Supplementary Information

Methylomew-wide association analysis of polygenic risk scores for depression

Shen X. et al.

## Methods

### Covarying confounders in the methylome-wide association study (MWAS)

#### Generation Scotland: **Scottish Family Health Study (GS)**

Technical confounders were pre-corrected. This was achived by residualising m-values against genetic relationship matrix (GRM), batch and estimated cell proportions. GRM was generated using Plink2.01 and was corrected in wave 1 only, as participants were unrelated in wave 2. Other covariates were fitted in the MWAS regression model, which include: age, sex, pack years, ever smoked tobacco and the first 20 principal components of the methylation data.

Ever smoked tobacco was an ordinal variable. Participants were asked to choose from one of the following responses: ‘Yes, currently smoke’ (= 3), ‘Yes, but stopped within the past 12 months’ (= 2), ‘Yes, but stopped for more than 12 months ago’ (= 1) and ‘No, never smoked’ (= 0). This variable was set as factor in all analyses using R.

Principal components of the methylation data was derived using the ‘FactoMineR’ R package2. This step was conducted on the m-values pre-corrected for technical confounders.

#### Lothian Birth Cohort (LBC) 1921 and LBC 1936

All participants were unrelated within each cohort itself and between the two cohorts. All covariates were fitted in the MWAS regression model, which include: estimated cell proportions, batch, age, sex, ever smoked tobacco and the first 20 principal components of the methylation data.

Ever smoked tobacco variable was ordinal and thus was set as factor in all analyses. Responses include: ‘Current smoker’ (= 2), ‘Previous smoker’ (= 1) and ‘Never smoked’ (= 0).

Principal components of the methylation data was derived using the same method as GS.

#### Avon Longitudinal Study of Parents and Children (ALSPAC)

MWAS was conducted on parents and youth respectively. For each of the MWAS, only unrelated participants were included in the analyses. Procedures for selecting unrelated sample can be found in3. Covariates that were fitted in the MWAS analysis include: estimated cell proportions, batch, age, sex, ever smoked tobacco and the first 15 variables derived from surrogate variable analysis.

Ever smoked tabacco was asked around the time of blood draw for DNAm processing. Participants were asked to choose from either smoked (= 1) or not smoked (= 0).

### Meta-analysis on MWAS results

Meta-analysis was conducted using METAL4.

For each individual analyses in the discovery and replication MWAS, meta-analysis was conducted to obtain an overall summary statistics of MWAS results. In the discovery analysis, MWAS on wave 1 and wave 2 GS data was conducted separately and then meta-analysed. In the replication analysis, MWAS was conducted on the two LBC cohorts together and ALSPAC adults. Summary statistics for replication analyses were then meta-analysed.

A final meta-analysis on the discovery and replication MWAS (LBC cohorts and ALSPAC adults) was conducted and presented in the Supplementary Information.

As there was extensive control for relatedness for MWAS in all cohorts, genomic control correction was not included in the analyses.

### Statistics for methylation quantitative trait loci (mQTL)

#### **Genetics of DNA Methylation Consortium** (GoDMC)

Meta-analysis of mQTL analyses was conducted using the same pipeline as described elsewhere5. For the present study, GS, LBC 1921, LBC 1936, GSK & MPIP, Munich Antidepressant Response Signature study (MARS) and Brisbane Systems Genetics Study were removed from the mQTL meta-analysis as they were also included in the MDD GWAS by Howard *et al.*6. Details for the cohorts can be found at the GoDMC website (<http://www.godmc.org.uk/cohorts.html>). As a result, a total of ~20,000 participants were left in the meta-analysis, with no overlapping participants with GS and no overlapping cohort with the MDD GWAS.

#### GS

Analysis of mQTL was conducted on DNAm data in GS for wave 1 and wave 2 separately. The OmicS-data-based Complex trait Analysis package was used for deriving mQTL summary statistics (<https://cnsgenomics.com/software/osca/>)7. DNAm data and covariates were kept consistent with the MWAS. Genetic data used for the mQTL analysis was also used for calculating polygenic risk scores for depression. Finally, meta-analysis between wave 1 and wave 2 was conducted on the mQTL summary statistics for each CpG probe.

## References

1. Yang, J., Lee, S., Goddard, M. & Visscher, P. GCTA: A tool for genome-wide complex trait analysis. *The American Journal of Human Genetics* **88**, 76–82 (2011).

2. Lê, S., Josse, J. & Husson, F. FactoMineR: AnRPackage for multivariate analysis. *Journal of Statistical Software* **25**, (2008).

3. Caramaschi, D. *et al.* Epigenome-wide association study of seizures in childhood and adolescence. *Clinical Epigenetics* **12**, (2020).

4. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

5. Min, J. *et al.* Genomic and phenomic insights from an atlas of genetic effects on dna methylation. (2020).

6. Howard, D. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience* **22**, 343–352 (2019).

7. Zhang, F. *et al.* OSCA: A tool for omic-data-based complex trait analysis. *Genome Biology* **20**, (2019).

##### Supplementary Table 6. Results for discovery MR of DNAm to depression.

| NSNP | Nsnp | Nnomhc | Nmhc |
| --- | --- | --- | --- |
| 5e-08 | 127 | 118 | 9 |
| 1e-05 | 615 | 602 | 13 |
| 1e-04 | 1457 | 1441 | 16 |
| 1e-03 | 4199 | 4176 | 23 |
| 0.01 | 14578 | 14530 | 48 |
| 0.05 | 38284 | 38185 | 99 |
| 0.1 | 58867 | 58742 | 125 |
| 0.5 | 151226 | 150977 | 249 |
| 1 | 199432 | 199113 | 319 |