Multiomic approach and Mendelian randomization analysis identify causal associations between blood biomarkers and subcortical brain structure volumes.

Pritesh Jain[1]**, Madison Yates[1]**, Carlos Rubin de Celis[1], Petros Drineas[2], Neda Jahanshad[3], Paul Thompson[3], Peristera Paschou[1]*

1. Department of Biological Sciences, Purdue University
2. Department of Computer Science, Purdue University
3. Mark and Mary Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of South California

* **Corresponding Author**: Dr. Peristera Paschou, Lilly Hall of Life Sciences, 915 W. State Street, Room 1-225, West Lafayette, IN 47907. E-mail: ppaschou@purdue.edu

**Equal contribution first authors.

Abstract

Alterations in subcortical brain structure volumes have been found to be associated with several neurodegenerative and psychiatric disorders. At the same time, genome-wide association studies (GWAS) have identified numerous common variants associated with brain structure. In this study, we integrate these findings, aiming to identify proteins, metabolites, or microbes that have a putative causal association with subcortical brain structure volumes via a two-sample Mendelian randomization approach. This method uses genetic variants as instrument variables to identify potentially causal associations between an exposure and an outcome. The exposure data that we analyzed comprised genetic associations for 2,994 plasma proteins, 237 metabolites, and 103 microbial genera. The outcome data included GWAS data for seven subcortical brain structure volumes including accumbens, amygdala, caudate, hippocampus, pallidum, putamen, and thalamus. Eleven proteins and six metabolites were found to have a significant association with subcortical structure volumes. We found causal associations between amygdala volume and granzyme A as well as association between accumbens volume and plasma protease c1 inhibitor. Among metabolites, urate had the strongest association with thalamic volume. No significant associations were detected between the microbial genera and subcortical brain structure volumes. We also observed significant enrichment for biological processes such as proteolysis, regulation of the endoplasmic reticulum apoptotic signaling pathway, and negative regulation of DNA binding. Our findings provide insights to the mechanisms through which brain volumes may be affected in the pathogenesis of neurodevelopmental and psychiatric disorders and point to potential treatment targets for disorders that are associated with subcortical brain structure volumes.

Introduction

Variations and dysfunctions of subcortical brain structures have been associated with numerous neurological and neuropsychiatric disorders such as Parkinson's disease, different types of dementia, insomnia, schizophrenia, autism spectrum disorder (ASD), depression and post-traumatic stress disorder (PTSD) [1–7]. However, it is largely unknown how abnormalities of specific subcortical structures influence different traits and what processes may influence disease-related changes in subcortical brain structures. Understanding the relationship between brain volume and structure and neurological disease would help us better determine the underlying pathophysiological pathways. Such analysis could also be important in clinical practice, providing biomarkers that could be useful in disease diagnosis and patient management as well as helping to identify treatment targets for the various disorders associated with abnormalities in subcortical brain structure.

Recent large-scale multicenter studies such as the Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) and UK Biobank (UKB) have put together neuroimaging and genomic data from tens of thousands of individuals and performed genome-wide association studies. This

has led to the identification of genetic variants that are associated with subcortical brain structure volumes [8–10]. These studies have been followed by transcriptomic and epigenomic analysis to identify genes and epigenetic markers associated with regional brain volumes [11–13]. However, studies seeking to identify associations between regional brain volumes and other biomarkers such as proteins, metabolites and the microbiome are limited.

Here, we seek to address this gap, exploring the role of the proteome, metabolome, and microbiome in mediating brain structure changes which could lead to neurological disease. Proteins are the final product of gene expression and are an important intermediary phenotype that can provide insight into the cellular processes and functions that influence human biology and disease pathophysiology [14]. On the other hand, metabolites are small molecules that are a product and intermediates of cellular metabolism and play a pivotal role in cellular and physiological processes [15, 16]. The observed levels of such metabolites in biofluids can elucidate these processes. Finally, the human microbiota play an important role in the fermentation of non-digestible substrates as well as providing protection against foreign pathogens [17, 18]. A number of studies have found that changes in the level of different proteins, metabolites and the composition of the gut microbiome are associated with different metabolic, immunological as well as neurological disorders [19–22]. The importance of the level of different metabolites such as glucose, lactate and pyruvate in the cerebrospinal fluid (CSF) is well known and they are established biomarkers to study inflammation and malignancies in the brain [23]. The gut-brain axis (GBA) - which consists of bidirectional communication between the central and the enteric nervous system - is heavily influenced by the gut microbiota [24], establishing the importance of the microbiome in neurological functions and disorders.

Although the levels of these biomarkers in the body (especially metabolites and gut microbiome) are heavily influenced by environmental factors such as diet, medication and lifestyle [25–28], twin and family-based studies show that genetics also play an important role and they are highly heritable [29–31]. With advancements in profiling methods, large-scale studies can measure the levels of thousands of proteins and the various metabolites circulating in the blood and identify genetic variants which influence the level of these biomarkers [14, 32, 33]. Genome-wide association studies have also been performed to identify genetic variants that are associated with the composition of various bacterial taxa in the gut microbiome [34]. With results from these multi-omic studies at hand, there is the opportunity to investigate potential causal associations between such biological markers and subcortical brain structure volumes, using a two-sample Mendelian randomization (MR) approach.

MR analysis is a genetic epidemiological method that can help to determine putative causal associations between an exposure and an outcome using genetic variants as instrument variables [35, 36]. The method is conceptually similar to a randomized controlled trial which is based on the idea that the individuals receiving the treatment/drug (the instrument variable) are assigned

randomly to the different groups [37]. Similarly, in MR studies, the SNPs are randomized by nature, assigned to offspring before birth and are not confounded by any environmental factor - thus satisfying the requirement of a randomized trial [36, 38]. This method is very powerful and can use the vast number of publicly available results of GWAS to identify causal associations between different exposures and outcomes. Indeed, studies undertaking this approach have identified causal associations between proteins and disorders such as depression, anorexia, ASD and many others [19, 39–41]. MR studies have also uncovered associations between the gut microbiome and autoimmune and cardiovascular disorders [42, 43]. MR studies for brain structures have also found causal associations between subcortical brain structure and neurological conditions like schizophrenia, anorexia, depression and other disorders[44–47]. However, so far no studies have examined associations between the different biomarkers and metrics of subcortical brain structures.

In this study, we sought to better understand the mechanisms and mediators that lead to the observed associations between brain structures and neurological and neuropsychiatric disease. In a systems biology approach, we integrated multi omic data with GWAS for subcortical brain volumes and employed a two-sample MR approach to ask if proteome, metabolome, and microbiome could be causally associated with volume of different subcortical brain structures. The central hypothesis of our study was that specific genetic variants influence subcortical brain volumes by altering levels of different biomarkers from the proteome, metabolome, or microbiome.

Methods

Study Design and Datasets

We applied a two-sample MR analysis to determine and identify causal associations between three multi-omic datasets (plasma proteome; metabolome; microbiome) and seven different subcortical brain structure volumes (accumbens, amygdala, caudate, hippocampus, pallidum, putamen and thalamus) using genetic variants as instrument variables. **Figure 1** shows the overall design of the analysis. The basic principle of MR is that SNPs (genetic instruments), which are significantly associated with modifiable exposure, would be causally associated with the exposure-related outcome. Three important assumptions are required for a valid genetic instrument and MR analysis. First, the instrument must be causally related to the exposure. Second, it must be independent of any confounders; and, finally, it should only be associated with the outcome through the exposure. In our current study, the genetic instruments for the different exposures were obtained from large-scale GWAS studies for each of the different omic datasets (information on these studies is shown in **Supplementary Table 1**). Overall, we obtained GWAS data on 2,994 plasma proteins, 237 blood metabolites and 103 microbial genera [32–34]. Our outcome dataset

included the GWAS summary statistics for the seven subcortical brain structure volumes (adjusted for intracranial volume) obtained from the ENIGMA consortium [8].

Selection of Genetic Instruments

The first step to performing MR analysis is the selection of instrument variables. We used a threshold of nominal significance ($P < 1 \times 10^{-5}$) to select SNPs from the GWAS summary statistics for each of the exposure variables. Ideally, genome-wide SNPs ($P < 5 \times 10^{-8}$) are used for MR analysis but a relatively relaxed threshold for the genetic instruments has been previously used in MR investigations when there were no or only a few genome wide SNPs available [40], [41, 48, 49]. To select independent SNPs, we performed LD clumping using PLINK2 with an r^2 threshold of 0.01 using the 1000 Genomes European dataset as the reference panel [50]. The next steps of the analysis were performed using the TwoSampleMR package in R [51]. Once the independent SNPs were selected, we harmonized the exposure and outcome datasets to match the effect alleles, obtained the SNP effects and corresponding standard errors, and removed ambiguous SNPs with intermediate allele frequencies. In cases where a SNP was not available in the outcome dataset, a proxy SNPs with high LD with main SNP was used (LD at $r^2 > 0.8$) for the analysis. We then evaluated the instrument strength of each of the exposures by estimating the proportion of variance explained by the SNPs (R^2) and the F-statistic for each of the variables [52]. Typically, an F statistic >10 is considered sufficiently informative for MR analysis [53]We extracted a range of seven to 84 SNPs for the proteome data with an average R^2 of 21% and the minimum F statistic was 20.56. The number of SNPs for the metabolites ranged from three to 241 with an average R^2 of 13.1% and a minimum F statistic of 20.52. Finally, for the various microbial genera we extracted 3 to 22, with an average R^2 of 3.2% and the lowest F-statistic of 20.46. The number of instrument variables, R^2 and F-statistics for each individual biomarker is shown in Additional file 1.

Two Sample MR Analysis and Statistical Validation

We used the inverse variance weighted (IVW) method of MR analysis to estimate the association between the different exposures and outcomes. The method provides a high power estimate and assumes that all the genetic instruments used for the analysis are valid. Significant associations of protein, metabolites and microbiomes with the different subcortical brain structures were identified after adjusting for multiple testing using the Benjamini-Hochberg false discovery rate (FDR) threshold of 0.05. We then performed downstream validation using other methods of MR estimation, heterogeneity analysis and pleiotropy analysis for the significant associations. Two methods - the weighted median method and MR-Egger method - were adopted as alternate methods to evaluate the robustness of causality and detect pleiotropy. These methods are useful to validate the results of the MR analysis in case we use SNPs that do not satisfy the assumptions for the analysis. The weighted median method provides a consistent estimate if less than 50% of the SNPs were invalid instruments[54] and the MR-Egger method was useful when up to 100% of the SNPs came from invalid instruments[55]. Cochran's Q test was performed to test for heterogeneity, and pleiotropy was tested by performing an MR-Egger Intercept test and a leave-

one out analysis. The directionality test to validate whether the genetic instruments were acting on the outcome through the exposure was tested using the MR Steiger directionality test, which calculates the variance explained in the exposure and the outcome by the instrumenting SNPs, and tests if the variance in the outcome is less than the exposure [56]. We also performed reverse MR analysis with the subcortical brain structure volume as exposure and the biomarkers as outcomes. This allows us to evaluate if there were any feedback loops between the brain structures and biomarker levels which could lead to false positive results. We used the same thresholds to select the genetic instruments from the GWAS studies of the subcortical structures and used the IVW method to estimate the association.

Functional Enrichment Analysis

Functional enrichment analysis was performed using the gProfiler tool [57]. We tested for enrichment across different gene ontology terms, KEGG and reactome pathway databases, protein complexes and human phenotype ontology databases. A Bonferroni threshold was used to correct for multiple testing for all pathways tested. The pathway and enrichment analysis for metabolites was performed using the MetaboAnalyst platform [58].

Results

Investigating the Causal Association between Proteome and Subcortical Brain Structures

Using two sample MR analysis, we tested for potentially causal associations between 2,994 proteins and seven subcortical brain volumes (Additional file 2). Eleven proteins showed significant causal association with one of the subcortical brain structures as shown in **Figure 2** and **Supplementary Table 2**. Agouti Signaling Protein (ASIP) had the strongest association with putamen volume, with increase in the protein expression resulting in decrease in putamen volume (Beta: -28, p-value: 1.2×10^{-8}). Plasma protease C1 inhibitor (SERPING1) and secretoglobin family 1C member 1 (SCGB1C1) were both found to be causally associated with accumbens volume, with the increase in expression of these proteins being associated with increase in the volume of accumbens (Beta: 6.3-9.7, p-value: 3×10^{-5} - 6.9×10^{-7}). Increase in Granzyme A (GZMA) levels was found to be significantly associated with increase in amygdala volume (Beta: 17, p-value: 1.43×10^{-5}). Two proteins had a significant causal association with caudate volume. Increase in Thioredoxin domain containing protein 12 (TXNDC12) levels was associated with increase in caudate volume (Beta: 11.7, p-value: 2.3×10^{-6}), whereas Transmembrane protease serine 11D (TMPRSS11D) had a negative association (Beta: -26.8, p-value: 7.1×10^{-7}). For the hippocampus, we found four proteins significantly associated and all of them had a negative association with volume of hippocampus. These included Copine-1 (CPNE1), Cardiotrophin-1 (CTF1), Selenoprotein S (VIMP) and and Protein CEI (C5orf38) (Beta: -21.2 to -25.9, p-value: 4.9×10^{-5} - 9.8×10^{-7}). Finally, we found that increases in Chymotrypsinogen B (CTRB1) were significantly

associated with decrease in the volume of thalamus (Beta: -23.9, *p*-value: 1.4×10^{-5}). No proteins were found to be significantly associated with pallidum volume after multiple testing corrections.

Investigating Causal Association between Metabolome and Subcortical Brain Structures

We proceeded to test for potentially causal association between metabolites and subcortical brain structure (Additional file 3). We found six metabolites to be significantly associated with one of the subcortical brain structure volumes (**Supplementary Table 3** and **Figure 3**). Among these, two metabolites had a causal association with amygdala volume. These included uridine levels which had a positive association (Beta: 255.9, *p*-value: 1.44×10^{-4}) and Arachidonate which had a negative association with amygdala volume (Beta: -110.4, *p*-value: 2.54×10^{-4}). We also found three metabolites significantly associated with thalamus volume which were Urate (Beta: -458.7, *p*-value: 3.7×10^{-5}), 1-arachidonoylglycerphosphocholine (Beta: 269.7, *p*-value: 1.1×10^{-4}) and N-acetylornithine (Beta: 72.4, *p*-value: 5.6×10^{-4}). Increase in mannose levels was found to be causally associated with increase in caudate volume (Beta: 244.7, *p*-value: 5.5×10^{-5}).

Investigating Causal Association between Microbiome and Subcortical Brain Structures

Here, we pursued MR analysis between 103 microbial genera as exposure and subcortical structure as outcome. Our analysis did not reveal any significant associations after multiple testing corrections (Additional file 4).

Heterogeneity, Sensitivity and Pleiotropy Analyses

To determine the robustness and the validity of our results, we performed downstream statistical analysis to further increase the confidence in the observed associations. For all the significant associations identified in the primary analysis, we repeated the MR analysis using other methods such as the weighted median method and the MR-Egger method. We found that the associations were largely consistent with effects in the same direction and a significant *p*-value for the proteins (**Supplementary Table 4**). The MR-Egger estimate between the metabolites and subcortical brain volumes was found to be non-significant (**Supplementary Table 5**). We then determined if there was any heterogeneity in the genetic instruments used by calculating the Cochran's *Q* statistic and found little to no evidence of heterogeneity (p-value: 0.094-0.99) for all proteins and metabolites (**Table 1A** and **1B**). Following this, we tested for pleiotropy of SNPs between exposure and outcome using the Egger intercept test and leave one out analysis. We found no evidence of pleiotropy (Egger Intercept *p*-value: 0.06-0.95) and leave one out analysis showed that removing any SNP did not greatly affect the association (**Table 1** and additional file 4). One of the assumptions of MR is that the instruments influence the exposure first and then the outcome through the exposure. To evaluate this, we used the MR-Steiger test which calculates the variance explained in the exposure and the outcome by the instrumenting SNPs, and tests if the variance in the outcome is less than the exposure. The test showed that for all the proteins and metabolites that had significant associations with subcortical volume, the variance of the genetic instruments in the

exposure is always greater than the outcome - thus validating the assumption of MR (**Supplementary Tables 6 and 7**).

Reverse Mendelian Randomization Analysis

We performed the MR analysis with the subcortical brain structure volumes as exposure and the significantly associated biomarkers as outcomes. The results showed that for all proteins except C5orf38, there was no reverse causation observed in our analysis (**Table 2A**), thus indicating the causal effects of the proteins on the subcortical brain volume were statistically robust and not false positives. No reverse association was found between subcortical brain volume and the six metabolites as well (**Table 2B**).

Functional Enrichment Analysis

Analysis of the associated proteins using the g:Profiler platform revealed significant enrichment for various Gene Ontology terms after adjusting for multiple testing (**Figure 4** and **Supplementary Table 8**). These included molecular functions such as endopeptidase activity, peptidase activity and hydrolase activity. We also observed significant enrichment for biological processes such as proteolysis, regulation of the endoplasmic reticulum apoptotic signaling pathway and negative regulation of DNA binding. Most of the proteins were enriched in the extracellular regions of the human system. No significant enrichment was observed for the metabolites across all metabolic pathways.

Discussion

Here, pursuing a systems biology, multi-omic approach, we sought to provide insights into the mechanisms and mediators that underlie known associations of brain structures and neuropsychiatric disease. To do this, we performed a two-sample MR analysis to identify potentially causal associations between the genetically predicted levels of different biomarkers (plasma proteome, blood metabolome and gut microbiome) and the volumes of seven subcortical brain structures. Analyzing available summary statistics from large-scale GWAS, we identified eleven proteins and six metabolites to have a significant causal association with at least one subcortical structure after correcting for multiple testing. Heterogeneity and pleiotropy analysis showed low to no deviation from null thus validating our associations as truly significant. Bi-directional MR analysis for the significant associations showed no reverse causation for any proteins or metabolites except one (C5orf38, which is an unknown protein). Finally, enrichment analysis of the association proteins showed significant enrichment for proteolytic processes including endopeptidase, peptidase and hydrolase activities. No significant causal associations were observed between different bacterial genera in the gut microbiome and subcortical brain structures.

The molecular functions and the roles of the different proteins identified in this analysis as causally associated with subcortical brain volumes point to various pathways and mechanisms that could also help explain the relationship between subcortical structures and neuropsychiatric disorders. For example, SERPING1 - which is a Plasma Protease inhibitor - is a glycosylated protein involved in the regulation of the complement cascade and has been previously found to be associated with influencing frontal cortical thickness [57, 58]. The complement system itself has been implicated in depression, schizophrenia, and other neurodegenerative disorders as well [61], [62]. The nucleus accumbens has been an important brain region for regulating behaviors related to schizophrenia, depression and addiction [61, 62] and our results indicate that this regulation is being driven by levels of SERPING1, which is causally associated with accumbens volume. Similar relationships can also be observed for many of our identified proteins. GZMA, which is a serine protease involved in pyroptosis [65], is also found to have a lower expression in patients with major depressive disorder (MDD) compared to healthy controls [22]. Patients with MDD also tend to have decreased amygdala volume [23] which, based on our results, could be driven by GZMA. Another interesting example is that of TXNDC12, which is a member of the thioredoxin (Trx) superfamily. The Trx system is an antioxidant system that is important in maintaining sulfhydryl homeostasis protecting against oxidative stress [66]. Studies have pointed to the role of Trx-mediated oxidative stress in Parkinson's disease-associated dopaminergic neuron degeneration, thus indicating that this protein might be an important regulator of the dopamine reward system [67], [68]. The caudate which is part of the striatum and connected to the substantia nigra is heavily involved in the reward system where the dopaminergic neurons are produced [69]. Changes in caudate volume have been found to be associated with disorders such as anorexia and Parkinson's disease [70], [71] and our results point to TXNDC12 being the mediator of these associations via the dopaminergic neuron generation and reward system.

Some of the proteins we identified had an established role in brain development. For example, the three proteins we found causally associated with hippocampus volume were Copine-1, Cardiotrophin-1 and Selenoprotein. Copine 1 is a calcium dependent phospholipid binding protein and plays a role in neuronal progenitor cell differentiation and induces neurite outgrowth [72]. Similar to copine-1, cardiotrophin-1 is also involved in the differentiation of neuronal stem cells via a protein kinase dependent signaling pathway [73]. Selenoprotein is an important protein that mediates the levels of selenium in the brain and is important for mammalian brain development and other functions such as antioxidant protection and synaptic signaling [74].

Apart from these proteins, we also identified six metabolites that were causally associated with subcortical brain volume. Previous studies have shown that these metabolites have an important role in the functioning of the central nervous system and are also associated with different neurological disorders. Uric acid is considered a key antioxidant in humans [75], but high levels of this metabolite are associated with increased risk of disorders such as ASD and ADHD [76], [77]. Both of these disorders are also associated with reduced thalamic volume [78], [79], which

could be explained by higher levels of uric acid based on our results. Interestingly, both uric acid and uridine are implicated in the development of Lesch-Nyhan syndrome which is a congenital disorder that affects brain structure and behavior of the affected individuals [80]. Other metabolites such as mannose and arachidonate which were identified in our study have also been found to be associated with disorders like anxiety and depression in mouse model systems [81]–[83].

There are certain limitations of this study. First, there were very few or no genome-wide significant SNPs to be used as instrument variables for many biomarkers in the MR analysis. To address this, we used a more exploratory threshold of 1e-05 for selecting genetic instruments, like previously done in multiple previous studies [40, 41, 48, 49]. We evaluated the strength for these genetic instruments using different statistical methods and found that they were valid for MR analysis. Second, the proteins and metabolites were quantified in the plasma for the GWAS analysis, which is a natural choice for biomarker-focused applications considering its convenience; however, we do not know whether these biomarkers would have had similar levels in specific brain regions, because of the existence of the blood-brain-barrier. To address this, we checked for the expression and presence of the different proteins and metabolites in the CNS. We found that most of them are highly expressed in different parts of the brain [84] and play an important role in its development and function (**Supplementary Table 9**). We would also like to point out that, we performed an MR study and identified several statistically causal risk factors associated with the subcortical brain volume, but these findings need further biological validation using experimental verification in cells and model systems. Based on statistical analysis, our study points to the most reliable targets for downstream investment, analysis and experimental validation and provides novel insights into the physiology of brain structures.

In conclusion, we identified several proteins and metabolites that are causally associated with the volume of subcortical brain structures. Our study highlighted the role of proteolytic and anti-oxidative components in the development and functioning of the brain. The biomarkers we identified could mediate the relationship between subcortical structures and different neurological and neuropsychiatric disorders. Future analysis could examine other characteristics of the brain such as neuronal activity, gray matter volume, and white matter connectivity which could further improve our understanding of the functioning of the central nervous system and its association to disease. The results of this study not only provide novel insight for understanding subcortical brain structure, but also help in uncovering potential diagnostic markers and drug targets for the many disorders that are associated with changes in brain structures.

Statement of Contribution

PJ, MY, CR, and PP designed the study, and performed primary analysis. All authors provided data, materials, and methods. All authors contributed to interpreting results, writing, reviewing, and editing the manuscript.

Conflict of Interest

The authors declare no conflicts of Interest.

Data Availability

The GWAS summary statistics used for the analysis were downloaded from publicly available sources. The subcortical brain volume GWAS were downloaded from the ENIGMA consortium (https://enigma.ini.usc.edu/). The proteome GWAS were downloaded from the SomaLogic Plasma Protein GWAS study (http://www.phpc.cam.ac.uk/ceu/proteins/). The metabolites GWAS were obtained from Metabolites GWAS server (http://metabolomics.helmholtz-muenchen.de/gwas/) and the microbiome GWAS results were downloaded from https://mibiogen.gcc.rug.nl/.

References

[1] N. I. Bohnen and R. L. Albin, "White matter lesions in Parkinson disease," *Nat. Rev. Neurol.*, vol. 7, no. 4, pp. 229–236, Apr. 2011, doi: 10.1038/nrneurol.2011.21.

[2] T. M. Nir *et al.*, "Effectiveness of regional DTI measures in distinguishing Alzheimer's disease, MCI, and normal aging," *NeuroImage Clin.*, vol. 3, pp. 180–195, 2013, doi: 10.1016/j.nicl.2013.07.006.

[3] A. N. Voineskos, "Genetic underpinnings of white matter 'connectivity': Heritability, risk, and heterogeneity in schizophrenia," *Schizophr. Res.*, vol. 161, no. 1, pp. 50–60, Jan. 2015, doi: 10.1016/j.schres.2014.03.034.

[4] D. van Rooij *et al.*, "Cortical and Subcortical Brain Morphometry Differences Between Patients With Autism Spectrum Disorder and Healthy Individuals Across the Lifespan: Results From the ENIGMA ASD Working Group," *Am. J. Psychiatry*, vol. 175, no. 4, pp. 359–369, Apr. 2018, doi: 10.1176/appi.ajp.2017.17010100.

[5] K. Zhao *et al.*, "Cortical thickness and subcortical structure volume abnormalities in patients with major depression with and without anxious symptoms," *Brain Behav.*, vol. 7, no. 8, p. e00754, Jun. 2017, doi: 10.1002/brb3.754.

[6] F. Emamian *et al.*, "Alterations of Subcortical Brain Structures in Paradoxical and Psychophysiological Insomnia Disorder," *Front. Psychiatry*, vol. 12, p. 661286, May 2021, doi: 10.3389/fpsyt.2021.661286.

[7] X. Wang *et al.*, "Cortical volume abnormalities in posttraumatic stress disorder: an ENIGMA-psychiatric genomics consortium PTSD workgroup mega-analysis," *Mol. Psychiatry*, vol. 26, no. 8, pp. 4331–4343, Aug. 2021, doi: 10.1038/s41380-020-00967-1.

[8] D. P. Hibar *et al.*, "Common genetic variants influence human subcortical brain structures," *Nature*, vol. 520, no. 7546, pp. 224–229, Apr. 2015, doi: 10.1038/nature14101.

[9] C. L. Satizabal *et al.*, "Genetic architecture of subcortical brain structures in 38,851 individuals," *Nat. Genet.*, vol. 51, no. 11, Art. no. 11, Nov. 2019, doi: 10.1038/s41588-019-0511-y.

[10] P. M. Thompson *et al.*, "ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries," *Transl. Psychiatry*, vol. 10, no. 1, Art. no. 1, Mar. 2020, doi: 10.1038/s41398-020-0705-1.

[11] B. Zhao *et al.*, "Transcriptome-wide association analysis of brain structures yields insights into pleiotropy with complex neuropsychiatric traits," *Nat. Commun.*, vol. 12, no. 1, Art. no. 1, May 2021, doi: 10.1038/s41467-021-23130-y.

[12] M. C. Barbu *et al.*, "Epigenome-wide association study of global cortical volumes in generation Scotland: Scottish family health study," *Epigenetics*, vol. 17, no. 10, pp. 1143–1158, Oct. 2022, doi: 10.1080/15592294.2021.1997404.

[13] T. Jia *et al.*, "Epigenome-wide meta-analysis of blood DNA methylation and its association with subcortical volumes: findings from the ENIGMA Epigenetics Working Group," *Mol. Psychiatry*, vol. 26, no. 8, Art. no. 8, Aug. 2021, doi: 10.1038/s41380-019-0605-z.

[14] P. E. Geyer, N. A. Kulak, G. Pichler, L. M. Holdt, D. Teupser, and M. Mann, "Plasma Proteome Profiling to Assess Human Health and Disease," *Cell Syst.*, vol. 2, no. 3, pp. 185–195, Mar. 2016, doi: 10.1016/j.cels.2016.02.015.

[15] A. P. Nath *et al.*, "An interaction map of circulating metabolites, immune gene networks, and

their genetic regulation," *Genome Biol.*, vol. 18, no. 1, p. 146, Aug. 2017, doi: 10.1186/s13059-017-1279-y.

[16] E. A. Miles and P. C. Calder, "Fatty Acids, Lipid Emulsions and the Immune and Inflammatory Systems," *Intraven. Lipid Emuls.*, vol. 112, pp. 17–30, 2015, doi: 10.1159/000365426.

[17] J. A. Gilbert, M. J. Blaser, J. G. Caporaso, J. K. Jansson, S. V. Lynch, and R. Knight, "Current understanding of the human microbiome," *Nat. Med.*, vol. 24, no. 4, Art. no. 4, Apr. 2018, doi: 10.1038/nm.4517.

[18] A. M. Valdes, J. Walter, E. Segal, and T. D. Spector, "Role of the gut microbiota in nutrition and health," *BMJ*, p. k2179, Jun. 2018, doi: 10.1136/bmj.k2179.

[19] C. Yang *et al.*, "Genomic atlas of the proteome from brain, CSF and plasma prioritizes proteins implicated in neurological disorders," *Nat. Neurosci.*, vol. 24, no. 9, pp. 1302–1312, Sep. 2021, doi: 10.1038/s41593-021-00886-6.

[20] R. B. Mofrad *et al.*, "Plasma proteome profiling identifies changes associated to AD but not to FTD," *Acta Neuropathol. Commun.*, vol. 10, no. 1, p. 148, Oct. 2022, doi: 10.1186/s40478-022-01458-w.

[21] M. S. Sabatine *et al.,* "Metabolomic Identification of Novel Biomarkers of Myocardial Ischemia," *Circulation*, vol. 112, no. 25, pp. 3868–3875, Dec. 2005, doi: 10.1161/CIRCULATIONAHA.105.569137.

[22] A. Vijay and A. M. Valdes, "Role of the gut microbiome in chronic diseases: a narrative review," *Eur. J. Clin. Nutr.*, vol. 76, no. 4, Art. no. 4, Apr. 2022, doi: 10.1038/s41430-021-00991-6.

[23] W.-M. Zhang and M. R. Natowicz, "Cerebrospinal fluid lactate and pyruvate concentrations and their ratio," *Clin. Biochem.*, vol. 46, no. 7, pp. 694–697, May 2013, doi: 10.1016/j.clinbiochem.2012.11.008.

[24] M. Carabotti, A. Scirocco, M. A. Maselli, and C. Severi, "The gut-brain axis: interactions between enteric microbiota, central and enteric nervous systems," *Ann. Gastroenterol. Q. Publ. Hell. Soc. Gastroenterol.*, vol. 28, no. 2, pp. 203–209, 2015.

[25] D. Rothschild *et al.*, "Environment dominates over host genetics in shaping human gut microbiota," *Nature*, vol. 555, no. 7695, Art. no. 7695, Mar. 2018, doi: 10.1038/nature25973.

[26] L. Maier and A. Typas, "Systematically investigating the impact of medication on the gut microbiome," *Curr. Opin. Microbiol.*, vol. 39, pp. 128–135, Oct. 2017, doi: 10.1016/j.mib.2017.11.001.

[27] K. M. Bermingham *et al.*, "Genetic and Environmental Contributions to Variation in the Stable Urinary NMR Metabolome over Time: A Classic Twin Study," *J. Proteome Res.*, vol. 20, no. 8, pp. 3992–4000, Aug. 2021, doi: 10.1021/acs.jproteome.1c00319.

[28] G. Nicholson *et al.,* "Human metabolic profiles are stably controlled by genetic and environmental variation," *Mol. Syst. Biol.*, vol. 7, no. 1, p. 525, Jan. 2011, doi: 10.1038/msb.2011.57.

[29] F. A. Hagenbeek *et al.*, "Heritability estimates for 361 blood metabolites across 40 genome-wide association studies," *Nat. Commun.*, vol. 11, no. 1, Art. no. 1, Jan. 2020, doi: 10.1038/s41467-019-13770-6.

[30] J. K. Goodrich *et al.*, "Genetic Determinants of the Gut Microbiome in UK Twins," *Cell Host Microbe*, vol. 19, no. 5, pp. 731–743, May 2016, doi: 10.1016/j.chom.2016.04.017.

[31] J. K. Goodrich *et al.*, "Human Genetics Shape the Gut Microbiome," *Cell*, vol. 159, no. 4, pp. 789–799, Nov. 2014, doi: 10.1016/j.cell.2014.09.053.

[32] S.-Y. Shin *et al.*, "An atlas of genetic influences on human blood metabolites," *Nat. Genet.*, vol. 46, no. 6, Art. no. 6, Jun. 2014, doi: 10.1038/ng.2982.

[33] B. B. Sun *et al.*, "Genomic atlas of the human plasma proteome," *Nature*, vol. 558, no. 7708, Art. no. 7708, Jun. 2018, doi: 10.1038/s41586-018-0175-2.

[34] A. Kurilshikov *et al.*, "Large-scale association analyses identify host factors influencing human gut microbiome composition," *Nat. Genet.*, vol. 53, no. 2, Art. no. 2, Feb. 2021, doi: 10.1038/s41588-020-00763-1.

[35] C. A. Emdin, A. V. Khera, and S. Kathiresan, "Mendelian Randomization," *JAMA*, vol. 318, no. 19, pp. 1925–1926, Nov. 2017, doi: 10.1001/jama.2017.17219.

[36] E. Sanderson *et al.*, "Mendelian randomization," *Nat. Rev. Methods Primer*, vol. 2, no. 1, Art. no. 1, Feb. 2022, doi: 10.1038/s43586-021-00092-5.

[37] E. Hariton and J. J. Locascio, "Randomised controlled trials—the gold standard for effectiveness research," *BJOG Int. J. Obstet. Gynaecol.*, vol. 125, no. 13, p. 1716, Dec. 2018, doi: 10.1111/1471-0528.15199.

[38] S. A. Swanson, H. Tiemeier, M. A. Ikram, and M. A. Hernán, "Nature as a trialist? Deconstructing the analogy between Mendelian Randomization and randomized trials," *Epidemiol. Camb. Mass*, vol. 28, no. 5, pp. 653–659, Sep. 2017, doi: 10.1097/EDE.0000000000000699.

[39] T. S. Wingo *et al.*, "Brain proteome-wide association study implicates novel proteins in depression pathogenesis," *Nat. Neurosci.*, vol. 24, no. 6, pp. 810–817, Jun. 2021, doi: 10.1038/s41593-021-00832-6.

[40] J. Yang *et al.*, "Mendelian randomization analyses reveal novel drug targets for anorexia nervosa," *Prog. Neuropsychopharmacol. Biol. Psychiatry*, vol. 112, p. 110427, Jan. 2022, doi: 10.1016/j.pnpbp.2021.110427.

[41] J. Yang *et al.*, "Association between plasma proteome and childhood neurodevelopmental disorders: A two-sample Mendelian randomization analysis," *EBioMedicine*, vol. 78, p. 103948, Apr. 2022, doi: 10.1016/j.ebiom.2022.103948.

[42] Q. Xu *et al.*, "Causal Relationship Between Gut Microbiota and Autoimmune Diseases: A Two-Sample Mendelian Randomization Study," *Front. Immunol.*, vol. 12, p. 746998, Jan. 2022, doi: 10.3389/fimmu.2021.746998.

[43] Y. Zhang *et al.*, "Causal associations between gut microbiome and cardiovascular disease: A Mendelian randomization study," *Front. Cardiovasc. Med.*, vol. 9, 2022, Accessed: Feb. 06, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fcvm.2022.971376

[44] O. Wootton, M. Campbell, N. Jahanshad, P. Thompson, D. Stein, and S. Dalvie, "CHARACTERIZING THE SHARED GENETIC INFLUENCES BETWEEN SCHIZOPHRENIA AND SUBCORTICAL BRAIN REGIONS," *Eur. Neuropsychopharmacol.*, vol. 63, p. e302, Oct. 2022, doi: 10.1016/j.euroneuro.2022.07.533.

[45] E. Walton *et al.*, "Exploration of Shared Genetic Architecture Between Subcortical Brain Volumes and Anorexia Nervosa," *Mol. Neurobiol.*, vol. 56, no. 7, pp. 5146–5156, Jul. 2019, doi: 10.1007/s12035-018-1439-4.

[46] X. Shen *et al.*, "A phenome-wide association and Mendelian Randomisation study of polygenic risk for depression in UK Biobank," *Nat. Commun.*, vol. 11, no. 1, Art. no. 1, May

2020, doi: 10.1038/s41467-020-16022-0.

[47] B.-S. Wu *et al.*, "Cortical structure and the risk for Alzheimer's disease: a bidirectional Mendelian randomization study," *Transl. Psychiatry*, vol. 11, no. 1, Art. no. 1, Sep. 2021, doi: 10.1038/s41398-021-01599-x.

[48] K. W. Choi *et al.*, "Assessment of Bidirectional Relationships Between Physical Activity and Depression Among Adults: A 2-Sample Mendelian Randomization Study," *JAMA Psychiatry*, vol. 76, no. 4, pp. 399–408, Apr. 2019, doi: 10.1001/jamapsychiatry.2018.4175.

[49] S. Sanna *et al.*, "Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases," *Nat. Genet.*, vol. 51, no. 4, Art. no. 4, Apr. 2019, doi: 10.1038/s41588-019-0350-x.

[50] A. Auton *et al.*, "A global reference for human genetic variation," *Nature*, vol. 526, no. 7571, Art. no. 7571, Oct. 2015, doi: 10.1038/nature15393.

[51] G. Hemani *et al.*, "The MR-Base platform supports systematic causal inference across the human phenome," *eLife*, vol. 7, p. e34408, May 2018, doi: 10.7554/eLife.34408.

[52] M.-J. A. Brion, K. Shakhbazov, and P. M. Visscher, "Calculating statistical power in Mendelian randomization studies," *Int. J. Epidemiol.*, vol. 42, no. 5, pp. 1497–1501, Oct. 2013, doi: 10.1093/ije/dyt179.

[53] S. Burgess, A. Butterworth, and S. G. Thompson, "Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data," *Genet. Epidemiol.*, vol. 37, no. 7, pp. 658–665, Nov. 2013, doi: 10.1002/gepi.21758.

[54] J. Bowden, G. Davey Smith, P. C. Haycock, and S. Burgess, "Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator," *Genet. Epidemiol.*, vol. 40, no. 4, pp. 304–314, May 2016, doi: 10.1002/gepi.21965.

[55] J. Bowden, G. Davey Smith, and S. Burgess, "Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression," *Int. J. Epidemiol.*, vol. 44, no. 2, pp. 512–525, Apr. 2015, doi: 10.1093/ije/dyv080.

[56] G. Hemani, K. Tilling, and G. Davey Smith, "Orienting the causal relationship between imprecisely measured traits using GWAS summary data," *PLOS Genet.*, vol. 13, no. 11, p. e1007081, Nov. 2017, doi: 10.1371/journal.pgen.1007081.

[57] U. Raudvere *et al.*, "g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update)," *Nucleic Acids Res.*, vol. 47, no. W1, pp. W191–W198, Jul. 2019, doi: 10.1093/nar/gkz369.

[58] Z. Pang *et al.*, "MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights," *Nucleic Acids Res.*, vol. 49, no. W1, pp. W388–W396, Jul. 2021, doi: 10.1093/nar/gkab382.

[59] G. Stelzer *et al.*, "The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses," *Curr. Protoc. Bioinforma.*, vol. 54, no. 1, Jun. 2016, doi: 10.1002/cpbi.5.

[60] D. M. Allswede *et al.*, "Complement Gene Expression Correlates with Superior Frontal Cortical Thickness in Humans," *Neuropsychopharmacology*, vol. 43, no. 3, Art. no. 3, Feb. 2018, doi: 10.1038/npp.2017.164.

[61] Y. Yi, Y. Liu, K. Wu, W. Wu, and W. Zhang, "The core genes involved in the promotion of depression in patients with ovarian cancer," *Oncol. Lett.*, vol. 18, no. 6, pp. 5995–6007, Dec.

2019, doi: 10.3892/ol.2019.10934.

[62] M. Druart and C. Le Magueresse, "Emerging Roles of Complement in Psychiatric Disorders," *Front. Psychiatry*, vol. 10, 2019, Accessed: Feb. 06, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyt.2019.00573

[63] L. Xu, J. Nan, and Y. Lan, "The Nucleus Accumbens: A Common Target in the Comorbidity of Depression and Addiction," *Front. Neural Circuits*, vol. 14, 2020, Accessed: Jan. 26, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fncir.2020.00037

[64] M. Forns-Nadal *et al.*, "Increased nucleus accumbens volume in first-episode psychosis," *Psychiatry Res. Neuroimaging*, vol. 263, pp. 57–60, May 2017, doi: 10.1016/j.pscychresns.2017.03.007.

[65] J. Lieberman, "Granzyme A activates another way to die," *Immunol. Rev.*, vol. 235, no. 1, pp. 93–104, May 2010, doi: 10.1111/j.0105-2896.2010.00902.x.

[66] L. Arodin, A. Miranda-Vizuete, P. Swoboda, and A. P. Fernandes, "Protective effects of the thioredoxin and glutaredoxin systems in dopamine-induced cell death," *Free Radic. Biol. Med.*, vol. 73, pp. 328–336, Aug. 2014, doi: 10.1016/j.freeradbiomed.2014.05.011.

[67] A. Garcia-Garcia, L. Zavala-Flores, H. Rodriguez-Rocha, and R. Franco, "Thiol-Redox Signaling, Dopaminergic Cell Death, and Parkinson's Disease," *Antioxid. Redox Signal.*, vol. 17, no. 12, pp. 1764–1784, Dec. 2012, doi: 10.1089/ars.2011.4501.

[68] Z. Liu *et al.*, "Overexpression of thioredoxin reductase 1 can reduce DNA damage, mitochondrial autophagy and endoplasmic reticulum stress in Parkinson's disease," *Exp. Brain Res.*, vol. 239, no. 2, pp. 475–490, Feb. 2021, doi: 10.1007/s00221-020-05979-5.

[69] M. E. Driscoll, P. C. Bollu, and P. Tadi, "Neuroanatomy, Nucleus Caudate," in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2022. Accessed: Jan. 26, 2023. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK557407/

[70] A. Gupta *et al.*, "Complex functional brain network properties in anorexia nervosa," *J. Eat. Disord.*, vol. 10, no. 1, p. 13, Feb. 2022, doi: 10.1186/s40337-022-00534-9.

[71] T. L. Pitcher *et al.*, "Reduced striatal volumes in Parkinson's disease: a magnetic resonance imaging study," *Transl. Neurodegener.*, vol. 1, p. 17, Aug. 2012, doi: 10.1186/2047-9158-1-17.

[72] N. Park, J. C. Yoo, J. Ryu, S.-G. Hong, E. M. Hwang, and J.-Y. Park, "Copine1 Enhances Neuronal Differentiation of the Hippocampal Progenitor HiB5 Cells," *Mol. Cells*, vol. 34, no. 6, pp. 549–554, Dec. 2012, doi: 10.1007/s10059-012-0235-7.

[73] L. Peng, X. Shu, C. Lang, and X. Yu, "Cardiotrophin-1 stimulates the neural differentiation of human umbilical cord blood-derived mesenchymal stem cells and survival of differentiated cells through PI3K/Akt-dependent signaling pathways," *Cytotechnology*, vol. 69, no. 6, pp. 933–941, Dec. 2017, doi: 10.1007/s10616-017-0103-6.

[74] N. D. Solovyev, "Importance of selenium and selenoprotein for brain function: From antioxidant protection to neuronal signalling," *J. Inorg. Biochem.*, vol. 153, pp. 1–12, Dec. 2015, doi: 10.1016/j.jinorgbio.2015.09.003.

[75] B. F. Becker, "Towards the physiological function of uric acid," *Free Radic. Biol. Med.*, vol. 14, no. 6, pp. 615–631, Jun. 1993, doi: 10.1016/0891-5849(93)90143-I.

[76] T. Page and M. Coleman, "Purine metabolism abnormalities in a hyperuricosuric subclass of autism," *Biochim. Biophys. Acta BBA - Mol. Basis Dis.*, vol. 1500, no. 3, pp. 291–296, Mar. 2000, doi: 10.1016/S0925-4439(99)00113-1.

[77] A. R. Sutin *et al.*, "Impulsivity is Associated with Uric Acid: Evidence from Humans and Mice," *Biol. Psychiatry*, vol. 75, no. 1, p. 10.1016/j.biopsych.2013.02.024, Jan. 2014, doi: 10.1016/j.biopsych.2013.02.024.

[78] R. Tamura, H. Kitamura, T. Endo, N. Hasegawa, and T. Someya, "Reduced thalamic volume observed across different subgroups of autism spectrum disorders," *Psychiatry Res. Neuroimaging*, vol. 184, no. 3, pp. 186–188, Dec. 2010, doi: 10.1016/j.pscychresns.2010.07.001.

[79] S. Xia, X. Li, A. E. Kimball, M. S. Kelly, I. Lesser, and C. Branch, "Thalamic Shape and Connectivity Abnormalities in Children with Attention Deficit/Hyperactivity Disorder," *Psychiatry Res.*, vol. 204, no. 2–3, pp. 161–167, Nov. 2012, doi: 10.1016/j.pscychresns.2012.04.011.

[80] H. A. Jinnah, "Lesch-Nyhan disease: from mechanism to model and back again," *Dis. Model. Mech.*, vol. 2, no. 3–4, pp. 116–121, 2009, doi: 10.1242/dmm.002543.

[81] K. Xu, M. Wang, W. Zhou, J. Pu, H. Wang, and P. Xie, "Chronic D-ribose and D-mannose overload induce depressive/anxiety-like behavior and spatial memory impairment in mice," *Transl. Psychiatry*, vol. 11, no. 1, p. 90, Feb. 2021, doi: 10.1038/s41398-020-01126-4.

[82] T. Larrieu and S. Layé, "Food for Mood: Relevance of Nutritional Omega-3 Fatty Acids for Depression and Anxiety," *Front. Physiol.*, vol. 9, 2018, Accessed: Feb. 06, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fphys.2018.01047

[83] H. Yu, X. Qin, Z. Yu, Y. Chen, L. Tang, and W. Shan, "Effects of high-fat diet on the formation of depressive-like behavior in mice," *Food Funct.*, vol. 12, no. 14, pp. 6416–6431, Jul. 2021, doi: 10.1039/d1fo00044f.

[84] M. Uhlén *et al.*, "Tissue-based map of the human proteome," *Science*, vol. 347, no. 6220, p. 1260419, Jan. 2015, doi: 10.1126/science.1260419.

Table 1: Statistical validation of MR results. The table shows the results of heterogeneity and pleiotropy tests performed for all biomarkers that had significant association with subcortical volume. Table 1A shows the results for proteins and 1B shows the results for the metabolites. Q refers to Cochran's Q estimate for the heterogeneity test; DF is the degree of freedom. The Int refers to the MR-Egger intercept for the pleiotropy test and SE is the standard error of the Intercept.

Biomarker ID (Exposure)	Region (Outcome)	Heterogeneity Test			Pleiotropy test		
		Q	Q_DF	p-val	Int	SE	p-val
(1A) Proteins							
SCGB1C1.5960.49.3	Accumbens	7.12	16	0.971	-0.12	0.88	0.893
SERPING1.4479.14.2	Accumbens	43.59	35	0.151	0.12	0.63	0.852
GZMA.3440.7.2	Amygdala	21.54	26	0.714	-0.45	2.26	0.841
TMPRSS11D.6547.83.3	Caudate	58.99	46	0.095	2.85	2.78	0.311
TXNDC12.4815.25.3	Caudate	87.65	80	0.261	-2.37	1.63	0.15
C5orf38.6378.2.3	Hippocampus	41.42	43	0.540	0.11	1.82	0.952
CPNE1.5346.24.3	Hippocampus	28.60	24	0.236	1.62	2.46	0.516
CTF1.13732.79.3	Hippocampus	34.77	28	0.177	-1.97	2.60	0.454
VIMP.11286.78.3	Hippocampus	24.13	28	0.675	-0.30	2.17	0.891
ASIP.5676.54.3	Putamen	55.63	49	0.239	3.79	1.97	0.061
CTRB1.5671.1.3	Thalamus	21.99	40	0.991	1.57	2.47	0.528
(1B) Metabolites							
Uridine	Amygdala	11.90	21	0.942	1.1	2.26	0.634
Arachidonate	Amygdala	20.25	28	0.855	0.41	1.01	0.686
Mannose	Caudate	32.76	30	0.333	4.71	2.81	0.104
Urate	Thalamus	23.63	27	0.650	-1.64	3.29	0.623
1-arachidonoylglycerphosphocholine	Thalamus	25.42	22	0.277	-0.28	2.64	0.915
N-acetylornithine	Thalamus	18.74	25	0.809	3.78	2.27	0.109

Table 2: Reverse MR analysis. The table shows the results of MR analysis with the subcortical brain structures as exposure and the biomarkers that were significant in the primary analysis as the outcomes. Table 2A shows the results for proteins and 2B shows the results for the metabolites. N SNPs is the number of genetic instruments used for the analysis.

| Region (Exposure)	Biomarker ID (Outcome)	N SNPs	Beta	SE	p-value
(2A) Proteins					
Accumbens	SCGB1C1.5960.49.3	7	5.84E-05	0.00143	0.967
Accumbens	SERPING1.4479.14.2	7	0.0024	0.00175	0.159
Amygdala	GZMA.3440.7.2	17	0.0001	0.00048	0.795
Caudate	TMPRSS11D.6547.83.3	24	0.00018	0.00021	0.377
Caudate	TXNDC12.4815.25.3	24	-0.00026	0.00023	0.268
Hippocampus	C5orf38.6378.2.3	18	-0.00066	0.00025	0.008
Hippocampus	CPNE1.5346.24.3	18	-0.00126	0.00119	0.288
Hippocampus	CTF1.13732.79.3	18	-0.00022	0.00025	0.369
Hippocampus	VIMP.11286.78.3	18	-0.00019	0.00025	0.435
Putamen	ASIP.5676.54.3	28	-0.00014	0.00017	0.410
Thalamus	CTRB1.5671.1.3	28	0.00005	0.00015	0.743
(2B) Metabolites					
Amygdala	Uridine	6	-3E-06	7.28E-05	0.967
Amygdala	Arachidonate	6	7.84E-05	0.0001	0.437
Caudate	Mannose	8	-8.23E-05	4.21E-05	0.051
Thalamus	Urate	12	-7.89E-06	1.69E-05	0.641
Thalamus	1-arachidonoylglycerphosphocholine	12	-4.13E-05	3.73E-05	0.267
Thalamus	N-acetylornithine	12	-3.60E-05	5.05E-05	0.475
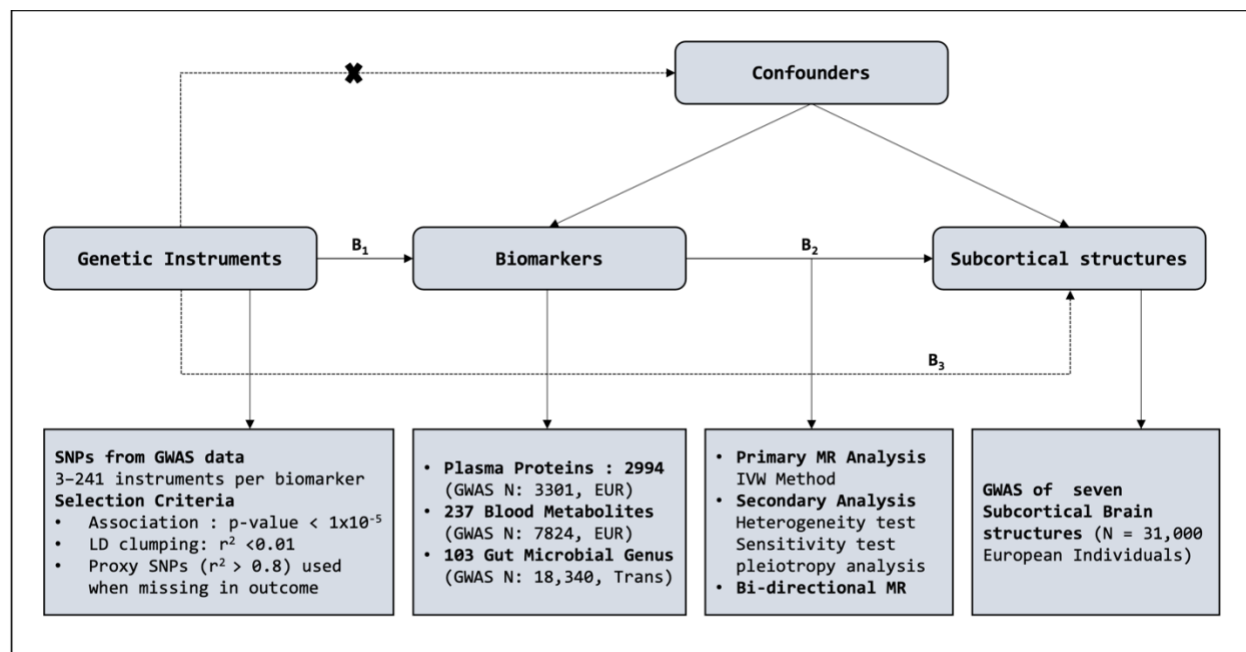
Figure 1: Study overview and design for MR analysis. SNP information for exposures and outcomes were extracted from GWAS summary statistics for each feature. B2 is the causal association of interest (Effect of Biomarkers on seven different subcortical brain structure volumes), estimated using B2 =B1/B3. B1 and B3 are the direct associations of the genetic variants on the exposure (biomarkers) and outcomes (subcortical structures) obtained from the GWAS studies. We also assume that the SNP instrument selected acts on the outcome only through exposure and not through any confounders. IVW: Inverse Variance Weighted.
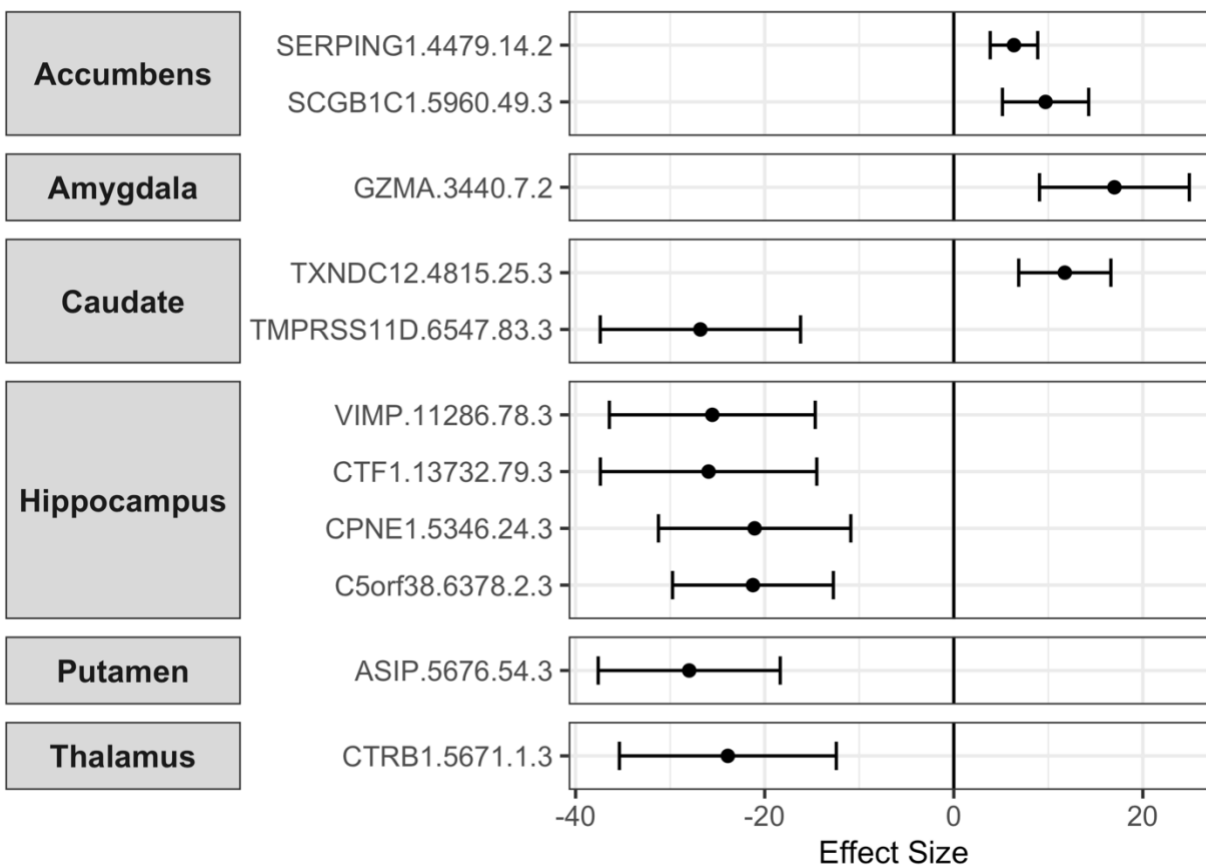
Figure 2: Significant causal associations between plasma proteins and subcortical brain structure volumes as uncovered via MR analysis. The Proteins were the exposures and the subcortical structures' volume as outcomes. The associations were significant after FDR corrections for multiple testing.
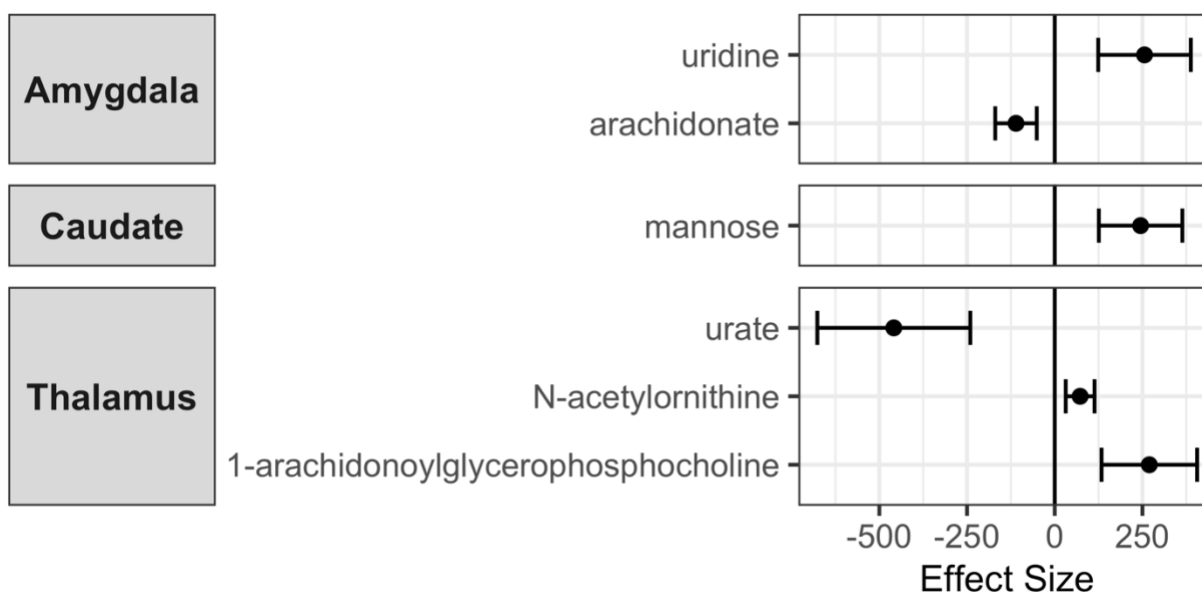
Figure 3: Significant causal associations between metabolites and subcortical brain structure volumes as uncovered via MR analysis. The metabolites were the exposures and the subcortical structures' volume as outcomes. The associations were significant after FDR corrections for multiple testing.
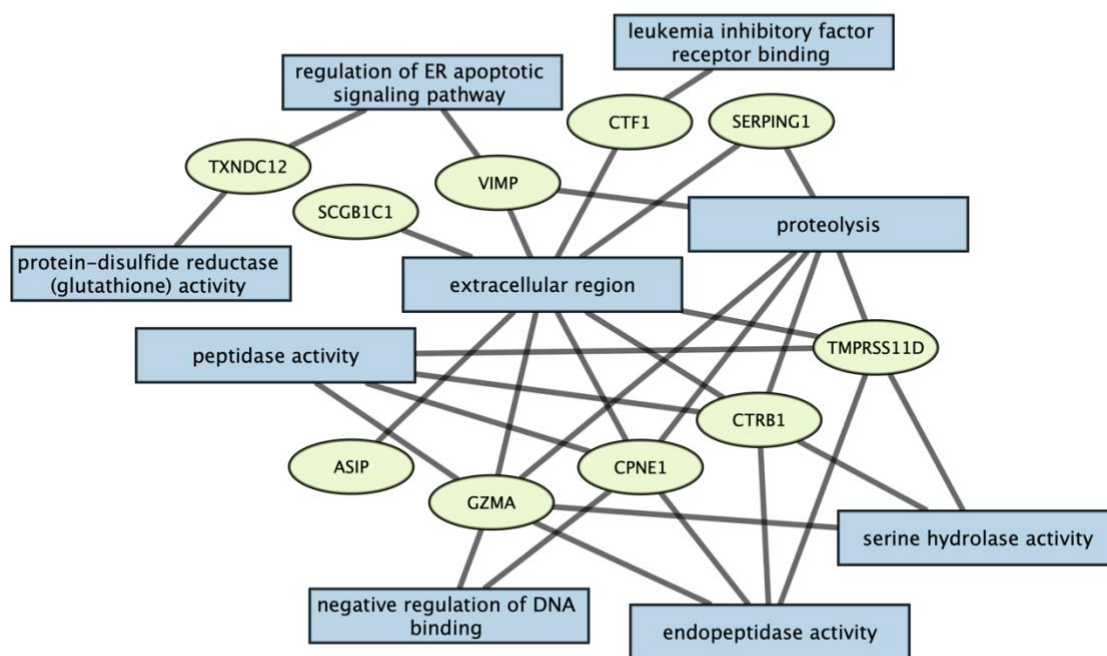
Figure 4: Enrichment analysis of proteins using the g:Profiler tool. The rectangles correspond to the various enriched Gene Ontology terms and the proteins associated with each term are shown in ellipses.