The Head and the Heart: Decoding Shared Genetic Architecture Between Coronary Artery Disease and Major Depressive Disorder

Koustav Chowdhury[1], Malia Schomaker[2], Lawrence Ong[3]

[1]Indian Statistical Institute, Kolkata, [2]University of Washington, Seattle, [3]University of California, San Diego

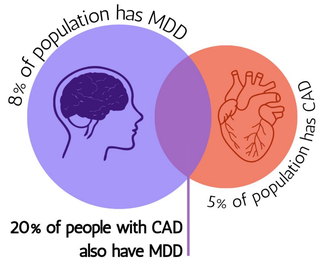
Department of Biostatistics, Yale School of Public Health, New Haven, CT

Yale SCHOOL OF PUBLIC HEALTH

Github Repository

BIGDATA SUMMER IMMERSION

Introduction



Coronary Artery Disease (CAD) and **Major Depression Disorder (MDD)** have a bidirectional relationship in patient care settings.[11]

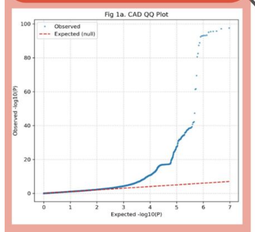
Research Question

Is this bi-directional relationship due to a shared genetic architecture? Or is this relationship simply environmental?

We compared separate CAD and MDD GWAS studies to locate SNPs that significantly affect both diseases, inferred their antagonist or synergistic pleiotropy, and examined their matching genes to make biological connections.

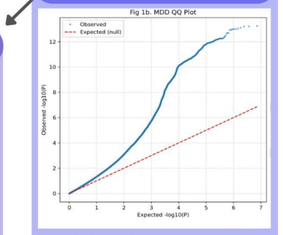
Methods

Coronary Artery Disease Data



184,305 individuals with 60,801 cases, 123,504 controls, and ~9.5 million SNPs.

Major Depressive Disorder Data



674,452 individuals with 166,773 cases, 507,679 controls, and ~7.3 million SNPs

1. LDSC Regression Score

Estimate heritability:

$$\mathbb{E}[\chi_j^2] \approx 1 + \frac{Nh^2}{M} \cdot \ell_j + \alpha$$

Find confounding measure:

The further the α term is from 1, the more sample overlap between studies.

Calculate genetic correlation:

$$\mathbb{E}[Z_j^{(1)} Z_j^{(2)}] = \frac{\sqrt{N_1 N_2}}{M} \cdot \rho_g \cdot \ell_j + \frac{\rho_g + \rho_e}{\sqrt{N_1 N_2}} \cdot N_s$$

2. LD Clumping

Select representative SNPs with the lowest p-value within a 250 kb window in LD ($r^2 \geqslant 0.1$) with other SNPs (p1: 1e-05 , p2: 1)

3. Effect Models

Compute **pooled estimates** from the two studies to create a comprehensive **additive genetic model:**

$$\mathbf{Y} = \mathbf{W}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Pooled estimates calculated as a weighted average of the effect estimates within the individual studies

Compare all SNPs at the significance level: **p-val < 5e-8**.

4. PLACO

Pleiotropic Analysis under Composite Null Hypothesis:

Test for genetic pleiotropy of representative SNPs for CAD and MDD. IDed SNPs where $FDR \leqslant 0.1$

$$H_0 : \beta_1\beta_2 = 0 \qquad H_1 : \beta_1\beta_2 \neq 0$$

$$\text{Estimated } p\text{-value} = F\left(\frac{z_1 z_2}{\sqrt{\text{Var}(Z_1)}}\right) + F\left(\frac{z_1 z_2}{\sqrt{\text{Var}(Z_2)}}\right) - F(z_1 z_2)$$

3a. Fixed

Assumes one true effect size (μ) exists with some error (ϵi).

$$\hat{\beta}_i = \mu + \epsilon_i, \quad \text{where } \epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$$

$$\hat{\mu} = \frac{\sum_{i=1}^k w_i \hat{\beta}_i}{\sum_{i=1}^k w_i} \quad \text{where} \quad w_i = \frac{1}{\sigma_i^2}$$

3b. Random

Assumes that the true effect size varies across studies and follows a distribution centered at a mean μ, with between-study variability τ^2

$$\hat{\beta}_i = \beta_i + \epsilon_i, \quad \text{where } \begin{cases} \epsilon_i \sim \mathcal{N}(0, \sigma_i^2) \\ \beta_i \sim \mathcal{N}(\mu, \tau^2) \end{cases}$$

$$\hat{\mu} = \frac{\sum_{i=1}^k w_i^* \hat{\beta}_i}{\sum_{i=1}^k w_i^*} \quad \text{where} \quad w_i^* = \frac{1}{\sigma_i^2 + \tau^2}$$

Significant SNPs to Investigate

The Additive Genetic Model

Let $\mathbf{y_1}=\mathbf{X\beta+\delta}$, $\mathbf{y_2}=\mathbf{Z\gamma+\varepsilon}$ where $\mathbf{y}_1$ and $\mathbf{y}_2$ are the vectors of standardized (possibly different) phenotypes for two studies, where $\mathbf{X}_{(N1\times M)}$ and $\mathbf{Z}_{(N2\times M)}$ are the genotype matrices consisting of the number of effect alleles per SNP, whose columns are also standardized, β and γ are the vector of effect sizes. We assume the following:

$$\mathbb{E}\begin{pmatrix}\beta\\ \gamma\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}, \text{Var}\begin{pmatrix}\beta\\ \gamma\end{pmatrix} = \frac{1}{M}\begin{pmatrix}h_1^2 I & \rho_g I\\ \rho_g I & h_2^2 I\end{pmatrix}, \mathbb{E}\begin{pmatrix}\delta\\ \epsilon\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}, \text{Var}\begin{pmatrix}\delta\\ \epsilon\end{pmatrix} = \begin{pmatrix}(1-h_1^2)I & \rho_e I\\ \rho_e I & (1-h_2^2)I\end{pmatrix}$$

where, $\mathbf{N}_1$,$\mathbf{N}_2$ are the # of individuals in each study, M = number of SNPs, $\mathbf{h}_1^2$, $\mathbf{h}_2^2$ are the heritabilities of the corresponding phenotypes, $\boldsymbol{\varrho}_g$ is the genetic covariance and $\mathbf{N}_s$ is the number of common individuals between the studies.[4]

Results

❑ **Figure 2: Miami Plot of CAD (top) and MDD (bottom) SNPs**

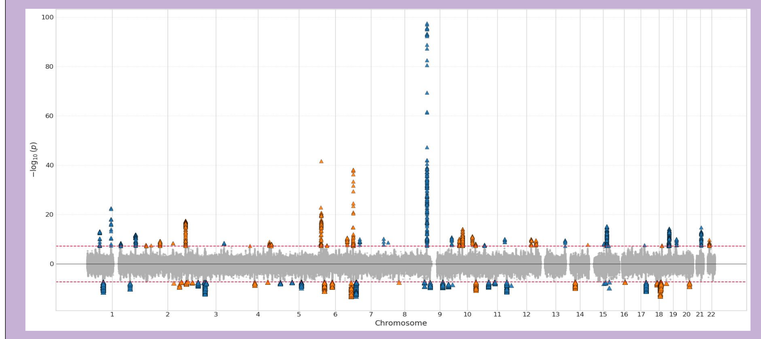


Fig 2. The top half shows results from the CAD dataset (2,213 significant SNPs), and the bottom half shows MDD results (3,189 significant SNPs). Each point represents a SNP plotted by chromosomal position (x-axis) and $-\log_{10}$(p-value) (y-axis). Significant SNPs have a p-value < 5e-8 and are colored. Insignificant SNPs are grey.

❑ **Table 1. LDSC Cross Trait Analysis Results**

	CAD	MDD
Observed h²	0.0597 (0.0044)	0.0366 (0.0018)
Lambda GC	1.0466	1.4423
Mean χ²	1.1249	1.5679
Intercept	0.9114 (0.0079)	1.0779 (0.0113)
Ratio	< 0	0.1371 (0.0199)

Genetic Covariance	
Total Observed scale gencov	**0.0026 (0.0009)**
Mean Z1*Z2	**0.0292**
Intercept	**0.0103 (0.004)**
Genetic Correlation	
Genetic Correlation	**0.0549 (0.0189)**
Z-score	**2.9066**
P	**0.0037**

Table 1. CAD and MDD cross-trait analysis using LDSC. The mean chi-sq is larger than 1 indicating an inflation in GWAS statistics; the intercept is close to 1, indicating minimal confounding bias. Both datasets show significant non-zero heritability and have positive correlation (r_g=0.0549; p < 0.05) indicating the comorbidity of both the diseases.

❑ **Figure 3: Bubble plot of Fixed, Random, and PLACO Combined Results**

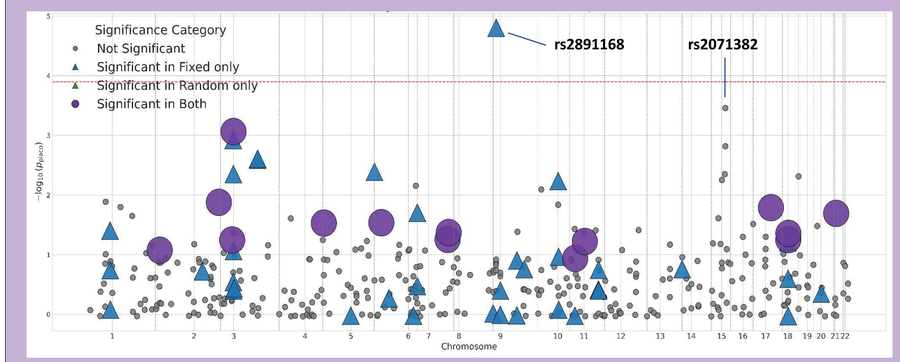


Fig 3. The red line is the Bonferroni Correction significance threshold (1.3e-04). The SNPs above the red line are significant for PLACO. The 51 blue triangles are SNPs that are significant in Fixed Effects Model ; the 14 purple circles are SNPs that are significant in both Fixed Effects and Random Effects models. There are no SNPs that are significant only in the Random Effects model.

Conclusions

Using PLACO[13], we identified two significant SNPs with a false discovery rate (FDR) at 10%:

$$FDR \leqslant 0.1$$

rs2891168 (G/A)*	**rs2071382** (T/C)
Linked Gene: **CDKN2B-AS1**	Linked Gene: **FES, FURIN**
Represents: 47 SNPs	Represents: 16 SNPs
Pooled β: 0.100 ± 0.093	Pooled β: 0.014 ± 0.039
Increased CAD risk **(+)**	Increased CAD risk **(+)**
Increased MDD risk **(+)**	Decreased MDD risk **(-)**

**Also significant under Bonferroni Correction*

Prior research links these genes to increased CAD risk[3], but evidence for involvement in MDD is limited. CDKN2B-AS1 is associated with myocardial infarction and co-occurring psychiatric conditions[8], while FES and FURIN have been linked to bipolar disorder and schizophrenia[9].

Limitations

- ❑ The estimated effect sizes are based on marginal models.
- ❑ Cross-trait analysis is performed on clumped data due to limitations of computational power when using the full datasets.
- ❑ Entire MDD Dataset not included due to data privacy restrictions.
- ❑ Random and Fixed Effect Models only detects unidirectional effect sizes of SNPs.

Acknowledgements

Special thanks to Dr. Hongyu Zhao for initializing this project, our GSIs Leqi Xu and Jiaqi Hu for their knowledge and mentorship, and Dr. Xiang Zhou for guiding us through the finish line. And thank you to the BDSY program team, especially Dr. Bhramar Mukherjee and Jackson Higginbottom.

References

1. Als, T. D., et. al. (2023). Depression pathophysiology, risk prediction of recurrence and comorbid psychiatric disorders using genome-wide analyses. Nature Medicine, 29(7), 1832–1844. https://doi.org/10.1038/s41591-023-02552-1
2. Baltramonaityte, V., et. al. (2023). A multivariate genome-wide association study of psycho-cardiometabolic multimorbidity. PLOS Genetics, 19(6), e1010508. https://doi.org/10.1371/journal.pgen.1010508
3. Broadbent, H. M., et. al. (2008). Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked SNPs in the ANRIL locus on chromosome 9p. Human Molecular Genetics, 17(6), 806–814. https://doi.org/10.1093/hmg/ddm352
4. Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P. R., ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, Duncan, L., Perry, J. R., Patterson, N., Robinson, E. B., Daly, M. J., Price, A. L., & Neale, B. M. (2015). An atlas of genetic correlations across human diseases and traits. Nature genetics, 47(11), 1236–1241. https://doi.org/10.1038/ng.3406
5. Chang CC, et. al. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience, 4.
6. Gaunt T, Rodríguez S, Day I (2007) Cubic exact solutions for the estimation of pairwise haplotype frequencies: implications for linkage disequilibrium analyses and a web tool 'CubeX'. BMC Bioinformatics, 8.
7. GISSI Prevenzione Investigators and SIBioC-GISSI Prevenzione Group, Gori, F., et. al. (2010). Common genetic variants on chromosome 9p21 are associated with myocardial infarction and type 2 diabetes in an Italian population. BMC Medical Genetics, 11(1), 60. https://doi.org/10.1186/1471-2350-11-60
8. Gori, F., Specchia, C., Pietri, S. et al. Common genetic variants on chromosome 9p21 are associated with myocardial infarction and type 2 diabetes in an Italian population. BMC Med Genet 11, 60 (2010). https://doi.org/10.1186/1471-2350-11-60
9. Hindley, G., et. al. (2021). Characterising the shared genetic determinants of bipolar disorder, schizophrenia and risk-taking. Translational Psychiatry, 11(1), 466. https://doi.org/10.1038/s41398-021-01576-4
10. Karamanavi, E., et. al. (2022). The fes gene at the 15q26 coronary-artery-disease locus inhibits atherosclerosis. Circulation Research, 131(12), 1004–1017. https://doi.org/10.1161/CIRCRESAHA.122.321146
11. Khawaja, Imran Shuja, et al. "Depression and Coronary Artery Disease." Psychiatry (Edgmont), vol. 6, no. 1, Jan. 2009, pp. 38–51. PubMed Central, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2719442/.
12. Purcell S, et al. (2007) PLINK: A Tool Set for Whole-Genome and Population-Based Linkage Analyses. American Journal of Human Genetics, 81.
13. Ray, D., & Chatterjee, N. (2020). A powerful method for pleiotropic analysis under composite null hypothesis identifies novel shared loci between Type 2 Diabetes and Prostate Cancer. PLOS Genetics, 16(12), e1009218. https://doi.org/10.1371/journal.pgen.1009218
14. https://www.ebi.ac.uk/gwas/
15. https://www.ncbi.nlm.nih.gov/gene/