

Supplementary methods

UK Biobank

In the UK Biobank (UKBB), electronic health records from the National Health Service registers are linked to the resource and enable follow-up of participants using ICD-9 and ICD-10 (International Classification of Diseases, ninth and tenth versions, respectively) codes. This, in addition to self-reported disease status at the time of recruitment, enables researchers define various disease phenotypes. Hospital admission records are linked to the UKBB through the data fields 41202 and 41204 representing primary and secondary disease diagnosis, respectively.

FinnGen

FinnGen study participants were genotyped using the Illumina and Affymetrix chip arrays (Illumina Inc., San Diego, and Thermo Fisher Scientific, Santa Clara, CA, USA, <https://www.thermofisher.com/>). The data was then imputed using the SISu v3 imputation panel (<http://sisuproject.fi>) resulting in 16,962,023 variants available for association analysis. Summary statistics provided for the FinnGen data association analyses were generated using the SAIGE software¹.

Multi-Trait Analyses of GWAS (MTAG)

MTAG performs an inverse-variance weighted meta-analysis of related phenotypes (ref). MTAG uses GWAS summary statistics (effect size estimates and their standard errors (BETA, SE) and the *P*-values of the associations). Formatting of the summary statistics was done as recommended by the authors². To account for sample overlap present in our UKBB analyses, MTAG applies the bivariate linkage disequilibrium (LD) score regression to estimate the correlation in GWAS estimation error due to sample overlap.

Mendelian randomization

Causality can be established in MR only if it uses valid instrumental variables (IV) defined by three core assumptions: (1) that they have a true effect on the exposure (the relevance assumption) (2) that they only affect the outcome through the risk factor (the exclusion restriction assumption) (3) that they are independent of any measured and unmeasured confounding factors of the exposure–outcome relationship (the independence assumption)³. The genetic

instruments for depression were from a recent large-scale GWAS of depression that reported 102 independent association signals⁴. They meta-analyzed data for 807,553 individuals, including 246,363 from UK Biobank, 23andMe and Psychiatric Genomics Consortium. The cases in the three contributing GWASs were defined using self-reported and clinical data⁴.

Two-step MR

For our mediation analysis we applied the Product method (also known as ‘product of coefficients’)⁵ in which the indirect effect is estimated as the product of coefficients of the total effect of exposure on the mediator (step 1 of two-step MR) and the total effect of the mediator on the outcome (step 2 of two-step MR), image below.

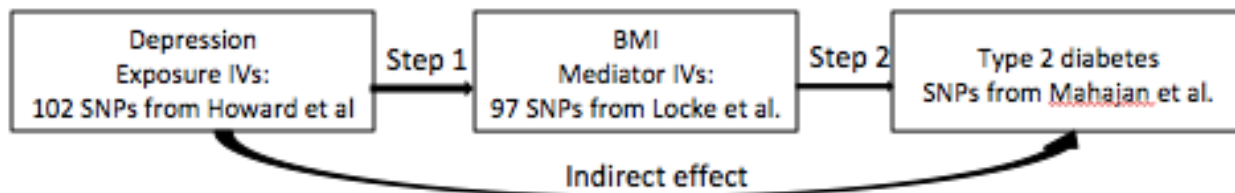


Figure showing the Two-step MR workflow

The estimation of the standard errors of the indirect effect were calculated using “Propagation of errors method”:

$$\text{se.indirect} = \sqrt{\text{se_total}^2 + \text{se_direct}^2}$$

$$\text{se.indirect} = \sqrt{\text{se_a}^2 + \text{se_b}^2}$$

where - se_total is the SE from the exposure -> outcome MR
 - se_direct is the SE from exposure + mediator -> outcome (of exposure->outcome)

The proportion of exposure effect mediated via a mediator is calculated as mediator indirect effect divided by the total effect beta.

LD clumping

To identify independent loci and their lead SNPs for each association result (SP-GWAS and MP-GWAS), we applied a LD-based clumping algorithm performed in PLINK⁶. The following parameters were set: *-clump-p1* 5×10^{-8} (significance threshold for the index SNP), *-clump-p2* 1×10^{-5} (significance threshold for the clumped SNPs), *-clump-r²* 0.01 (LD threshold for clumping) and *-clump-kb* 500kb (distance threshold for clumping). Clumped signals were annotated to their nearest gene using the ANNOVAR software⁷.

References

1. Zhou, W. *et al.* Scalable generalized linear mixed model for region-based association tests in large biobanks and cohorts. *Nat Genet* **52**, 634–639 (2020).
2. Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet* **50**, 229–237 (2018).
3. Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N. & Smith, G. D. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat Med* **27**, 1133–1163 (2008).
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. Burgess, S. *et al.* Dissecting causal pathways using mendelian randomization with summarized genetic data: Application to age at menarche and risk of breast cancer. *Genetics* **207**, 481–487 (2017).
6. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* (2007) doi:10.1086/519795.
7. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164–e164 (2010).

List of supplementary figures

Supplementary Figure 1. Mendelian Randomization analysis outline between depression and type 2 diabetes

Supplementary Figure 2. A) Scatter plot for MR analyses of the causal effect of Depression on Type 2 Diabetes.
B) Scatter plot for MR analyses of the causal effect of Type 2 Diabetes on Depression.

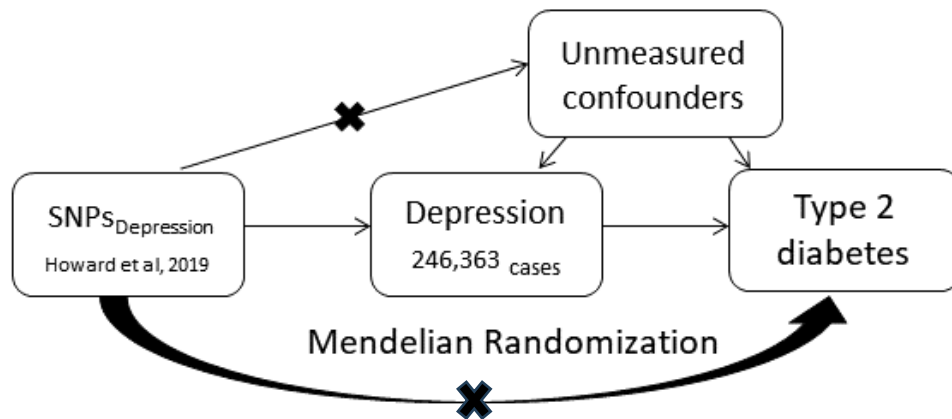
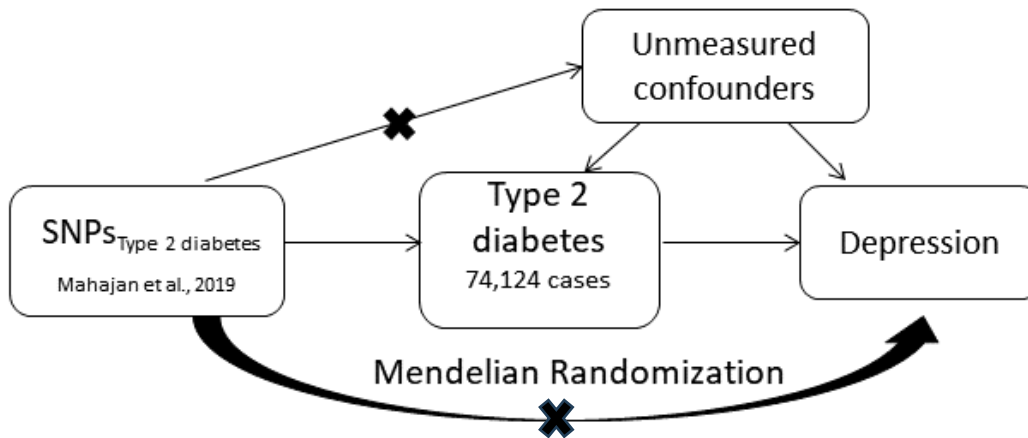
Supplementary Figure 3. Leave-one-out analysis: each row represents a MR analysis of Depression on Type 2 Diabetes

Supplementary Figure 4. Manhattan plots for type 2 diabetes, PHQ-9 and MDD GWAS in the UK Biobank

Supplementary Figure 5. Manhattan plots for type 2 diabetes and MDD after MP-GWAS in MTAG

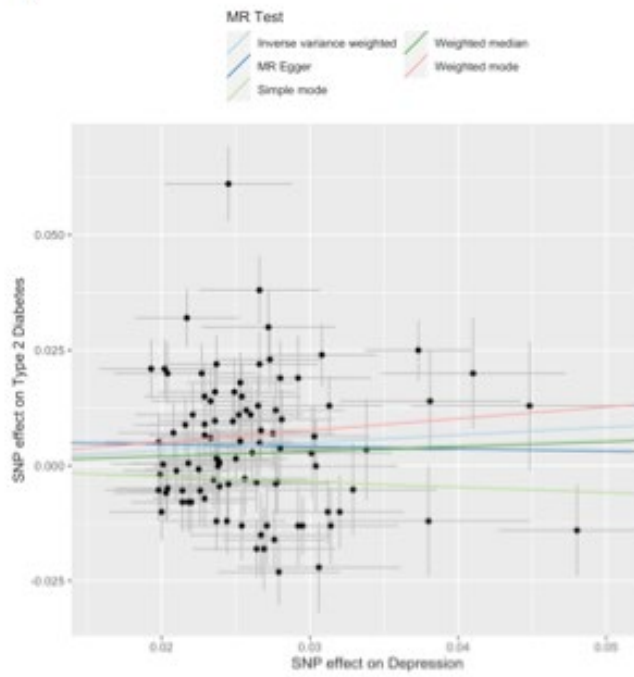
Supplementary Figure 6. Manhattan plots for type 2 diabetes and MDD after MP-GWAS in FinnGen dataset

Supplementary Figure 7. Volcano plots of association in tissues implicated in both type 2 diabetes and PHQ-9

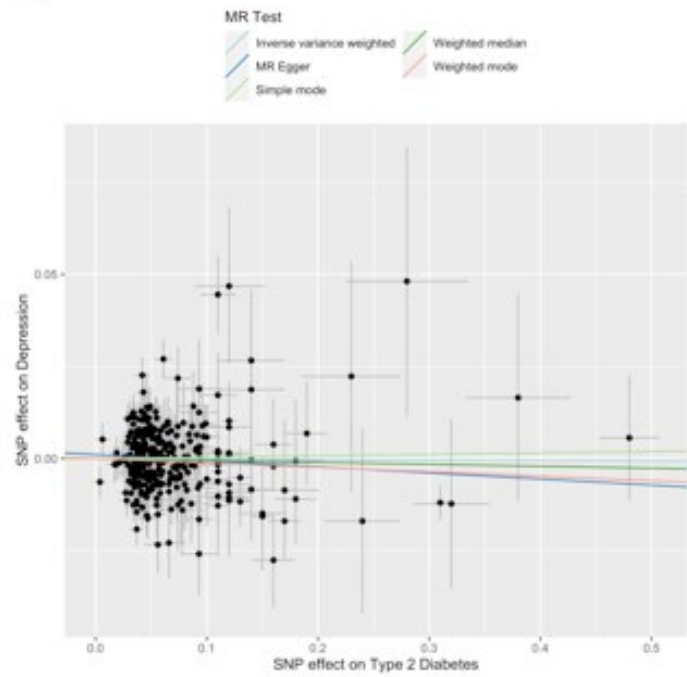
A**B**

Supplementary Figure 1. Mendelian Randomization analysis to explore causality between depression and type 2 diabetes. (A) IV estimator is calculated as the beta coefficient from the association of $GRS_{\text{depression}}$ with type 2 diabetes divided by the beta coefficient from the association of $GRS_{\text{Type 2 diabetes}}$ with depression. (B) The relationship of type 2 diabetes with depression.

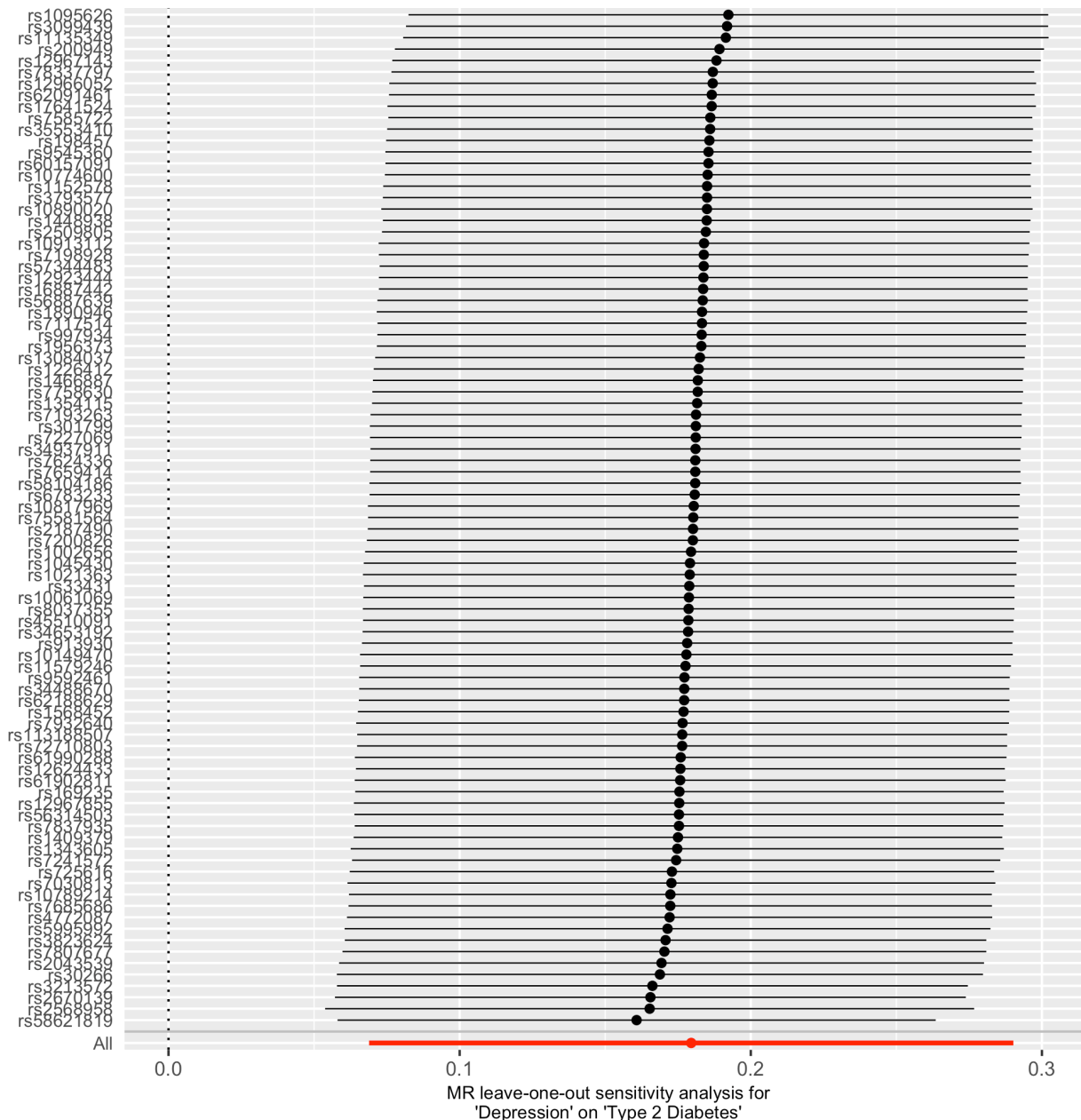
A)



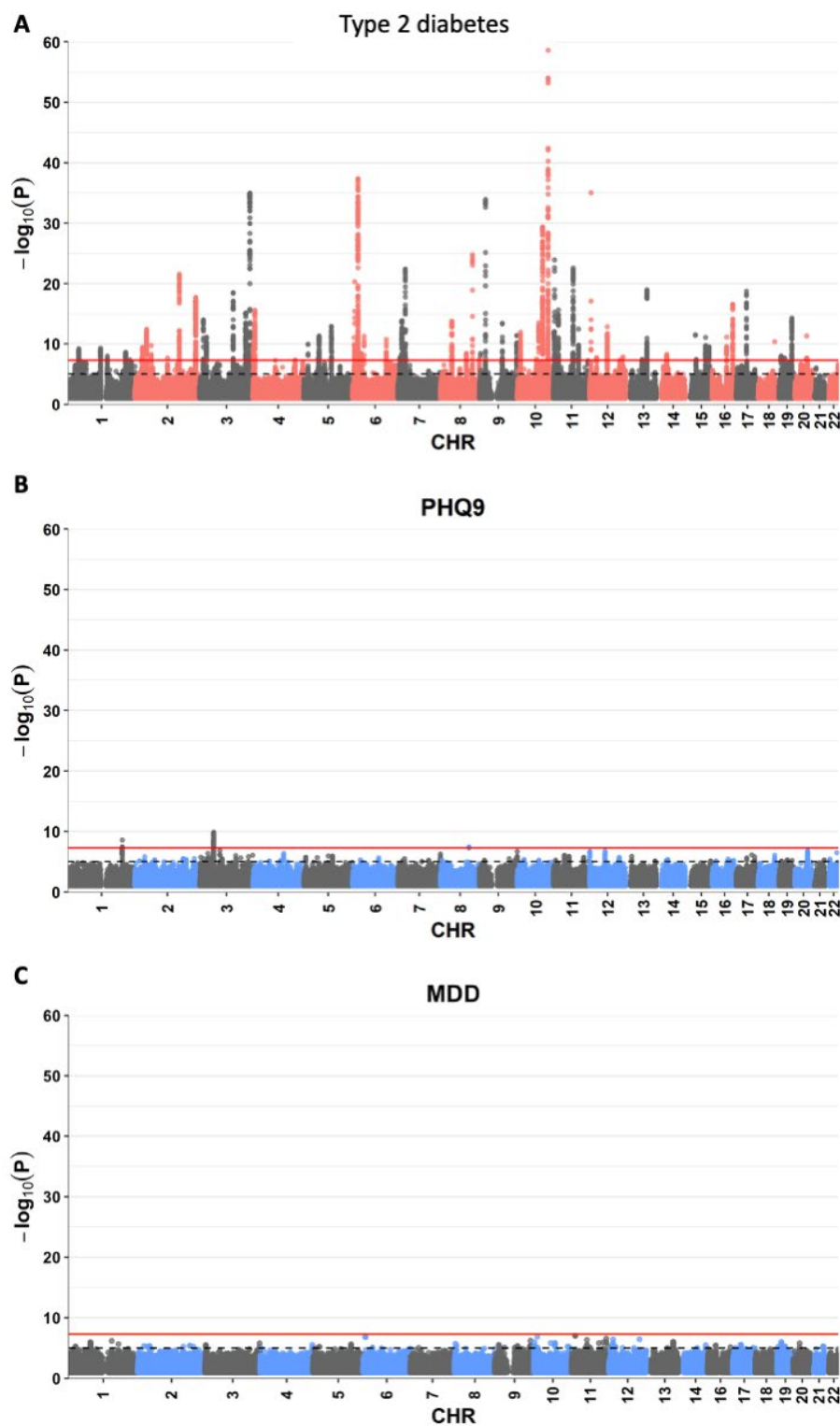
B)



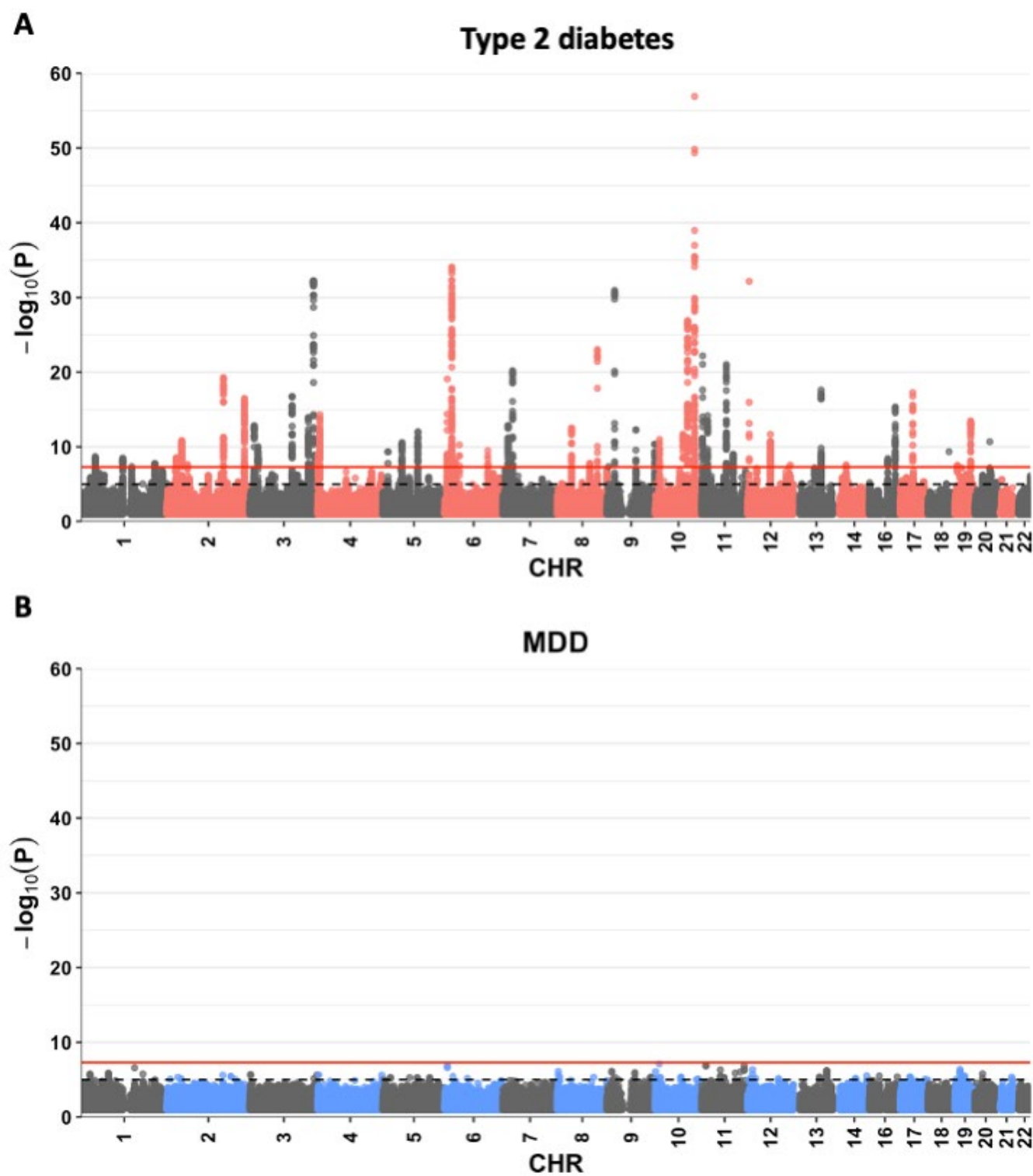
Supplementary Figure 2. A) Scatter plot for MR analyses of the causal effect of Depression on Type 2 Diabetes. B) Scatter plot for MR analyses of the causal effect of Type 2 Diabetes on Depression. Analyses were conducted using IVW, MR-Egger, simple mode, weighted median and weighted mode methods. The slope of each line corresponding to the estimated MR effect per method.



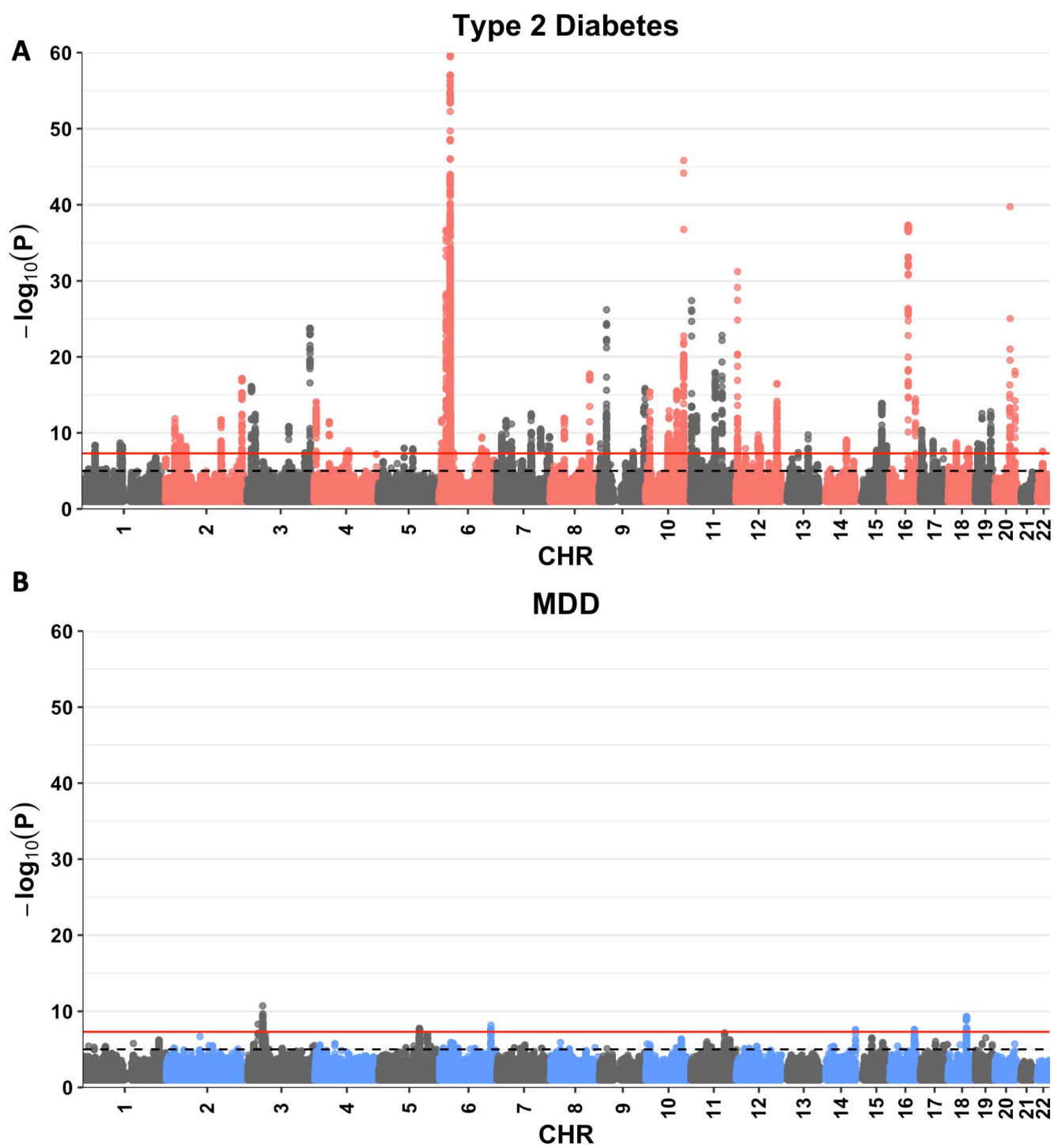
Supplementary Figure 3. Leave-one-out analysis: each row represents a MR analysis of Depression on Type 2 Diabetes using all instruments expect for the SNP listed on the y-axis. The point represents the odds ratio with that SNP removed and the line represents 95% confidence interval.



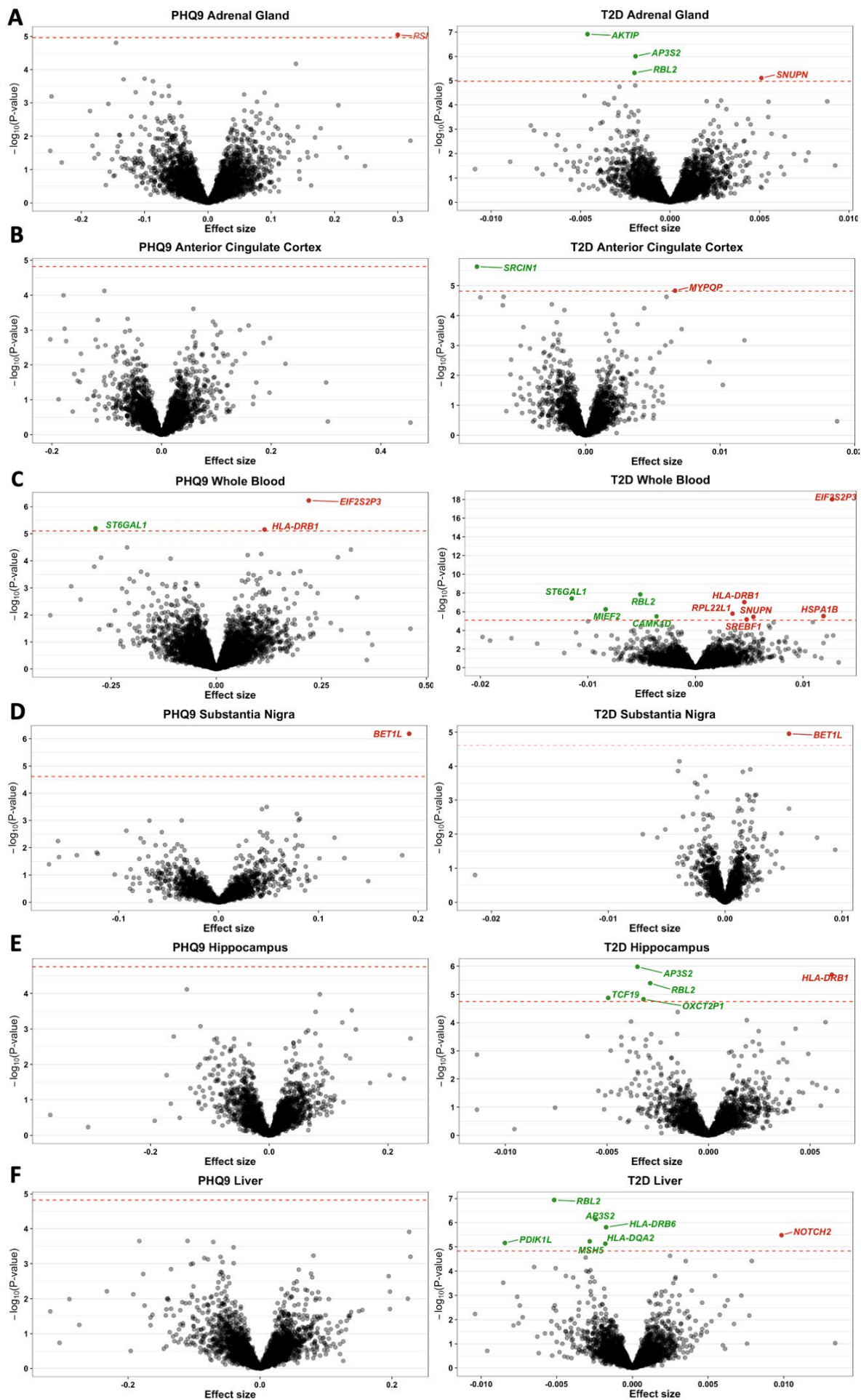
Supplementary Figure 4. Manhattan plots for (A) Type 2 diabetes, (B) PHQ-9 and (C) MDD SP-GWAS in the UK Biobank. The red horizontal line shows genome-wide significance threshold ($P < 5 \times 10^{-8}$). Grey dashed horizontal lines show suggestive genome-wide significance threshold ($P < 1 \times 10^{-5}$)

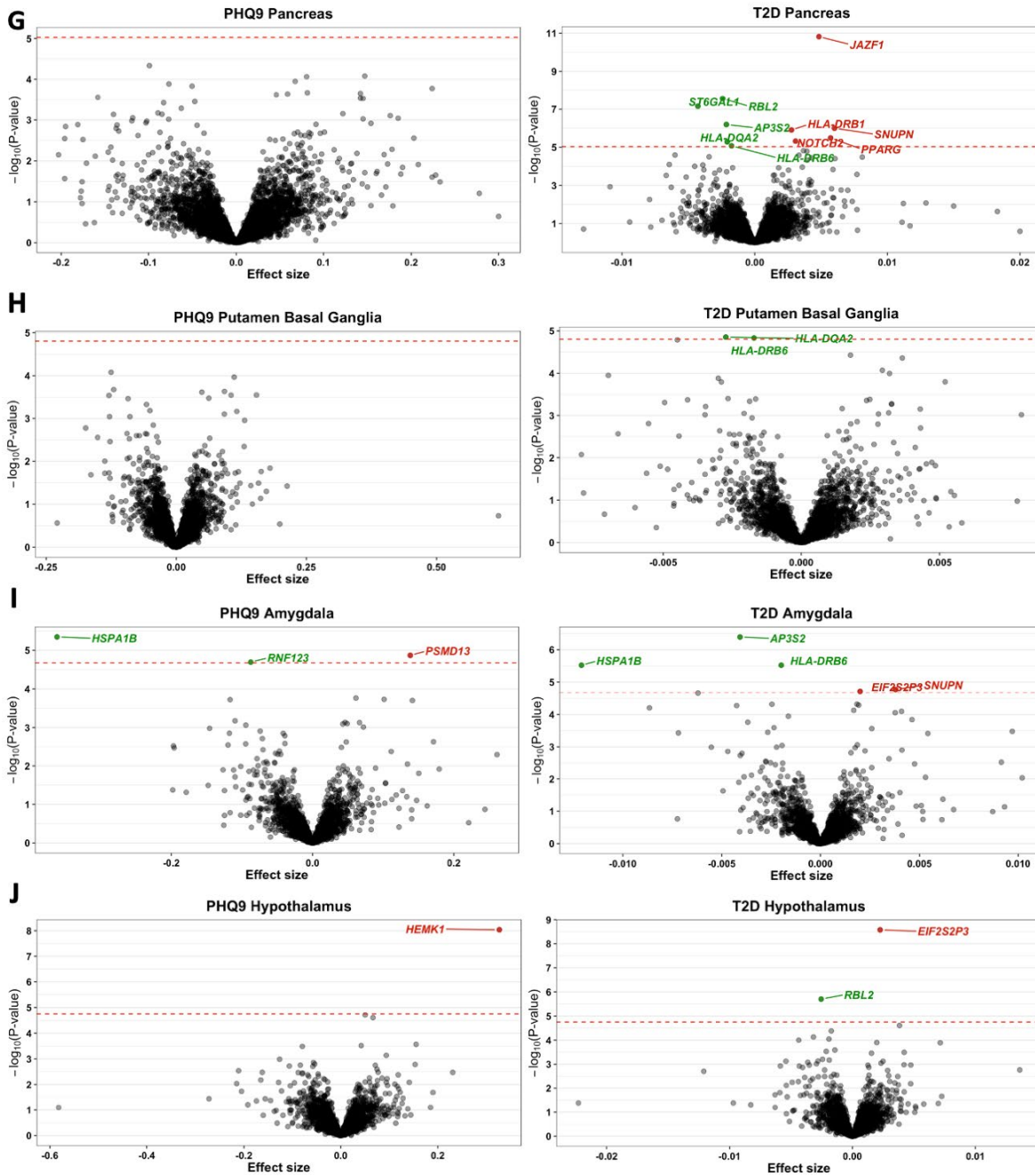


Supplementary Figure 5. Manhattan plots for (A) type 2 diabetes and (B) MDD after MP-GWAS in MTAG. The red horizontal line shows genome-wide significance threshold ($P < 5 \times 10^{-8}$). Grey dashed horizontal lines show suggestive genome-wide significance threshold ($P < 1 \times 10^{-5}$)



Supplementary Figure 6. Manhattan plots for A) type 2 diabetes and B) MDD after MP-GWAS in FinnGen dataset. Red horizontal line shows genome-wide significance threshold ($P < 5 \times 10^{-8}$). Grey dashed horizontal lines show suggestive genome-wide significance threshold ($P < 1 \times 10^{-5}$)





Supplementary Figure 7. Volcano plots of association in tissues implicated in both type 2 diabetes and PHQ-9. Each point on the plot represents an association result for one gene, where the effect size for the association of predicted gene expression and the phenotype of interest is on the x-axis and the $-\log_{10}(\text{p-value})$ of the association is shown on the y-axis. The red dashed line represents a Bonferroni corrected threshold for statistical significance for each tissue analyzed. Statistically significant genes for which predicted gene expression is increased are shown in red, and genes for which predicted gene expression is decreased are shown as green. Statistically insignificant genes are shown in black. A = adrenal gland, B = anterior cingulate cortex, C = whole blood, D = substantia nigra, E = hippocampus, F = liver, G = pancreas, H = putamen basal ganglia, I = amygdala, J = Hypothalamus