disorder [6]. Other genes that located outside of the MHC region, such as *MAD1L1*, *RERE*, *SORCS3* and *ANKK1*, are associated with neuronal development and guidance of neuronal growth [58], transcriptional processes [59, 60] and other risk factors for depression, for example, obesity, smoking and abnormal physical development [61].

DNAm is associated with both genetic and environmental factors that collectively contribute to disease liability [62]. The present study focuses on investigating the associations between genetic risk of depression and DNAm, as well as the mechanism of genetic risk that penetrates through DNAm to disease liability. To our knowledge, this is the first such investigation using PRS for depression, which is likely to be due to a lack of statistical power within previously available samples. However, small-scale twin studies, although using a different design, have shown consistent findings with the present study. For example, differential methylation in CpG probes mapping near *MAD1L1* were found in affected depression patients compared with their unaffected monozygotic twins [63]. The finding was replicated later in a large-scale study of 724 twin pairs [64]. Compared to the previous and smaller-scale twin studies that showed small numbers of findings, the present study showed novel associations implicating genes involved in brain maturation and synaptic processing. This may indicate a broader mechanistic and potentially mediating role for DNAm in conferring the downstream effects of genetic risk. Our findings also highlight that DNAm may facilitate the functional interpretation of genetic risk loci.

MR 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, three probes showed an

independent causal effect on depression. Genes annotated with these independent probes are associated with phenotypes such as lower total brain volume [65], higher C-reactive protein [66], obesity [67] and adverse lifestyle factors such as smoking [68], which are implicated in both patients with depression and those who have been exposed to early environmental risks, such as childhood trauma [69]. These phenotypes have also been shown to have causal effects on depression in previous MR studies [4, 5].

Statistical evidence was stronger for the causal effect from DNAm 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 4 to 22 and N_{SNP} for depression to DNAm was 122). 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 depression [70].

The present study utilises large samples with replication analyses yielding highly consistent results. One limitation for interpreting the current findings is that DNAm 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, 71]. 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 addition, findings from the present study were supported by European samples. Future studies regarding other ancestry groups are necessary for identifying more generalisable genetic-epigenetic associations.

Conclusions

In the current study, we demonstrate that genome-wide genetic risk variants for depression show widespread methylome-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.

Abbreviations

DNAm: DNA methylation; MWAS: Methylome-wide association study; CpG: Cytosine-guanine dinucleotide; PRS: Polygenic risk score(s); MR: Mendelian randomisation; GS: Generation Scotland: Scottish Family Health Study; LBC: Lothian Birth Cohort; ALSPAC: Avon Longitudinal Study of Parents and Children; GoDMC: Genetics of DNA Methylation Consortium; SD: Standard deviation; mQTL: Methylation quantitative trait loci; MHC: Major histocompatibility complex; HLA: Human leukocyte antigen.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-022-01039-5>.

Additional file 1: Figure S1. Manhattan plots for discovery MWAS on GS. **Figure S2.** Quantile-quantile plots for methylome-wide association studies (MWAS) of polygenic risk scores (PRS) for depression at p threshold of 5e-8 on Generation Scotland: Scottish Family Health Study (GS). **Figure S3.** Supplementary MWAS investigating the MHC region. **Figure S4.** Manhattan plots for replication MWAS on LBC 1921, LBC 1963 and ALSPAC adults. **Figure S5.** Quantile-quantile plots for replication MWAS of PRS for depression at p threshold of 5 × 10⁻⁸ on Lothian Birth Cohort (LBC) 1921, LBC 1936 and Avon Longitudinal Study of Parents and Children (ALSPAC) adults. **Figure S6.** Manhattan plots for replication MWAS ALSPAC children. **Figure S7.** Quantile-quantile plots for replication MWAS of PRS for depression at p threshold of 5 × 10⁻⁸ on ALSPAC children. **Figure S8.** Regional association plots for the genetic loci showed associations with methylation levels at CpGs located in more than half of the distal autosomal chromosomes (window = 1Mb). **Figure S9.** Scatter plot for discovery Mendelian randomisation (MR) of DNA methylation (DNAm) to depression. **Figure S10.** Scatter plot for replication MR of DNAm to depression. **Figure S11.** Scatter plot for MR of liability of depression to DNAm. **Figure S12.** Supplementary MWAS on the unrelated sample in GS (N=6777). **Table S1.** Number of genetic variants used for calculating PRSs. **Table S2.** Association between PRSs and prevalence depression. **Table S3.** Association between PRSs and DNAm-estimated white-blood cell proportions. **Table S4.** Genomic inflation factor for discovery, adult replication (LBC 1921 + LBC 1936 + ALSPAC adults) and adolescent replication (ALSPAC adolescents) MWAS. **Table S5.** Top CpG probes associated with PRS at pT = 5e-8. **Table S6.** Results for discovery MR of DNAm to depression. **Table S7.** Results for replication MR of DNAm to depression. **Table S8.** Results for multivariable MR of DNAm to depression. **Table S9.** Results for MR of liability of depression to DNAm.

Additional file 2: Table S10. Association statistics for the discovery MWAS and replication MWASs in the adult and adolescent samples.

Additional file 3. Results for colocalisation analysis between depression and mQTL.

Acknowledgements

We thank the participants and team members for their ongoing contribution to the recruitment, data management and technical and legal support for GS. We are grateful to all the participants and team members of LBC1936 and LBC1921. We thank 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.

Authors' contributions

XS and AMM conceived the study concept. XS conducted analysis for GS, LBC1921, LBC1936 and meta-analyses. DC conducted MWAS for ALSPAC. MJA developed part of the study pipeline. RMW and SWM contributed to data preparation for GS. JLM and GH contributed to data preparation for GoDMC. SHE, SRC and REM contributed to data preparation for LBC1921 and LBC1936. AK and DC contributed to data preparation for ALSPAC. KLE, IJD, CLR, REM, KLE, HCW and MCB edited the manuscript. AMM oversaw and supervised the project. All authors read and approved the final manuscript.

Funding

This work is supported by two Wellcome Trust grants to AMM (Investigator Award in Science: 220857/Z/20/Z and Strategic Award 104036/Z/14/Z). For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. In addition, DNAm profiling was supported by funding from NARSAD (Ref 27404, DMH) and the Royal College of Physicians of Edinburgh (SIM Fellowship, HCW). Genotyping of the GS samples was funded by the MRC and Wellcome Trust [104036/Z/14/Z]. GS also receives support from the Chief

Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006].

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 SEH), 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. SRC and IJD were supported by a National Institutes of Health (NIH) research grant R01AG054628, and SRC 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. Part of this data was collected using REDCap, see the REDCap website for details <https://projectredcap.org/resources/citations/>. CLR, DC, JLM and GH are supported by the MRC Integrative Epidemiology Unit at the University of Bristol (MC_UU_000115/5).

Availability of data and materials

All codes used for generating the PRS, preparing genetic data and analysis have been stored in a publicly available GitHub repository in the GitHub repository: https://github.com/xshen796/MDD_PRS_MWAS [53]. A detailed summary of scripts used for each analysis can be found in the wiki page for the GitHub repository: https://github.com/xshen796/MDD_PRS_MWAS/wiki. Summary statistics for the association analyses conducted in the present study can be found in Additional file 2: Table S10. Summary statistics for depression GWAS that was used for generating the PRS can be found in the URL: <https://datashare.ed.ac.uk/handle/10283/3203>. PRS for depression has been previously developed and validated by Howard et al. [33] in GS.

According to the terms of consent, access to any form of individual-level data requires application for each individual cohort. Access to individual-level genetic, DNAm data and phenotypes need to be approved by the GS Access Committee (<https://www.ed.ac.uk/generation-scotland/for-researchers/access>, mailto: access@generationscotland.org). Data dictionary for GS is available at the URL: <https://datashare.ed.ac.uk/handle/10283/2988>. Access to LBC1921 and LBC1936 must be approved by the LBC research team. A guideline for accessing LBC data can be found in the URL: <https://www.ed.ac.uk/lothian-birth-cohorts/data-access-collaboration>. Data structure, application procedure and contact details are described in the guideline. Access to ALSPAC data requires approved application. Data dictionary and requirements for data access are described in detail in the URL: <http://www.bristol.ac.uk/alspac/researchers/access/>.

Declarations

Ethics approval and consent to participate

Written consent was obtained from all participants for each individual cohort in the present study. Consent for biological samples has been collected for research use. The study was conducted in accordance with the Declaration of Helsinki. Details for ethical approvals are included in the methods section and below:

Generation Scotland (GS): Written consent was obtained from all participants in GS. The study was approved by the NHS Tayside Research Ethics committee (05/s1401/89).

Lothian Birth Cohort (LBC) 1921 and LBC1936: 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).

The Avon Longitudinal Study of Parents and Children (ALSPAC): 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. 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. Consent for biological samples has been collected in accordance with the Human Tissue Act (2004).

Consent for publication

Not applicable.

Competing interests

AMM previously received support from The Sackler Trust, Illumina and speaker fees from Illumina and Janssen. The remaining authors declare that they have no competing interests.

Author details

¹Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Morningside Park, Edinburgh EH10 5HF, UK. ²College of Life and Environmental Sciences, Psychology, University of Exeter, Exeter, UK. ³Medical Research Council Integrative Epidemiology Unit, Bristol Medical School, Population Health Sciences, University of Bristol, Bristol, UK. ⁴Centre for Clinical Brain Sciences, Chancellor's Building, 49 Little France Crescent, Edinburgh BioQuarter, Edinburgh, UK. ⁵Lothian Birth Cohorts group, Department of Psychology, The University of Edinburgh, Edinburgh, UK. ⁶Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK.

Received: 2 August 2021 Accepted: 14 March 2022

Published online: 31 March 2022

References

- Marcus M, Yasamy MT, van Ommeren M, Chisholm D, Saxena S. Depression: a Global Public Health Concern; 2012. p. 6–8.
- Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry*. 2000;157(10):1552–62.
- Sullivan PF, Daly MJ, O'Donovan M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat Rev Genet*. 2012;13(8):537–51 <https://doi.org/10.1038/nrg3240>.
- Howard DM, Adams MJ, Clarke TK, Hafferty JD, Gibson J, Shirali M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci*. 2019;22(3):343–52 <https://doi.org/10.1038/s41593-018-0326-7>.
- Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, & Bacanu SA. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet*. 2018;50(May):668–81.
- International T, Consortium S, International Schizophrenia C, Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;10(AuGuST):748–52 <https://doi.org/10.1038/nature08185>.
- Choi SW, Mak TS-H, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc*. 2020;15(9):2759–72 <https://doi.org/10.1038/s41596-020-0353-1>.
- Shen X, Howard DM, Adams MJ, Hill WD, Clarke TK, McIntosh AM, et al. A phenome-wide association and Mendelian randomisation study of polygenic risk for depression in UK Biobank. *Nat Commun*. 2020;11(1):1–16. <https://doi.org/10.1038/s41467-020-16022-0>
- Barbu MC, Shen X, Walker RM, Howard DM, Evans KL, Whalley HC, et al. Epigenetic prediction of major depressive disorder. *Mol Psychiatry*. 2020;1–12 <https://doi.org/10.1038/s41380-020-0808-3>.
- Clark SL, Hattab MW, Chan RF, Shabalina AA, Han LKM, Zhao M, et al. A methylation study of long-term depression risk. *Mol Psychiatry*. 2020;25(6):1334–43 <https://doi.org/10.1038/s41380-019-0516-z>.

11. Zeng Y, Amador C, Xia C, Marioni R, Sproul D, Walker RM, et al. Parent of origin genetic effects on methylation in humans are common and influence complex trait variation. *Nat Commun.* 2019;10(1):1–13 <https://doi.org/10.1038/s41467-019-09301-y>.
12. Van Dongen J, Nivard MG, Willemsen G, Hottenga JJ, Helmer Q, Dolan CV, et al. Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat Commun.* 2016;7 <https://doi.org/10.1038/ncomms11115>.
13. Min J, Hemani G, Hannon E, Dekkers K, Castillo-Fernandez J, Luijk R, et al. Genomic and phenomic insights from an atlas of genetic effects on DNA methylation. *Nat Genet.* 2020;25:81.
14. Hannon E, Spiers H, Viana J, Pidsley R, Burrage J, Murphy TM, et al. Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nat Neurosci.* 2015;19(1):48–54 <https://doi.org/10.1038/nn.4182>.
15. Hannon E, Dempster E, Viana J, Burrage J, Smith AR, Macdonald R, et al. An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. *Genome Biol.* 2016;17(1):176 <https://doi.org/10.1186/s13059-016-1041-x>.
16. Hannon E, Schendel D, Ladd-Acosta C, Grove J, Hansen CS, Andrews SV, et al. Elevated polygenic burden for autism is associated with differential DNA methylation at birth. *Genome Med.* 2018;10(1):1–13 <https://doi.org/10.1186/s13073-018-0527-4>.
17. Sullivan PF, Posthuma D. Biological pathways and networks implicated in psychiatric disorders. *Curr Opin Behav Sci.* 2015;2:58–68 <https://doi.org/10.1016/j.cobeha.2014.09.003>.
18. Sullivan PF, Geschwind DH. Defining the genetic, genomic, cellular, and diagnostic architectures of psychiatric disorders. *Cell.* 2019;177(1):162–83 <https://doi.org/10.1016/j.cell.2019.01.015>.
19. Smith BH, Campbell H, Blackwood D, Connell J, Connor M, Deary IJ, et al. Generation Scotland: The Scottish Family Health Study; a new resource for researching genes and heritability. *BMC Med Genet.* 2006;7(1):74 <https://doi.org/10.1186/1471-2350-7-74>.
20. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, et al. Cohort profile: Generation Scotland: Scottish family health study (GS: FHS). The study, its participants and their potential for genetic research on health and illness. *Int J Epidemiol.* 2013;42(3):689–700 <https://doi.org/10.1093/ije/dys084>.
21. Deary IJ, Whiteman MC, Starr JM, Whalley LJ, Fox HC. The impact of childhood intelligence on later life: following up the Scottish Mental Surveys of 1932 and 1947. *J Pers Soc Psychol.* 2004;86(1):130–47 <https://doi.org/10.1037/0022-3514.86.1.130>.
22. Deary IJ, Whalley LJ, Starr JM. A lifetime of intelligence: follow-up studies of the Scottish mental surveys of 1932 and 1947. 2009; <https://doi.org/10.1037/11857-000>
23. Fraser A, Macdonald-wallis C, Tilling K, Boyd A, Golding J, Davey smith G, et al. Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol.* 2013;42(1):97–110 <https://doi.org/10.1093/ije/dys066>.
24. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al. Cohort profile: The 'Children of the 90s'-The index offspring of the Avon Longitudinal Study of Parents And Children. *Int J Epidemiol.* 2013;42(1):111–27 <https://doi.org/10.1093/ije/dys064>.
25. Cox SR, Bastin ME, Ferguson KJ, Maniega SM, MacPherson SE, Deary IJ, et al. Brain white matter integrity and cortisol in older men: the Lothian Birth Cohort 1936. *Neurobiol Aging.* 2015;36(1):257–64 <https://doi.org/10.1016/j.neurobiolaging.2014.06.022>.
26. Hillary R, Trejo-Banos D, Kousathanas A, McCartney D, Harris S, Stevenson A, et al. Integrative omics approach to identify the molecular architecture of inflammatory protein levels in healthy older adults. *bioRxiv.* 2020:1–15 <https://doi.org/10.1101/2020.02.17.952135>.
27. Northstone K, Lewcock M, Groom A, Boyd A, Macleod J, Timpson N, et al. The Avon Longitudinal Study of Parents and Children (ALSPAC): an update on the enrolled sample of index children in 2019. *Wellcome Open Res.* 2019;4 <https://doi.org/10.12688/wellcomeopenres.15132.1>.
28. Relton CL, Gaunt T, McArdle W, Ho K, Duggirala A, Shihab H, et al. Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES). *Int J Epidemiol.* 2015;44(4):1181–90 <https://doi.org/10.1093/ije/dyv072>.
29. Davies G, Tenesa A, Payton A, Yang J, Harris SE, Liewald D, et al. Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Mol Psychiatry.* 2011;16(10):996–1005 <https://doi.org/10.1038/mp.2011.85>.
30. Kwong ASF, Morris TT, Pearson RM, Timpson NJ, Rice F, Stergiakouli E, 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; <https://doi.org/10.1111/jcpp.13422>.
31. Gunderson KL. Whole-genome genotyping on bead arrays. *Methods Mol Biol.* 2009;529:197–213 https://doi.org/10.1007/978-1-59745-538-1_13.
32. McCartney DL, Min JL, Richmond RC, Lu AT, Sobczyk MK, Davies G, et al. Genome-wide association studies identify 137 genetic loci for DNA methylation biomarkers of aging. *Genome Biol.* 2021;22(1):25 <https://doi.org/10.1186/s13059-021-02398-9>.
33. Howard DM, Adams MJ, Clarke TK, Hafferty JD, Gibson J, Shirali M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci.* 2019;22(3):343–52 <https://doi.org/10.1101/433367>.
34. Fortin JP, Fertig E, Hansen K. shinyMethyl: Interactive quality control of Illumina 450K DNA methylation arrays in R. *F1000Research.* 2014;3:1–11.
35. Pidsley R, Wong Y, CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics.* 2013;14(1):293 <https://doi.org/10.1186/1471-2164-14-293>.
36. Min JL, Hemani G, Smith GD, Relton C, Suderman M. Meffil: Efficient normalization and analysis of very large DNA methylation datasets. *Bioinformatics.* 2018;34(23):3983–9 <https://doi.org/10.1093/bioinformatics/bty476>.
37. Walker RM, MacGillivray L, McCafferty S, Wrobel N, Murphy L, Kerr SM, et al. Assessment of dried blood spots for DNA methylation profiling. *Wellcome Open Res.* 2019;4:44 <https://doi.org/10.12688/wellcomeopenres.15136.1>.
38. Jaffe AE, Gao Y, Deep-Soboslay A, Tao R, Hyde TM, Weinberger DR, et al. Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nat Neurosci.* 2015;19(1):40–7 <https://doi.org/10.1038/nn.4181>.
39. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics.* 2010;11(1):587 <https://doi.org/10.1186/1471-2105-11-587>.
40. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, et al. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet.* 2011;43(6):519–25 <https://doi.org/10.1038/ng.823>.
41. Zhang Q, Marioni RE, Robinson MR, Higham J, Sproul D, Wray NR, et al. Genotype effects contribute to variation in longitudinal methylome patterns in older people. *Genome Med.* 2018;10(1):75 <https://doi.org/10.1186/s13073-018-0585-7>.
42. Shah S, McRae AF, Marioni RE, Harris SE, Gibson J, Henders AK, et al. Genetic and environmental exposures constrain epigenetic drift over the human life course. *Genome Res.* 2014;24(11):1725–33 <https://doi.org/10.1101/gr.176933.114>.
43. McCartney DL, Walker RM, Morris SW, McIntosh AM, Porteous DJ, Evans KL. Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genomics Data.* 2016;9:22–4 <https://doi.org/10.1016/j.gdata.2016.05.012>.
44. Willer CJ, Li Y, Abecasis GR, Overall P. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26(17):2190–1.
45. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47 <https://doi.org/10.1093/nar/gkv007>.
46. Hansen KD. IlluminaHumanMethylationEPICanno.ilm10b4.hg19: annotation for Illumina's EPIC methylation arrays. R Packag version 060; 2017.
47. Phipson B, Maksimovic J, Oshlack A. MissMethyl: An R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics.* 2016;32(2):286–8 <https://doi.org/10.1093/bioinformatics/btv560>.
48. Hemani G. GWAS summary data sources connected to analytical tools. Github. 2020; <https://github.com/MRCIEU/gwasglue>.
49. Wallace C, Rotival M, Cooper JD, Rice CM, Yang JHM, McNeill M, et al. Statistical colocalization of monocyte gene expression and genetic risk variants for type 1 diabetes. *Hum Mol Genet.* 2012;21(12):2815–24 <https://doi.org/10.1093/hmg/dds098>.

50. Hemani G, Tilling K, Davey SG. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet*. 2017;13(11):e1007081 <https://doi.org/10.1371/journal.pgen.1007081>.
51. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7 <https://doi.org/10.7554/eLife.34408>.
52. Zhang F, Chen W, Zhu Z, Zhang Q, Nabais MF, Qi T, et al. OSCA: a tool for omic-data-based complex trait analysis. *Genome Biol*. 2019;20(1):107 <https://doi.org/10.1186/s13059-019-1718-z>.
53. Shen X. DNA methylome-wide association study of genetic risk for depression. Github. 2022; https://github.com/xshen796/MDD_PRS_MWAS.
54. Bigdeli TB, Ripke S, Peterson RE, Trzaskowski M, Bacanu SA, Abdellaoui A, et al. Genetic effects influencing risk for major depressive disorder in China and Europe. *Transl Psychiatry*. 2017;7(3):61 <https://doi.org/10.1038/tp.2016.292>.
55. Guo H, Fortune MD, Burren OS, Schofield E, Todd JA, Wallace C. Integration of disease association and eQTL data using a Bayesian colocalisation approach highlights six candidate causal genes in immune-mediated diseases. *Hum Mol Genet*. 2015;24(12):3305–13 <https://doi.org/10.1093/HMG/DDV077>.
56. Gay NR, Gloude-mans M, Antonio ML, Abell NS, Balliu B, Park Y, et al. Impact of admixture and ancestry on eQTL analysis and GWAS colocalization in GTEx. *Genome Biol*. 2020;21(1):1–20 <https://doi.org/10.1186/S13059-020-02113-0>.
57. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, et al. Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet*. 2016;9(5):436–47 <https://doi.org/10.1161/CIRCGENETICS.116.001506>.
58. Sandrini L, Di Minno A, Amadio P, Ieraci A, Tremoli E, Barbieri SS. Association between obesity and circulating brain-derived neurotrophic factor (BDNF) levels: systematic review of literature and meta-analysis [Internet]. *Int J Mol Sci*. 2018; <https://doi.org/10.3390/ijms19082281>.
59. Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics*. 2014;13(2):397–406 <https://doi.org/10.1074/mcp.M113.035600>.
60. Plaster N, Sonntag C, Schilling TF, Hammerschmidt M. REREa/atrophin-2 interacts with histone deacetylase and Fgf8 signaling to regulate multiple processes of zebrafish development. *Dev Dyn*. 2007;236(7):1891–904 <https://doi.org/10.1002/dvdy.21196>.
61. Mansur RB, Brietzke E, McIntyre RS, Cao B, Lee Y, Japiassú L, et al. BDNF and BMI effects on brain structures of bipolar offspring: results from the global mood and brain science initiative. *Acta Psychiatr Scand*. 2017;136(6):607–14 <https://doi.org/10.1111/acps.12822>.
62. Januar V, Saffery R, Ryan J. Epigenetics and depressive disorders: a review of current progress and future directions. *Int J Epidemiol*. 2015;44(4):1364–87 <https://doi.org/10.1093/IJE/DYU273>.
63. Byrne EM, Carrillo-Roa T, Henders AK, Bowdler L, McRae AF, Heath AC, et al. Monozygotic twins affected with major depressive disorder have greater variance in methylation than their unaffected co-twin. *Transl Psychiatry*. 2013;3(6):e269 <https://doi.org/10.1038/tp.2013.45>.
64. Starnawska A, Tan Q, Soerensen M, McGue M, Mors O, Børglum AD, et al. Epigenome-wide association study of depression symptomatology in elderly monozygotic twins. *Transl Psychiatry*. 2019;9(1):1–14 <https://doi.org/10.1038/s41398-019-0548-9>.
65. Xu Y, Xiao G, Liu L, Lang M. Zinc transporters in Alzheimer's disease [Internet]. *Mol. Brain*. 2019; <https://doi.org/10.1186/s13041-019-0528-2>.
66. Noh H, Paik HY, Kim J, Chung J. The alteration of zinc transporter gene expression is associated with inflammatory markers in obese women. *Biol Trace Elem Res*. 2014;158(1):1–8 <https://doi.org/10.1007/s12011-014-9902-1>.
67. Starnawska A, Tan Q, Soerensen M, McGue M, Mors O, Børglum AD, et al. Epigenome-wide association study of depression symptomatology in elderly monozygotic twins. *Transl Psychiatry*. 2019;9(1):214 <https://doi.org/10.1038/s41398-019-0548-9>.
68. Zhong R, Chen X, Chen X, Zhu B, Lou J, Li J, et al. MAD1L1 Arg558His and MAD2L1 Leu84Met interaction with smoking increase the risk of colorectal cancer. *Sci Rep*. 2015;5(1):1–9 <https://doi.org/10.1038/srep12202>.
69. Pariante CM, Lightman SL. The HPA axis in major depression: classical theories and new developments. *Trends Neurosci*. 2008;31(9):464–8 <https://doi.org/10.1016/j.tins.2008.06.006>.
70. Mooney MA, Ryabinin P, Wilmot B, Bhatt P, Mill J, Nigg JT. Large epigenome-wide association study of childhood ADHD identifies peripheral DNA methylation associated with disease and polygenic risk burden. *Transl Psychiatry*. 2020;10(1):1–12 <https://doi.org/10.1038/s41398-020-0710-4>.
71. Braun PR, Han S, Hing B, Nagahama Y, Gaul LN, Heinzman JT, et al. Genome-wide DNA methylation comparison between live human brain and peripheral tissues within individuals. *Transl Psychiatry*. 2019;9(1):1–10 <https://doi.org/10.1038/s41398-019-0376-y>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

