利用台灣人體資料庫對遺傳變異、表觀基因組及社會心理層面進行重鬱症之綜合評估

A multifactorial risk assessment of depression integrating genomic, epigenomic, and psychosocial factors in the Taiwanese population

Institute: Department of Public Health, National Taiwan University

Student: Shih-Hsiang Lo **Supervisor**：Yen-Chen Anne Feng Sc.D.

Introduction & Background

Major Depressive Disorder (MDD) is a complex psychiatric illness and a leading cause of global disease burden, especially in Asia.
Although twin studies suggest ~35–40% heritability, SNP-based models explain <10% of the variance.
Therefore, this study integrates genetic, epigenetic, and psychosocial factors to better predict MDD risk in the Taiwanese population.

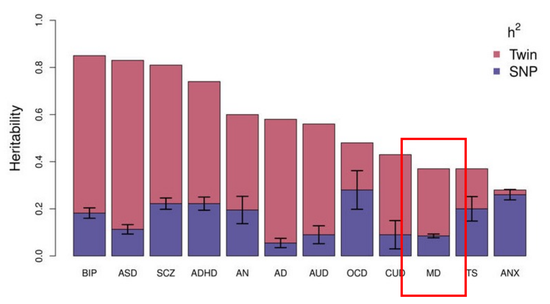
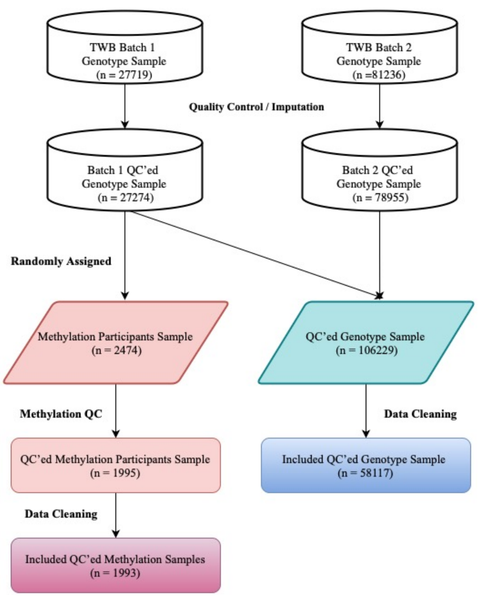


Figure 1. Heritability estimates of MDD (O'Connell et al, 2021)

Study Design & Method

1. Data Source & Sample Selection



Data were obtained from the Taiwan Biobank (TWB), a population-based cohort of Taiwanese adults aged 30–70 years.

After quality control and data cleaning, 58,117 individuals with genotype data and 1,993 individuals with methylation data were included (Figure 2).

Phenotype was defined based on *self-reported physician diagnosis history of depression*, which, while not adhering strictly to DSM criteria, has been adopted as a proxy for MDD.

Figure 2. Flowchart of quality control, data cleaning process and sample size information

2. Genetic and Epigenetic Risk Estimation (PRS and MRS)

To estimate individuals' genetic and epigenetic risk for depression, this study utilized polygenic risk scores (PRS) and methylation risk scores (MRS), respectively. PRS captures inherited genetic predisposition, while MRS reflects epigenetic variation potentially shaped by environmental exposures.

(1) Polygenic Risk Score(PRS)

PRS were constructed using summary statistics from a multi-ancestry genome-wide association study (GWAS) of MDD (Meng et al., 2024), applying both the clumping-and-thresholding (C+T) method via PRSice-2 and the Bayesian shrinkage method via PRS-CS.

Clumping + Thresholding (C+T)
- SNPs selected by LD clumping and p-value thresholding
- Best-fit threshold selected via PRSice-2
- Formula:

$$PRS_i = \sum_{j=1}^{N} \hat{\beta}_j \cdot X_{ij}$$

Bayesian Continuous Shrinkage Prior (PRS-CS):
- SNP effect sizes shrunk using external LD panel
- Posterior weights estimated via PRS-CS

$$\text{Prior:} \quad \beta_j \sim \mathcal{N}\left(0, \frac{\sigma^2}{N}\phi\psi_j\right)$$

$$\text{Posterior effect:} \quad p(\beta_j|\hat{\beta}_j, X)$$

(2) Methylation Risk Score (MRS)

To capture epigenetic contributions to depression risk, methylation risk scores (MRS) were constructed using CpG sites identified from three published epigenome-wide association studies (EWAS) conducted in European or North American populations: Crawford et al. (2018), Starnawska et al. (2019), and Li et al. (2022).

EWAS detect differential DNA methylation patterns associated with disease vulnerability, potentially influenced by environmental exposures.

- CpG sites grouped into co-methylated regions using R package CoMeBack
- Lowest p-value CpG per region retained
- P-value thresholding applied; optimal cutoff selected via correlation with depression
- Final MRS = weighted sum of selected CpG methylation levels

$$MRS_i = \sum_{i=1}^{m} w_i c_i$$

w_i: the estimated effect size or direction of association for CpG site i from EWAS results.

c_i: the DNA methylation level (beta value) of CpG site i for individual i (from 0 - 1).

3. Evaluation of Model Fit and Predictive Performance

To evaluate the combined and individual contributions of genetic, epigenetic, and psychosocial factors to depression risk, stepwise logistic regression models were built.
Genotype and methylation samples were modeled separately, and model fit and predictive ability were assessed.

Genotype Sample (N = 58,117)

- **Model 1:** Age, sex (Male as ref.), marital status (Married as ref.), first 20 PCs (Principal Component of all SNPs), batch effect
- **Model 2:** + Higher education (Yes as ref.)
- **Model 3:** + Monthly income (10K NTD)
- **Model 4:** + Significant interaction terms

- Added PRS (C+T, PRS-CS) to evaluate genetic contribution
- Model fit assessed by Nagelkerke's R^2; prediction evaluated by 5-fold cross-validated AUC

Methylation Sample (N = 1,993)

- **Model 1:** Age, sex, marital status, top 20 PCs
- **Model 2:** + Higher education

- Added PRS (C+T, PRS-CS) to examine genetic effects
- Added MRS (from Crawford, Starnawska, Li EWAS results) to examine epigenetic effects
- Model fit assessed by Nagelkerke's R^2; prediction evaluated by 5-fold cross-validated AUC

Results & Conclusion

Genotype Results:

- Among all genotype-based models, Model 4 (including psychosocial variables, interaction terms, and PRS-CS) demonstrated the best model fit (Nagelkerke $R^2 = 0.0425$) and predictive ability (AUC = 0.6476).
- Adding PRS-CS improved model fit by 9.5%, outperforming PRSice-2 across all models.
- One standard deviation increase in PRS was significantly associated with increased MDD risk (OR = 1.17, 95% CI: 1.12–1.22, $p < 0.0001$).
- Several significant psychosocial factors and interactions were identified

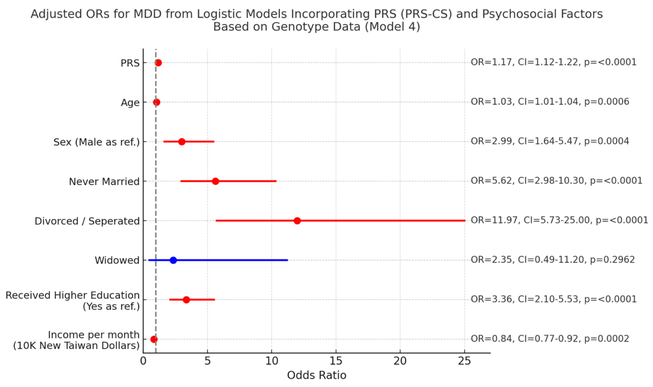


Figure 3. Adjusted ORs of Genotype Model 4 Predictors after adding PRS calculated by PRS-CS

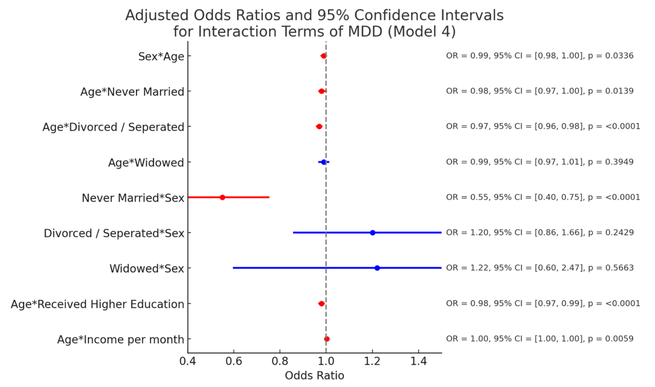


Figure 4. Interaction Terms in Genotype Model 4 after adding PRS calculated by PRS-CS

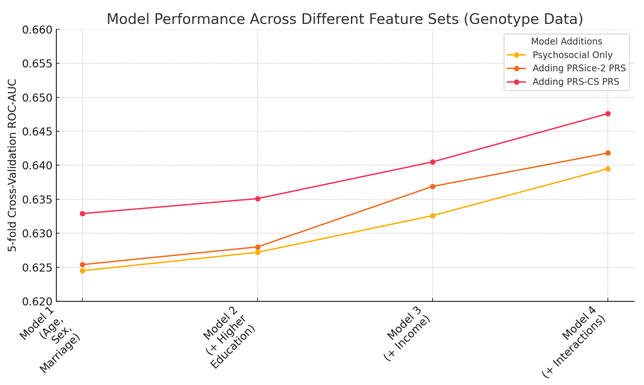


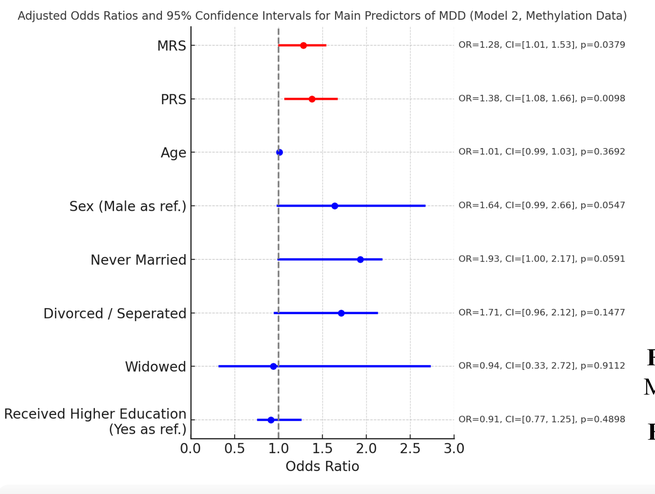
Figure 5. Prediction performance (ROC-AUC) across different models of genotype data

Integrating genetic and psychosocial factors—especially PRS from PRS-CS—substantially improved model fit and depression prediction.

A multifactorial approach (Model 4) yielded the best performance. Results from methylation-based models (MRS) are presented below.

Methylation Results:

- Among MRS models, MRS calculated by Starnawska et al. (2019) EWAS achieved the best fit (Nagelkerke $R^2 = 0.0738$, OR = 1.28, $p = 0.0379$).
- Adding this MRS improved model fit by 9.7%, outperforming other EWAS-based MRS.
- In contrast to genotype data, the C+T method (PRSice-2) outperformed PRS-CS in methylation-based models, improved model fit by 19.3%.
- Model 2 with PRS (PRSice-2) and MRS (Starnawska) achieved the best fit and predictive performance.



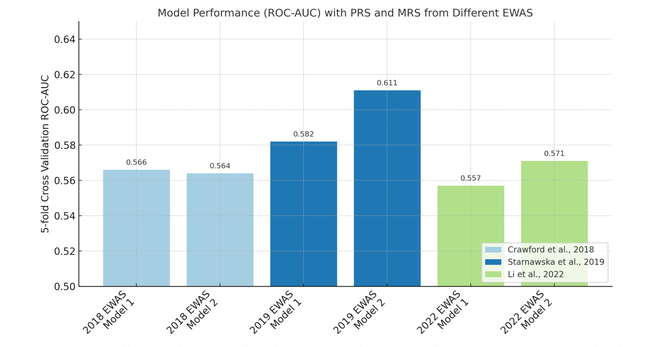


Figure 6. Adjusted ORs for key predictors of MDD in methylation Model 2, adding PRS by PRSice-2 and MRS by Starnawska EWAS

Figure 7. Model ROC-AUC with different EWAS-based MRS + PRS by PRSice-2 combinations

Discussion

This study supports a multifactorial approach to depression prediction by integrating genetic, epigenetic, and psychosocial factors. While PRS-CS yielded the best model fit and AUC in genotype-based data, the simpler C+T method performed better in methylation-based models, suggesting that PRS methods should be flexibly chosen based on data characteristics. The inclusion of MRS, particularly from Starnawska et al. (2019), further improved model fit and predictability. Together, these findings highlight the potential of combining PRS and MRS to enhance model performance in mental health risk prediction.