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 Plink2.0 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 package (ref: Lê S, Josse J, Husson F (2008). “FactoMineR: A Package for Multivariate Analysis.” *Journal of Statistical Software*, **25**(1), 1–18. doi: [10.18637/jss.v025.i01](https://doi.org/10.18637/jss.v025.i01).). 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 in {ref: <https://clinicalepigeneticsjournal.biomedcentral.com/articles/10.1186/s13148-019-0793-z>}. 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 METAL(<https://genome.sph.umich.edu/wiki/METAL_Documentation>).

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 LBC 1921 and LBC 1936 and ALSPAC adults, and then was 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)

##### Supplementary Figure 1.

