

Investigation of Causal Relationship between Plasma Metabolites and Bipolar Disorder: A Two-Sample Mendelian Randomization Approach

Muchi Tang^{1*}

¹ Shenyang Pharmaceutical University, Liaoning, Shenyang, China

Email Address

t13080780425@163.com (Muchi Tang)

*Correspondence: t13080780425@163.com

Received: 23 March 2024; Accepted: 16 April 2024; Published: 18 May 2024

Abstract:

This study employed a two-sample Mendelian randomization (MR) approach to investigate the causal relationship between plasma metabolites and bipolar disorder using large-scale genome-wide association studies (GWAS). The study design was based on the three key assumptions of MR: stable association between genetic variants and exposure, absence of confounding factors related to genetic variants, and full mediation of genetic variants' effects on the outcome through the exposure. The study adhered to the guidelines for MR research. Data sources included the plasma metabolite study by Chen et al. (2023) and the bipolar disorder GWAS analysis by Hou et al. (2016). R software and relevant packages were used for data processing and analysis, employing methods such as inverse-variance weighted (IVW), weighted median (WM), simple mode (SM), weighted mode (WM), and MR-Egger regression. The results identified 17 plasma metabolites significantly associated with bipolar disorder.

Keywords:

Mendelian Randomization, Bipolar Disorder, Plasma Metabolites, Genome-Wide Association Studies (GWAS), Instrumental Variable Selection

1. Introduction

Bipolar Disorder (BD) is a severe mental illness characterized by recurrent episodes of depression and mania or hypomania. The disease typically begins between the ages of 15 and 25, with depressive symptoms being the most common initial presentation. Patients spend approximately 75% of their time in a depressive state. Early diagnosis and treatment can significantly improve the prognosis, but in reality, it takes an average of about 9 years after the first depressive symptoms appear to receive the correct diagnosis and treatment. Bipolar Disorder not only reduces the life expectancy of patients (by about 12 to 14 years), but the mortality rate from cardiovascular diseases is also 1.6 to 2 times higher than that of the general population, occurring on

average 17 years earlier. Additionally, the prevalence of metabolic syndrome, obesity, smoking, and type 2 diabetes in patients with Bipolar Disorder is higher than that of the general population, all of which are risk factors for early mortality. The suicide rate in Bipolar Disorder is about 0.9%, which is much higher than the 0.014% in the general population, with 15% to 20% of patients ending their lives by suicide [1]. The genetic predisposition to Bipolar Disorder is quite obvious, affecting about 50 million people worldwide. In recent years, with the advancement of technology and methodology, especially the establishment of international alliances and large population gene banks, we have gained a deeper understanding of the genetic etiology of Bipolar Disorder. Studies have shown that Bipolar Disorder is not only highly heterogeneous and polygenic in its genetic makeup but also has significant genetic overlap with other mental disorders [2].

Metabolites are substances produced and utilized during the process of metabolism in organisms, microorganisms, and external substances [3]. Metabolomics is a widely used technique for analyzing the metabolites present in biological fluids, cells, and tissues, and is frequently employed in the identification of biomarkers [4]. By leveraging the metabolites obtained from human studies, we can establish connections between diseases and metabolic pathways, providing new insights and possibilities for early diagnosis and potential treatment of diseases [5].

However, a major challenge in the identification and validation of metabolites is the variability in metabolite levels between individuals. This variability is caused by differences in genetic factors and environmental exposures, making validation of biomarkers difficult [6]. To address this issue, the Mendelian Randomization (MR) method has emerged. MR utilizes genetic variations as instrumental variables (IVs) to assess causal relationships between exposure factors and outcomes. By leveraging summary data from genome-wide association studies (GWAS), MR provides an alternative approach to determine causal relationships in observational studies [7]. Recently, GWAS has extended its scope to include metabolomics, offering new opportunities to investigate causal relationships between metabolites and disease outcomes.

In this study, we employed a Two-Sample Mendelian Randomization (TSMR) approach to investigate the causal relationship between plasma metabolites and Bipolar Disorder (BD). Through a bidirectional study design, we aim to explore the impact of metabolites on BD and the influence of BD on metabolites, gaining deeper insights into this complex mental disorder. Our research will provide new perspectives and possibilities for early diagnosis and potential treatment of Bipolar Disorder. By employing the Two-Sample Mendelian Randomization method, we will provide more reliable evidence for the causal relationship between metabolites and BD, guiding future clinical practice and personalized therapies.

2. Data and Methods

2.1. Study Design

A two-sample Mendelian randomization (MR) approach was employed in this experiment to investigate the relationship between exposure (plasma metabolites) and outcome (Bipolar disorder), initially utilizing large-scale genome-wide association studies (GWAS). Subsequently, reverse MR analysis was conducted to explore the relationship between exposure (Bipolar disorder) and outcome (plasma metabolites). The study design was based on three key assumptions of MR [8]: (1) genetic variants

should exhibit robust associations with the exposure; (2) genetic variants should not be associated with any confounding factors; (3) the influence of genetic variants on the outcome should be fully mediated by the exposure, excluding any relevance to alternative pathways. This study adhered to the guidelines outlined for MR studies [9].

2.2. Data Source

This study utilized data from multiple sources, the overview of which is provided in Table 1. The plasma metabolite data was obtained from the study by Chen et al. [10], which involved a cohort of approximately 8,000 individuals of European ancestry. Summary statistics from the genome-wide association study (GWAS) on plasma metabolites can be accessed through the GWAS catalog (<https://www.ebi.ac.uk/gwas/>). The access IDs for these GWAS data range from GCST90199621 to GCST90201020. The bipolar disorder data required for the MR analysis was derived from the GWAS analysis conducted by Hou et al. (<https://gwas.mrcieu.ac.uk/>, Dataset: ebi-a-GCST003724). This analysis included 34,950 individuals of European ancestry, comprising 7,647 individuals with bipolar disorder and 27,303 individuals without bipolar disorder. Through genetic analysis, a total of 9,483,147 single nucleotide polymorphisms (SNPs) associated with bipolar disorder were identified[11]. Data numbers and related information are shown in Table 1.

Table 1. The comprehensive delineation of data sources employed in this multifaceted bidirectional Mendelian randomization investigation.

Variables	PubMed ID	Journal	Study accession	Year of publication	Sample size
Plasma metabolites	36635386	Nature Genetics	GCST90199621 to GCST90201020	2023	Approximately 8000 individuals of European descent
Bipolar disorder	27329760	Human Molecular Genetics	ebi-a-GCST003724	2016	34950 individuals of European descent, comprising 7,647 individuals with Bipolar disorder 27,303 individuals without Bipolar disorder

2.3. Instrumental Variable Selection

The instrumental variable (IV) selection process primarily consists of the following steps:

- (1) Screening of IVs associated with the exposure by extracting SNPs with a P-value < 1×10^{−5} from the GWAS data on Plasma metabolites.
- (2) Removal of SNPs exhibiting linkage disequilibrium within the aforementioned IVs, using a threshold of r² < 0.001 and kb < 10000.
- (3) Extraction of the corresponding IVs from the GWAS data on bipolar disorder outcome (9,483,147 SNPs), excluding palindromic SNPs and removing IVs related to the outcome.
- (4) Calculation of F statistics to assess the strength of the obtained SNPs, and exclusion of weak IVs with F < 10. The formula for calculating F statistics is as follows [12]:

$$F = \frac{R^2 \times (N - K - 1)}{K(1 - R^2)}$$

(1)

$$R^2 = \frac{2 \times \text{EAF} \times (1 - \text{EAF}) \times \beta^2}{2 \times \text{EAF} \times (1 - \text{EAF}) \times N \times \text{SE}^2 + 2 \times \text{EAF} \times (1 - \text{EAF}) \times \beta^2} \quad (2)$$

In this formula:

N represents the sample size of the exposure factor,

K represents the number of instrumental variables,

R^2 represents the proportion of variance in the exposure factor explained by the instrumental variables,

EAF represents the effect allele frequency,

β represents the effect size, and

SE represents the standard error of the effect size.

2.4. Mendelian Randomization Analysis and Result Presentation

This study utilized R software (version 4.3.2) along with the TwoSampleMR package (version 0.5.11), as well as VariantAnnotation, gwasglue, and TwoSampleMR packages [13], for data processing and analysis. To further assess the reliability and robustness of the results, heterogeneity analysis and multiple testing were conducted.

The analysis primarily employed methods such as inverse-variance weighted (IVW), weighted median (WM), Simple mode (SM), Weighted mode (WM), and MR-Egger regression (MR-Egger). The IVW results were used as the main indicator for evaluating causal effects, while the other methods were used to assist in assessing the consistency of effect sizes (β values) with the IVW, indicating the robustness of the results. The IVW method constrained the regression intercept to zero under the assumption that all SNPs were valid and independent instrumental variables, and it used the inverse of the outcome variance as weights for fitting. When each genetic variant satisfied the assumptions of instrumental variables, the IVW method provided the most accurate results and was considered the gold standard for evaluating causal effects [14]. Additionally, FDR correction was applied to the IVW results for multiple testing, considering a statistically significant causal relationship between metabolites and bipolar disorder when the FDR-corrected p-value was less than 0.05. In order to visualize the results more effectively, this study employed scatter plots, forest plots, and funnel plots. The scatter plots were used to display the distribution of Mendelian randomization results (Figure 3, Figure 4, Figure 5). The forest plots provided a visual representation of the impact of each instrumental variable on the outcome (Figure 2). To further evaluate the stability of the findings, a leave-one-out sensitivity analysis was conducted, and corresponding charts were generated (Figure 6, Figure 7, Figure 8). The leave-one-out sensitivity analysis allows for the assessment of the influence of each observation on the results, thereby validating their stability.

2.5. Heterogeneity and Multieffect Analysis

In order to assess the consistency and stability of the results, heterogeneity analysis and multiple testing were conducted. Heterogeneity was evaluated using Cochran's Q test [15], while multiple testing was assessed using the MR-PRESSO test and the intercept term of MR-Egger regression. A p-value greater than 0.05 indicated the absence of heterogeneity or multiple effects [16]. To eliminate the influence of significant outliers on the results, MR-PRESSO analysis was performed. This involved fitting a regression model to identify potential horizontal pleiotropic outliers

and exclude them from the analysis. This step helped to mitigate the interference of outliers on the MR estimation results, improving the accuracy and reliability of the analysis. A p-value below 0.05 was considered statistically significant. Data filtering and arrangement were performed using software packages such as reshape2, circlize, and ComplexHeatmap. A circular plot was generated to visualize the p-values of different methods used to assess causal effects (Figure 1).

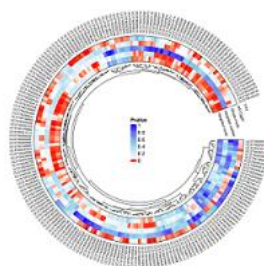


Figure 1. The association between plasma metabolites and Bipolar disorder was first investigated. A total of 375 unique plasma metabolites were preliminarily identified as significantly associated with Bipolar disorder.

3. Results

3.1. Instrumental Variables

Among the 1400 metabolites considered, there were a total of 34843 SNPs associated with a P-value less than 1×10^{-5} . After conducting linkage disequilibrium analysis and excluding confounding SNPs, a subset of 29302 SNPs was included for further analysis. Among these SNPs, 11815 were found in the GWAS data for coronary heart disease, resulting in a site coverage rate of 40.32%. The median number of instrumental variables corresponding to each metabolite was 10.

3.2. Mendelian Randomization Analysis Results

The results obtained through inverse-variance weighted (IVW), weighted median (WM), Simple mode (SM), weighted mode (WM), and MR-Egger regression (MR-Egger) with p-values less than 0.05 initially identified 375 unique plasma metabolites significantly associated with bipolar disorder, as shown in Figure 1. Further analysis focused only on the results selected by IVW ($p < 0.05$), while excluding instrumental variables (IVs) with horizontal pleiotropy. Finally, 17 plasma metabolites were identified, including Linolenate [α or γ ; (18:3n3 or 6)] levels; Arachidate (20:0) levels; N-lactoyl valine levels; N-methyl-2-pyridone-5-carboxamide levels; X-15486 levels; Ribitol levels; Quinolinate levels; Nisinate (24:6n3) levels; 1-palmitoyl-2-docosaheptaenoyl-gpc (16:0/22:6) levels; X-17690 levels; 1-stearoyl-2-arachidonoyl-GPI (18:0/20:4) levels; Bilirubin (E,Z or Z,E) levels; Bilirubin (Z,Z) to taurocholate ratio; Phosphate to serine ratio; Pristanate levels; Dihomo-linolenate (20:3n3 or n6) levels; 1-palmitoleoyl-2-linolenoyl-GPC (16:1/18:3) levels. These plasma metabolites were significantly associated with the occurrence of bipolar disorder.

Based on the primary criterion of PIVW < 0.05 , 17 metabolites were found to have a causal association with bipolar disorder. Among them, 11 plasma metabolites were negatively correlated with the risk of bipolar disorder, while the remaining 6 plasma metabolites were positively correlated with the risk. The metabolites with their corresponding GCST IDs and associations are as follows: GCST90199748:

Linolenate [alpha or gamma; (18:3n3 or 6)] levels; GCST90200446: Arachidate (20:0) levels; GCST90200200: N-lactoyl valine levels; GCST90199778: N-methyl-2-pyridone-5-carboxamide levels; GCST90200533: X-15486 levels; GCST90199642: Ribitol levels; GCST90200348: Quinolinatate levels; GCST90200122: Nisinate (24:6n3) levels; GCST90200043: 1-palmitoyl-2-docosaheptaenoyl-gpc (16:0/22:6) levels; GCST90200555: X-17690 levels; GCST90199649: 1-stearoyl-2-arachidonoyl-GPI (18:0/20:4) levels. On the other hand, the following metabolites were associated with an increased risk of bipolar disorder: GCST90200687: Bilirubin (E,Z or Z,E) levels; GCST90200829: Bilirubin (Z,Z) to taurocholate ratio; GCST90200863: Phosphate to serine ratio; GCST90200409: Pristanate levels; GCST90199820: Dihomo-linolenate (20:3n3 or n6) levels; GCST90200102: 1-palmitoleoyl-2-linolenoyl-GPC (16:1/18:3) levels. Figure 2 illustrates the strongest protective metabolite with a causal effect being N-lactoyl valine levels (OR=0.58671103, 95% CI: 0.420 to 0.819, PIVW=0.002, PFDR=0.161419827), followed by Linolenate [alpha or gamma; (18:3n3 or 6)] levels (OR=0.589219083, 95% CI: 0.435 to 0.797, PIVW<0.001, PFDR=0.085761011), and N-methyl-2-pyridone-5-carboxamide levels (OR=0.604004939, 95% CI: 0.441 to 0.827, PIVW=0.002, PFDR=0.161419827). The most significant risk-related metabolite was Pristanate levels (OR=1.527837815, 95% CI: 1.162 to 2.009, PIVW=0.002, PFDR=0.193613074), followed by Dihomo-linolenate (20:3n3 or n6) (OR=1.50451743, 95% CI: 1.163 to 1.946, PIVW=0.002, PFDR=0.162221316). Among the other analysis methods, the various MR analysis methods generally exhibited consistent directions with the IVW results (Figure 2), indicating reliable findings.



Figure 2. Causal relationship between plasma metabolites (exposures) and bipolar affective disorder (outcome).

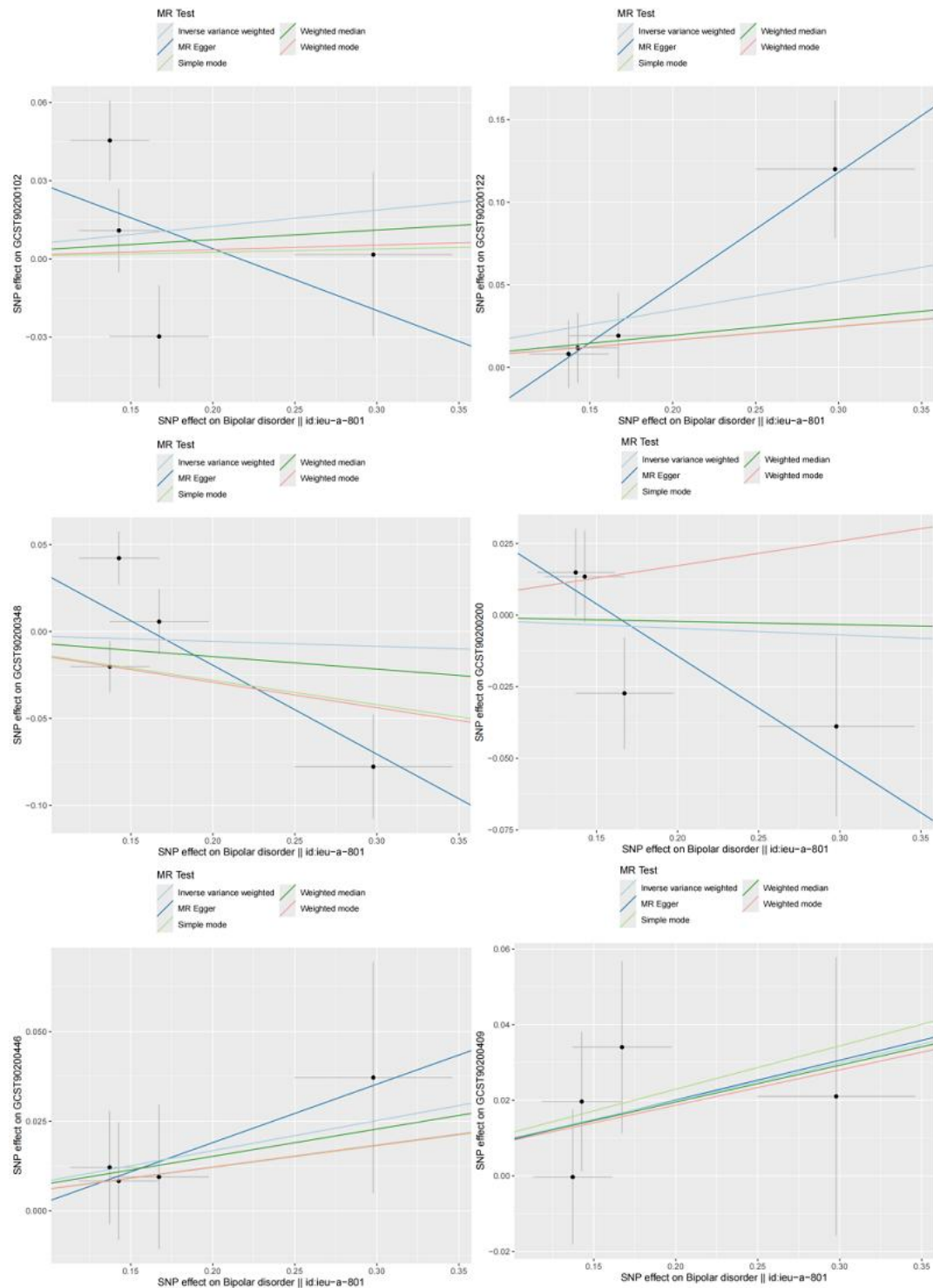


Figure 3. MR scatter plot of 17 known metabolites with risk of bipolar disorder (GCST90200102; GCST90200446; GCST90200122; GCST90200348; GCST90200200; GCST90200409).

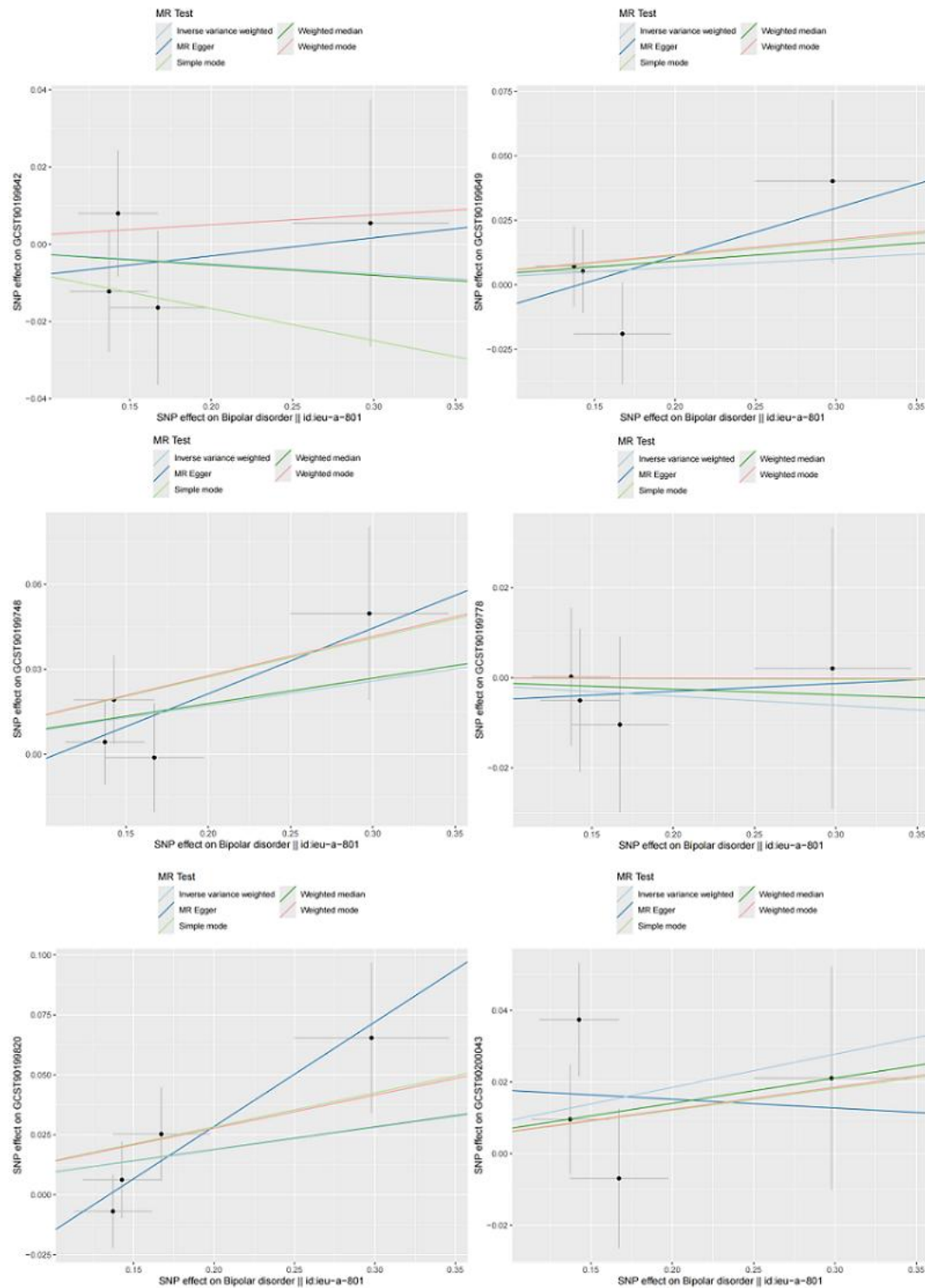


Figure 4. MR scatter plot of 17 known metabolites with risk of bipolar disorder (GCST90199649;GCST90199642; GCST90199748; GCST90200043; GCST90199778; GCST90199820).

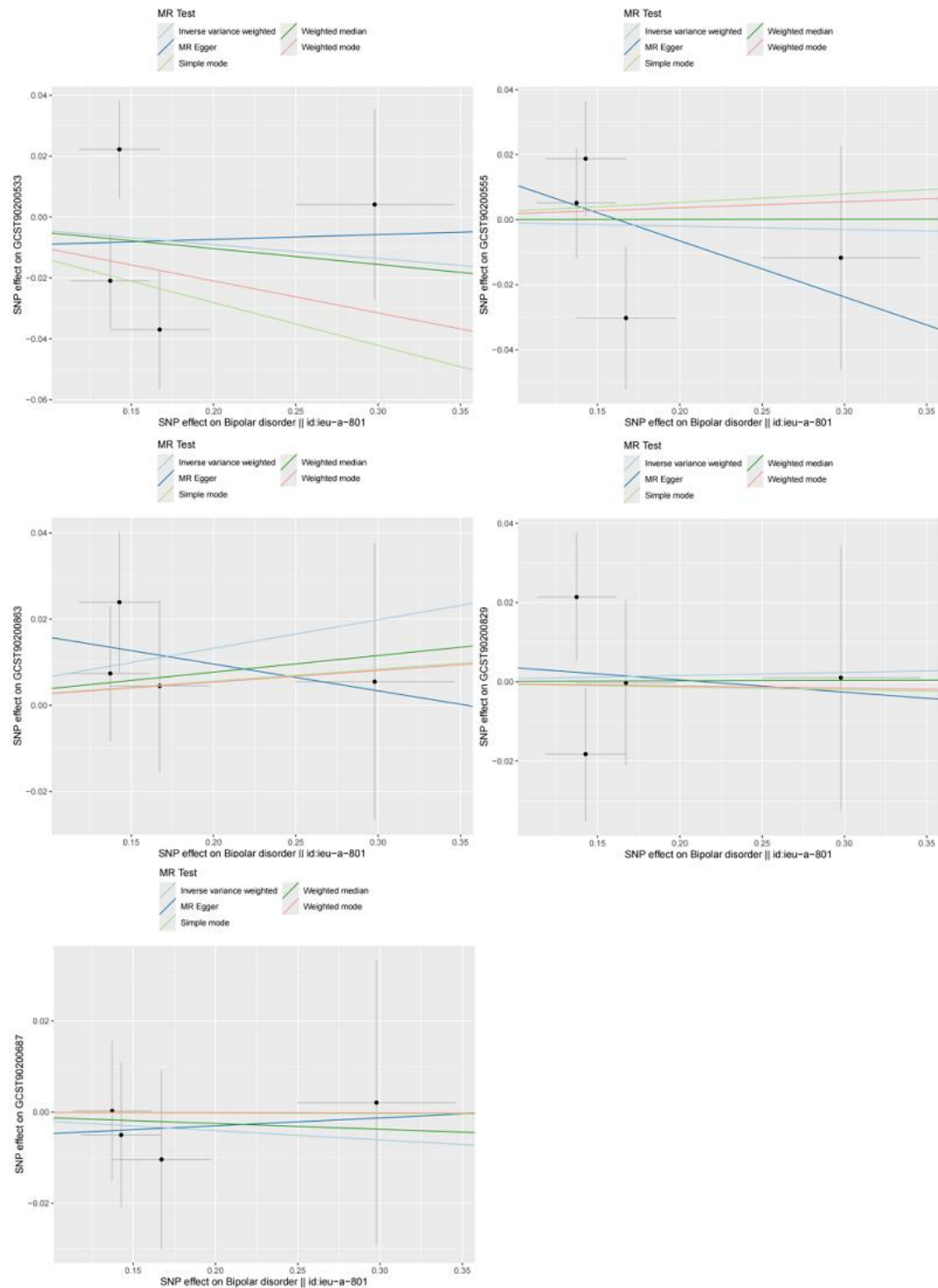


Figure 5. MR scatter plot of 17 known metabolites with risk of bipolar disorder (GCST90200533; GCST90200687; GCST90200555; GCST90200863; GCST90200829).

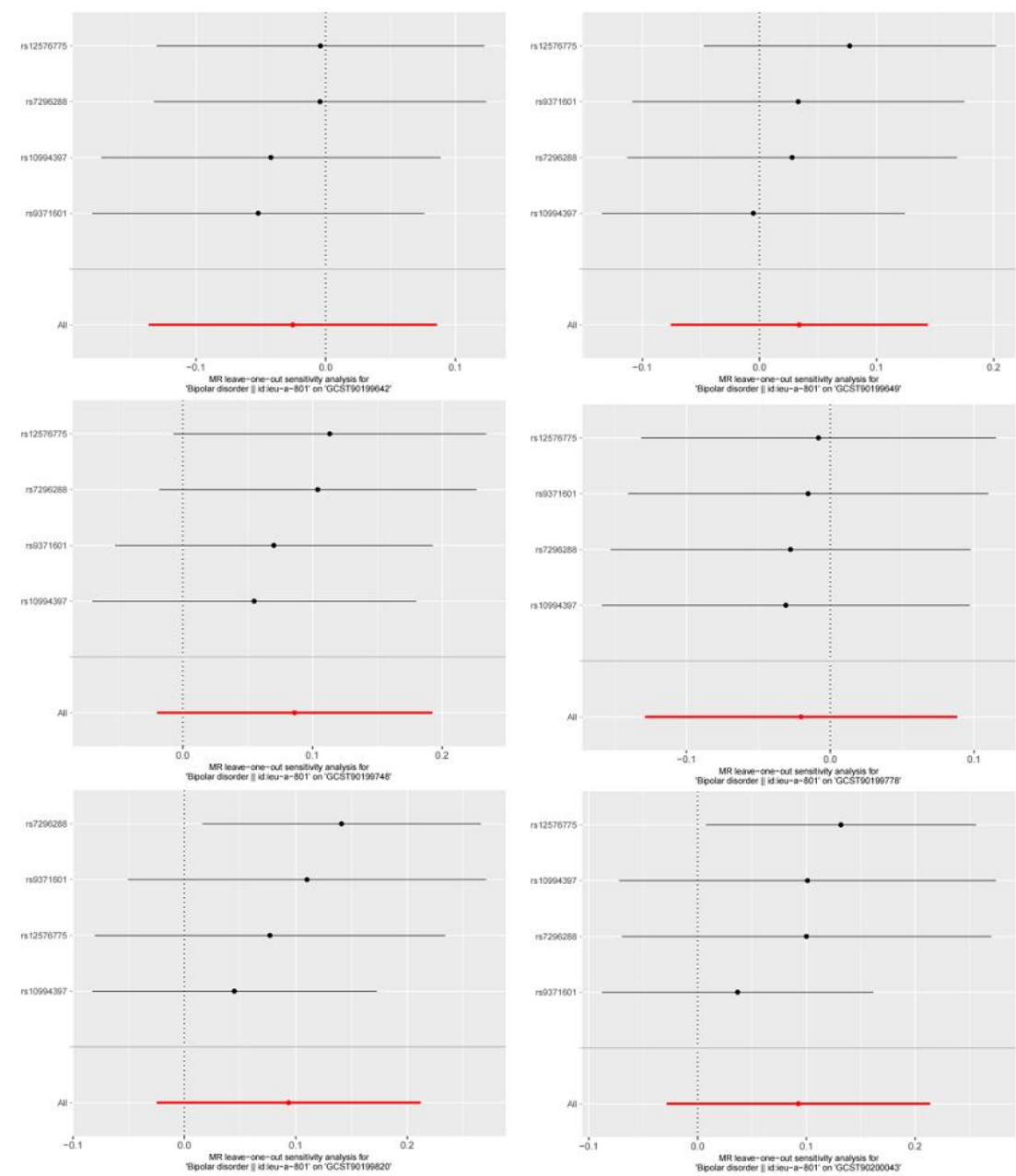


Figure 6. Results of leave-one-out sensitivity analysis of metabolites and bipolar disorder risk (GCST90199649; GCST90199642; GCST90199748; GCST90200043; GCST90199778; GCST90199820).

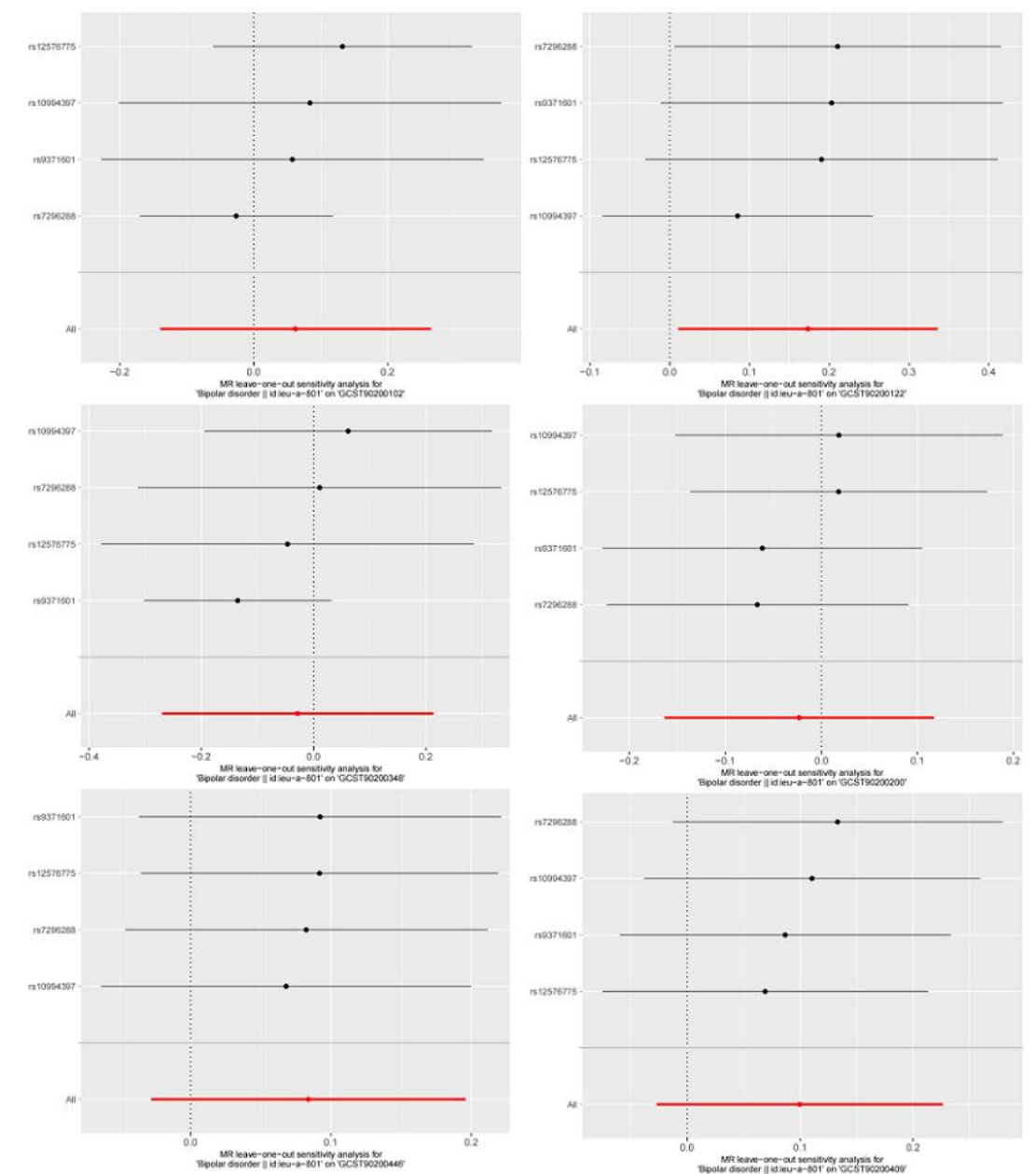


Figure 7. Results of leave-one-out sensitivity analysis of metabolites and bipolar disorder risk (GCST90200102; GCST90200446; GCST90200122; GCST90200348; GCST90200200; GCST90200409).

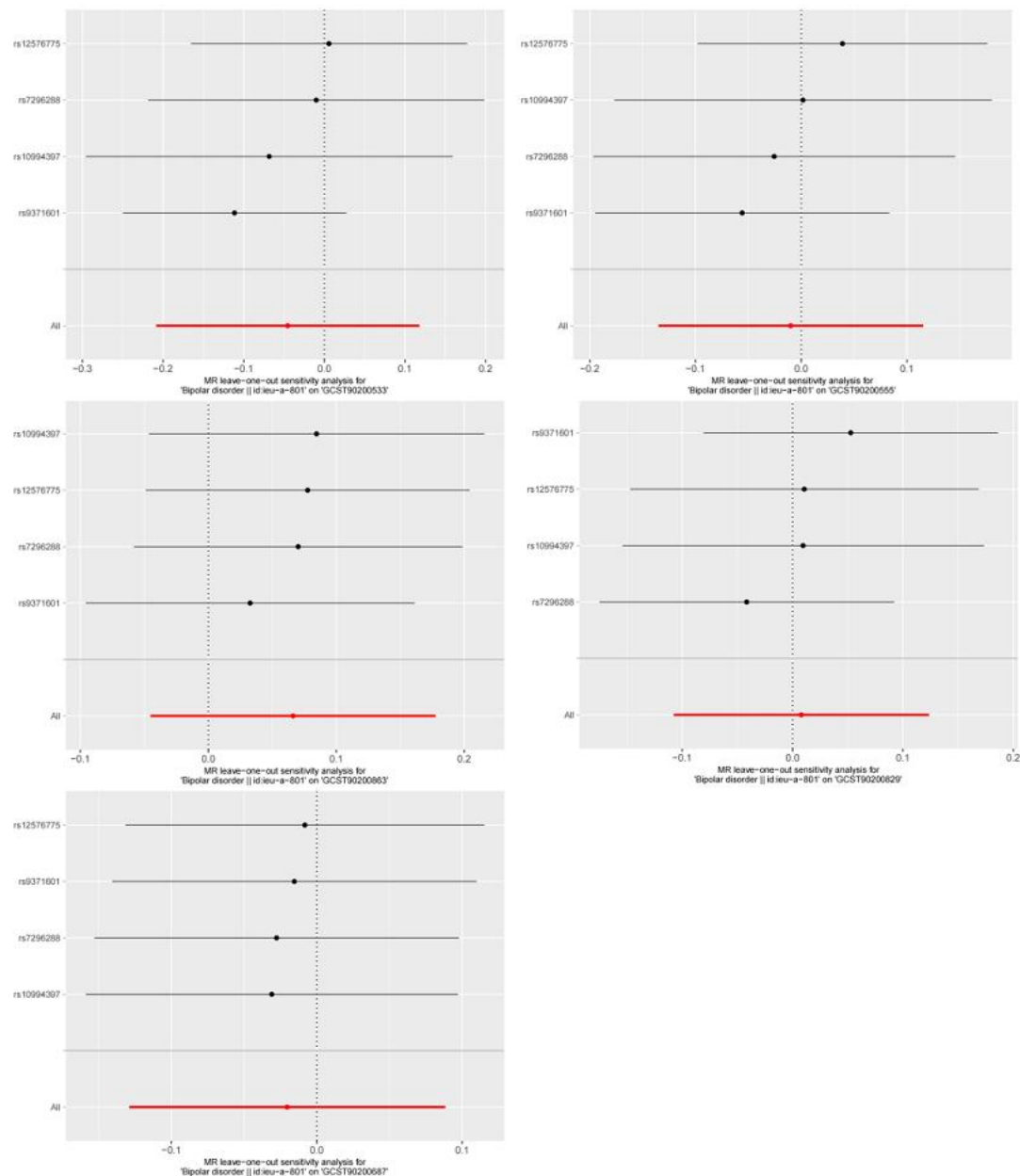


Figure 8. Results of leave-one-out sensitivity analysis of metabolites and bipolar disorder risk (GCST90200533; GCST90200687; GCST90200555; GCST90200863; GCST90200829).

3.4. Heterogeneity and pleiotropy tests

In the heterogeneity and pleiotropy tests, no evidence of heterogeneity was found in the results of the 17 known metabolites, indicating that the findings were not influenced by heterogeneity. Both the MR-Egger regression and MR-PRESSO methods did not detect any horizontal pleiotropy in the metabolite results. The leave-one-out plots displayed in Figure 6, Figure 7, Figure 8 showed that the overall results of beta values and 95% CIs did not exhibit significant changes after excluding individual SNPs, indicating the absence of biased SNPs. The combination of multiple methods consistently indicated the absence of heterogeneity and pleiotropy in the results, thereby demonstrating the robustness of the findings.

4. Conclusions

Through two-sample MR analysis, this study identified 17 plasma metabolites that showed a significant causal association with bipolar disorder. Among these, 11 plasma metabolites were negatively associated with the risk of bipolar disorder, while the remaining 6 metabolites showed a positive association. Notably, N-lactoyl valine levels, Linolenate [α or γ ; (18:3n3 or 6)] levels, and N-methyl-2-pyridone-5-carboxamide levels were identified as the plasma metabolites with the strongest protective effects. The reliability and stability of the findings were confirmed through consistent results from multiple analysis methods.

Further analyses for heterogeneity and pleiotropy did not reveal any evidence of heterogeneity or pleiotropic effects in the results. MR-Egger regression and MR-PRESSO methods did not detect horizontal pleiotropy in the plasma metabolite results. Sensitivity analysis demonstrated that excluding individual single nucleotide polymorphisms (SNPs) did not significantly affect the overall results, indicating the absence of biased SNPs. The combination of multiple methods consistently supported the robustness of the results, confirming the absence of heterogeneity and pleiotropy.

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this article.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

References

- [1] Andrew, A.N.; Bruno, A.O.Köhler-Forsberg, et al. Diagnosis and Treatment of Bipolar Disorder. *Jama*, 2023, 330(14), 1370-1380.
- [2] Kevin, S O'Connell.; Brandon, J.C. Genetic contributions to bipolar disorder: current status and future directions. *Psychological Medicine*, 2021, 51(13), 2156-2167.
- [3] Maik, P.; Isobel, D.S.; Johannes, Raffler, et al. Plasma metabolites to profile pathways in noncommunicable disease multimorbidity. *Nature Medicine*, 2021, 27(3), 471-479.
- [4] Schrimpe-rutledge, A.C.; Codreanu, S.G.; Sherrod, Stacy-D, el. Untargeted Metabolomics Strategies-Challenges and Emerging Directions. *Journal of the American Society for Mass Spectrometry*, 2016, 12, 1897-1905.
- [5] Peter, C.G.; Sean, E.; Jaclyn, N.T. et al. Metabolic pathways and immunometabolism in rare kidney diseases. *Annals of the Rheumatic Diseases*, 2018, 8, 1226-1233.
- [6] Juliet Richetto,Urs Meyer. Epigenetic Modifications in Schizophrenia and Related Disorders: Molecular Scars of Environmental Exposures and Source of Phenotypic Variability. *Biological Psychiatry*, 2021, 89(3), 215-226.
- [7] Jared, G.; Maina, Z.N.; Balkhiyarova, A.N. et al. Bidirectional Mendelian Randomization and Multiphenotype GWAS Show Causality and Shared

- Pathophysiology Between Depression and Type 2 Diabetes. *Diabetes Care*, 2023, 46(9), 1707-1714.
- [8] Stephen, B.; George, D.S.; Neil, M.D, et al. Guidelines for performing Mendelian randomization investigations: update for summer 2023. *Wellcome Open Research*, 2019, 4, 186.
- [9] Veronika, W.S.; Rebecca, C.R.; Benjamin, A.R.W. et al. Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization. *Jama*, 2021, 326(16), 1614-1621.
- [10]Chen, Y.C.; Tianyuan, L.; Ulrika, P.K. et al. Genomic atlas of the plasma metabolome prioritizes metabolites implicated in human diseases. *Nature Genetics*, 2023, 55(1), 44-53.
- [11]Hou, L.P.; Sarah, E.B.; Nirmala, A. et al. Genome-wide association study of 40,000 individuals identifies two novel loci associated with bipolar disorder. *Human Molecular Genetics*, 2016, 25(15), 3383-3394.
- [12]Nikos, P.; Niki, D.; Konstantinos, K.T. et al. Physical activity and risks of breast and colorectal cancer: a Mendelian randomisation analysis. *Nature Communications*, 2020, 11(1), 597.
- [13]Danielle, R.; Chirag, J.P. Conducting a Reproducible Mendelian Randomization Analysis Using the R Analytic Statistical Environment. *Current Protocols in Human Genetics*, 2019, 101(1), e82.
- [14]Panagiota, P.; Roxanna, S.; Korologou, L.; Laura, D.H. et al. Causal effects of circulating cytokine concentrations on risk of Alzheimer's disease and cognitive function. *Brain, Behavior, and Immunity*, 2022, 104, 54-64.
- [15]Niu, P.P.; Song, B.; Wang, X. el. Serum Uric Acid Level and Multiple Sclerosis: A Mendelian Randomization Study. *Frontiers in genetics*, 2020, 11, 254.
- [16]Wu, K.H.; Sun, Q.H.; Liu, D.X. el. Genetically predicted circulating levels of cytokines and the risk of oral cavity and pharyngeal cancer: a bidirectional mendelian-randomization study. *Frontiers in genetics*, 2023, 14, 1321484.



© 2024 by the author(s); licensee International Technology and Science Publications (ITS), this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (<http://creativecommons.org/licenses/by/4.0/>)