Brain structure and function show distinct relations with genetic predispositions to mental health and cognition

Shu Liu¹, Dirk J.A. Smit¹, Abdel Abdellaoui¹*, Guido A. van Wingen¹*, & Karin J.H. Verweij¹*

¹ Amsterdam UMC, University of Amsterdam, Department of Psychiatry, Amsterdam Neuroscience, Amsterdam, the Netherlands

*These authors jointly supervised the work.

Corresponding authors:
Shu Liu, MSc and Guido van Wingen, PhD
Department of Psychiatry
Amsterdam UMC
University of Amsterdam
Meibergdreef 5
1100 DD Amsterdam
The Netherlands
Email: s.liu@amsterdamumc.nl or g.a.vanwingen@amsterdamumc.nl

Running title: Brain relations to mental health and cognition
Abstract

Background: Mental health and cognitive achievement are partly heritable, highly polygenic, and associated with brain variations in structure and function. However, the underlying neural mechanisms remains unclear.

Methods: We investigated the association between genetic predispositions to various mental health and cognitive traits and a large set of structural and functional brain measures from the UK Biobank (N=36,799). We also applied linkage disequilibrium score regression to estimate the genetic correlations between various traits and brain measures based on genome-wide data. To decompose the complex association patterns, we performed a multivariate partial least squares model of the genetic and imaging modalities.

Results: The univariate analyses showed that certain traits were related to brain structure (significant genetic correlations with total cortical surface area from $r_g = -0.101$ for smoking initiation to $r_g = 0.230$ for cognitive ability), while other traits were related to brain function (significant genetic correlations with functional connectivity from $r_g = -0.161$ for educational attainment to $r_g = 0.318$ for schizophrenia). The multivariate analysis showed that genetic predispositions to attention deficit hyperactivity disorder, smoking initiation, and cognitive traits have stronger associations with brain structure than with brain function, whereas genetic predispositions to most other psychiatric disorders have stronger associations with brain function than with brain structure.

Conclusions: These results reveal that genetic predispositions to mental health and cognitive traits have distinct brain profiles.

Keywords: structural MRI, resting-state, genetics, polygenic risk, mental health, cognition.
**Introduction**

Twin and family studies have revealed that mental health and cognitive traits are partly heritable (1). Genome-wide association studies (GWAS) have identified many single-nucleotide polymorphisms (SNPs) across the whole genome associated with these traits(2). These complex traits are influenced by many genes that individually explain a very small part of the variance, as well as the environment. These genetic effects influence other underlying traits and may also affect which environments people get exposed to(3). The polygenic nature poses a challenge to understanding the genetic architecture and biological mechanisms underlying specific traits(4).

A promising approach that we will explore here is the use of imaging genetics to examine brain-based intermediate phenotypes, also known as endophenotypes(5,6). The effects of genes are not expressed directly at the level of behavior, but are more likely mediated by their molecular and cellular effects on brain development and information processing(7,8). The neural mechanisms through which genetic variants travel to influence one’s cognitive capacity and mental health remain largely unknown.

The cerebral cortex is the outer layer of the brain that contains the neuronal cell bodies, also known as gray matter. It forms a folded sheet that plays a major role in multiple aspects of higher cognitive function. A recent study by Grasby et al. reported that variation in total cortical surface area (CSA) was genetically correlated with cognitive function, depression, and attention deficit hyperactivity disorder (ADHD), but not with many other psychiatric disorders(9). Other studies also found significant genetic associations between total brain volume and cognitive traits including general cognitive function, educational outcomes, intelligence, and numerical reasoning(10,11). For brain function, a recent study detected significant genetic correlations with many complex traits, such as education, cognitive performance, ADHD, major depressive disorder, schizophrenia, sleep, and neuroticism(12). In addition, psychiatric disorders, especially schizophrenia and bipolar disorder, have been consistently found to be genetically
associated with variation in brain function(13–15), but inconsistent results regarding the association between polygenic risk for psychiatric disorders and brain structure were observed in previous studies(16–18). Besides psychiatric disorders and cognitive traits, substance use traits have also been reported to have shared genetic etiology with cortical brain morphology(19).

While studies combining genetic and neuroimaging data have made considerable progress in revealing neural correlates of various traits, there are still some limitations: 1, the sample sizes of previous studies have been relatively limited, resulting in insufficient statistical power; 2, genetic correlations have been used to estimate the magnitude of the overlap between brain measures and complex traits in genetic influences, but have been shown to be less powerful in detecting a genetic association between two traits than methods based on individual-level genotype data(20,21); 3, and importantly, to our knowledge, no studies have combined both structural and functional neuroimaging information jointly with multiple mental health outcomes in multivariate analyses, leaving open whether certain predispositions are predominantly linked to brain structure or function.

In the present study, we systematically investigated the genetic associations of multiple mental health and cognitive traits with brain structure and function. Specifically, we included seven mental health disorders (ADHD, Alzheimer’s disease, autism spectrum disorder (ASD), bipolar disorder, eating disorder, major depression, schizophrenia), five substance use traits (cigarettes per day, smoking initiation, smoking cessation, caffeine consumption, alcoholic drinks per week), and two cognitive traits (cognitive ability and educational attainment). The 14 traits were highly heritable and had large enough GWAS discovery sample(22). Measures of brain structure and function were derived from T1-weighted and resting-state functional MRI. We used both PGS(23) analyses and genome-wide genetic correlations based on GWAS summary statistics(24). In detail, we calculated PGSs for 14 traits which represent the aggregated genetic
effects across the genome, and assessed their associations with structural and functional neuroimaging measures. In addition, we also estimated genetic correlations of the mental health and cognitive traits with neuroimaging features based on genome-wide SNP data using linkage disequilibrium score (LDSC) regression. Finally, we modeled the joint genetic associations of the mental health and cognitive traits with brain structure and function, using a robust partial least squares (PLS) regression(25). To decompose the association patterns of various traits with brain structure and function, PLS regression is used to identify a limited number of latent variables that link the many genotype variables to the many imaging phenotypes.
Methods

Participants

The data used in this study are from UK Biobank(26). We used the latest available imaging release (on January 2020) including 36,799 unrelated European ancestry participants from across the United Kingdom with both genetic and imaging data (19,512 females and 17,287 males, mean age = 54.90 years, SD = 7.43 years, range from 40 to 70 years). The National Health Service North West Centre for Research Ethics Committee provided UK Biobank project with ethical approval (reference: 11/NW/0382) (http://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/brain_mri.pdf).

Imaging procedures

Detailed information about the image acquisition protocols and processing pipeline is provided at: http://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/brain_mri.pdf. For brain structure, there were three global measures (total cortical surface area (CSA), intra-cranial volume (ICV) and average cortical thickness (CT)). Apart from them, we also focused on the regional measures including the volumes of 14 subcortical regions, 66 surface area of parcels identified on the cortical surface, and cortical thickness within these areas. The cortical parcellations were based on the Desikan-Killiany (DK) atlas(27). For brain function, there were amplitude within 21 functional networks and functional connectivities (FCs) between these networks derived from rsfMRI. Additional details on structural and functional measures are reported in the Supplementary Information (SI).

Polygenic scores (PGSs)

We conducted the quality control procedures on genotype data (SI Methods). PGSs were then calculated by summing the number of risk alleles an individual carries weighted by the strength of the association of each allele with a specific trait. The strength of association for each SNP
is based on the effect size estimate from existing GWAS summary statistics. We used the summary-data based best linear unbiased prediction (SBLUP) approach (28) to create PGSs (SI Methods and Table S1). The scores were computed on a set of 1,312,100 autosomal HapMap 3 SNPs with a minor allele count of >5, an info score of >0.3, Hardy–Weinberg equilibrium (HWE) P-value ($P_{HWE}$) < $10^{-6}$ and missingness < 0.05.

Genetic relationships analysis

We used the SciPy Stats package in python to calculate the Pearson correlation coefficients of PGSs for the 14 traits with global brain measures, the volumes of 14 subcortical regions as well as CSA and CT for the 66 DK atlas regions, and brain functional measures which includes amplitude of 21 functional networks and 210 FCs (Figure S1). We regressed out the covariates of the 14 PGSs and all brain measures by the estimation of residuals (Table S2). There were age, sex, age*sex, age$^2$, age$^2$*sex, head positions, the top 25 genetic principal components (PCs), rsfMRI head motions, ICV, total CSA, and average CT, which were consistent with the main covariates used in the GWASs by Smith (29) and Elliott (30). We applied false discovery rate (FDR) multiple testing correction with a significance threshold of 0.0167 (0.05/3) to account for the above three sets of analyses: global structural measures, regional structural measures, or functional measures.

In addition, we conducted the GWASs for brain imaging measures in the UK Biobank sample, and used LDSC regression (24) to calculate their genetic correlations with mental health disorders, substance use, or cognitive traits (SI Methods). Covariates like those applied in the Pearson correlation analysis were regressed out.

Joint partial least squares modeling analysis

To test for differences in relationships of 14 mental health and cognitive traits with brain structure and function, we applied joint partial least squares (PLS) modeling of the genetic and imaging modalities, which find the multidimensional direction in the genetic space that explains
the maximum multidimensional variance direction in the imaging space (31). The detailed procedures of PLS method are represented in **Figure S2**. We performed a canonical symmetric version of the PLS regression from the scikit-learn package in python (32) to find linear relations between two multivariate datasets (Figure 1A). In short, the input features included PGSs for the 14 mental health and cognitive traits as one set of variables and 377 structural and functional measures as the other. PLS modeling produces a weight for each input feature that represents its relative importance for describing the global joint multimodal relationship (25). Comparing the weights helps identify PGSs that are linked to the patterns of brain structure or function. Age, sex, age*sex, age^2, age^2*sex, head positions, and the top 25 genetic PCs were added as covariates.

A robust approach for the stable estimation and interpretation of PLS weights was applied in this study (Figure 1B-D), following a previously described procedure (25). Briefly, a stability selection procedure was conducted to assess the reproducibility and robustness of the PLS parameters. The whole sample was randomly split in half and the PLS model was run on each sampled subgroup independently. Through 5000 repetitions of this procedure, we estimated a confidence measure for each input feature ranging between 0.0 and 1.0, indicating its probability containing highly reproducible PLS weights. Therefore, the measure can be regarded as the importance of input feature. Similar to the study by Lorenzi and Altmann (25), the number of components was also initially set as five. In addition, we used the permutation test (n = 1000) to assess the significance of the estimated selection probability of PGSs and neuroimaging measures through randomly shuffling the PGSs of individuals to conduct the PLS model for 1000 times.
Results

Genetic relationships with global brain structure

The PGSs showed many significant associations with total CSA and ICV, but none with average CT (Table 1, Table S3). Higher polygenic risks for ADHD, Alzheimer’s disease, major depression, cigarettes per day, and smoking initiation were strongly associated with lower total CSA. Conversely, higher PGSs for smoking cessation, alcoholic drinks per week, cognitive ability, and educational attainment were strongly associated with higher total CSA. Similarly, higher polygenic risks for ADHD, smoking initiation, and Alzheimer’s disease were significantly associated with lower ICV, whereas higher PGSs for cognitive ability, educational attainment, and eating disorder were significantly associated with higher ICV. Polygenic risks for ASD, bipolar disorder, schizophrenia and caffeine consumption were not significantly associated with any global measure of brain structure.

LDSC regression analysis showed the highest SNP-based heritability estimates for total CSA (37.8%) and ICV (35.4%) (Table S4, Figure S3). Moreover, the results of the genetic correlation analysis for global structural measures were largely in line with those of the PGS analysis (Table S5). Total CSA and ICV showed highly significant genetic correlations with ADHD ($r_g = -0.170$ and $r_g = -0.159$), cognitive ability ($r_g = 0.230$ and $r_g = 0.228$) and educational attainment ($r_g = 0.211$ and $r_g = 0.159$), but no significant correlations were observed with average CT. Most psychiatric disorders (ASD, bipolar disorder, eating disorder, schizophrenia) were not genetically correlated with measures of global brain structure.

Genetic relationships with regional brain structure

We found six significant associations for subcortical volumes controlling for ICV (Figure S4, Table S6). Notably, volume of the thalamus was negatively associated with genetic vulnerability to schizophrenia and positively with genetic predisposition to educational
attainment. In addition, the PGS for major depression was positively associated with volume of the left caudate nucleus, the PGS for Alzheimer’s disease was negatively associated with volume of the left putamen, and the PGS for cigarettes per day showed negative associations with volume of the right putamen. As for genetic correlation analysis, we only found a significant correlation between volume of the thalamus and schizophrenia ($r_g = -0.135$) (Figure S5, Table S7).

For CSA of brain regions, a total of 27 significant associations were found with the 14 PGSs controlling for total CSA, which were mainly identified on educational attainment, smoking initiation and cognitive ability (Figure 2A-B, Table S8, SI Results). The genetic correlation analysis found largely comparable results though with fewer significant associations, with significant genetic correlations of cognitive ability and educational attainment with CSA of the pars orbitalis of the inferior frontal gyrus ($r_g = 0.147$ and $r_g = 0.170$) (Figure S6, Table S9). Although the global measure of CT did not show any significant associations with PGSs, we found 27 significant associations with regional CT controlling for average CT (Figure 2C-D, Table S10, SI Results). Similarly, the genetic correlation analysis showed a negative correlation of educational attainment with CT of the inferior parietal cortex and two positive correlations with CT of the superior temporal regions. (Figure S7, Table S11).

Associations without correction for global brain structure are presented in the supplementary materials (Figures S8-10, Tables S12-14). Those results showed considerable overlap with those for global brain structure.

**Genetic relationships with brain function**

We examined the associations between the 14 PGSs and these 231 functional measures. We found 22 significant associations between PGSs and amplitude within 21 networks (Figure 3, Figure S11, Table S15, SI Results). Most of them were identified on bipolar disorder,
schizophrenia, major depression, and educational attainment. Although the pattern of results was comparable, the genetic correlation analysis showed no significant associations between network amplitudes and the 14 mental health and cognitive traits (Figure S12, Table S16). In addition, a total of 133 significant associations between PGSs and FCs were observed (Figure 3, Table S17, SI Results), which were also mainly identified on educational attainment, schizophrenia, major depression and ASD. The genetic correlation analysis only showed a few significant genetic correlations between FCs and schizophrenia, cognitive ability, and educational attainment (Figure S13, Table S18).

Multivariate modeling of joint relationships between PGSs and brain measures

The univariate analyses reported above showed complex genetic relationships of multiple traits with brain structure and function. In order to decompose the association patterns of various traits with brain structure and function, we performed a data-driven multivariate statistical analysis of the genetics and imaging data with split-half cross-validation on basis of partial least squares (PLS) regression(25) (Figure 1). We focused on the first five components which ensured that we can get stable and robust latent variables(25). This method with inherent replication identifies multivariate components that link multiple traits with multiple brain imaging variables.

Figure 4A and 4B respectively show the mean weights for the first PLS genotype and phenotype component, which reflect the importance of each input variable for the PGSs-imaging relationships in the PLS model. Figure 4C and 4D show the probabilities for PGSs and imaging measures receiving the largest PLS weights in two random groups across 5000 repetitions. PGSs for ADHD, educational attainment, smoking initiation, and cognitive ability had higher weights and were consistently selected for the first genotype component (selection probability > 50%). The corresponding imaging component was mainly associated with larger
CSA of many cortical regions at the threshold of 50% selection probability, especially the frontal and temporal regions, as well as larger volume of the thalamus, whereas regional CT and functional measures had low weights and were not consistently selected. Similarly, Figure 4E-H show the weights and probabilities for the second PLS genotype and phenotype component. The genetic patterns of PGSs for psychiatric disorders (schizophrenia, major depression, bipolar disorder, and ASD) corresponded consistently to lower amplitude of 5 functional networks and to 10 both stronger and weaker FCs at the threshold of 50% selection probability. All structural measures had low weights and were not consistently selected. The selected nodes (resting-state networks) included the visual system (node 2 and 4), somatosensory system (node 10), cerebellum (node 15), and basal ganglia (node 18). FCs included 5 positive and 5 negative associations across the entire brain, including networks often associated with psychiatric disorders such as the default mode network, salience network, and ventral and dorsal attention networks. Through the permutation test (n = 1000) on selection probabilities, the selected PGSs and brain measures in the first two components were significantly different from chance (Figure S14 and Table S19).

The first two components explained most of the total variance, 52.4% and 11% respectively, and the PGSs and brain measures could be stably and consistently selected across 5000 repetitions compared to other components, and were stable across a range of thresholds (SI Results).
Discussion

We systematically investigated the genetic relationships of the 14 mental health and cognitive traits with 380 structural and functional brain measures using neuroimaging and genetic data from 36,799 participants. The results show that genetic predispositions to ADHD, educational attainment, smoking initiation, and cognitive ability are more strongly associated with brain structure than with brain function, whereas genetic predispositions to most psychiatric disorders (schizophrenia, major depression, bipolar disorder, and ASD) are more strongly associated with brain function than with brain structure. To our knowledge, this is the first study providing evidence for these differential associations, and suggests that there are distinct neural pathways through which genes eventually influence different mental health and cognition traits.

The univariate correlation analyses showed various and complex associations between the 14 PGSs and 380 brain measures. We also estimated the SNP-based genetic correlations between these traits and neuroimaging features and found generally consistent results, although fewer associations were significant due to lower power of these analyses. In order to decompose the joint genetic association patterns into underlying components, we performed a data-driven multivariate statistical analysis (PLS regression) of the genetics and imaging data. The first PLS component confirmed that ADHD was more strongly associated with brain structure than with brain function. In fact, previous studies consistently reported significant genetic sharing between ADHD and total CSA and ICV(9,33,34), whereas only a few FCs were significantly associated with the PGS for ADHD(35). Epidemiologic and clinical studies have found that genetic risk factors that affect the structure and functional capacity of brain networks are involved in the etiology of ADHD(36,37). Many genes that have been associated with ADHD, such as FOXP2, SORCS3, and DUSP6, are highly expressed in the brain and implicated in neurodevelopmental processes(36).
Comparable results were obtained for smoking initiation, which showed a similar association pattern with the brain measures as ADHD. Although the molecular mechanisms underlying these associations have yet to be elucidated, a previous gene set analysis of enriched biological pathways provided some possible clues that brain dopamine receptor signaling is associated with smoking initiation(38). For example, BDNF regulates synaptic plasticity and survival of cholinergic and dopaminergic neurons(39). Genetic variants within the BDNF gene were found to be genome-wide significantly associated with smoking initiation(40) and the gene was highly expressed in the prefrontal cortex, which might be implicated in the cognitive-enhancing effects of nicotine(41). This is in line with our finding that the PGS for smoking initiation was especially associated with the CSA of several frontal regions.

In addition, genetic predispositions to cognitive traits were significantly associated with both brain structure and function, though the first PLS component showed that the association with brain structure was stronger. Our results for cognitive traits are in line with previous studies in which the PGS for educational attainment was associated with larger brains(42,43) and CSA and CT in some frontal and temporal regions mediated the association between the PGS for intelligence and the g-factor(44), but no large-scale study provided credible and sufficient evidence on relationships of PGSs for cognitive traits with brain function. Previous studies have presented some possible explanations for the associations, describing that biological annotation of genes associated with educational attainment were involved in brain-development processes and neuron-neuron communication(45,46).

The second PLS component shows that PGSs for most psychiatric disorders (schizophrenia, major depression, bipolar disorder, and ASD) were more strongly associated with brain function than with brain structure. Previous studies for the genetic relationships between psychiatric disorders and global brain structure produced mixed results, likely because of limited sample sizes. For example, some studies reported that a decrease in total brain volume
or cortical thickness was significantly associated with an increased polygenic risk for schizophrenia(17,47), whereas another study with a relatively large sample found no significant associations(16). With respect to brain function, previous studies indicated that PGSs for ASD, bipolar disorder, major depression, and schizophrenia were associated with functional measures(6,48,49). Overall, although our results do support a limited link between PGSs for psychiatric disorders and brain structure, results from our multivariate analysis highlights that the association with brain function is stronger when both are considered together. The stronger associations between PGSs and brain function may indicate that genetic predispositions to psychiatric disorders are likely to cause difficulties in information integration across the brain, which in turn makes individuals more prone to developing a psychiatric disorder(13).

Several interesting genetic relationships with different brain measures were identified in this study. For example, higher PGS for schizophrenia was significantly associated with lower CT of orbitofrontal regions. We also found that PGSs for ASD, bipolar disorder, and major depression showed positive associations with CSA of cingulate regions. The orbitofrontal cortex and cingulate cortex are involved in decision making, emotion regulation, response to reward, learning and memory(50–52). For substance use traits, we found that smoking and drinking behaviors were genetically associated with prefrontal brain structure, which are involved in self-control and risk-taking (53).

The heritability estimates (see Figure S3) differed for the various brain measures. It was highest for total CSA and ICV and somewhat lower for average CT, which may have contributed to the significant associations of PGSs with total CSA and ICV but not average CT. Also the regional measures showed differences in heritability, but there was no apparent relationship between these heritability estimates and the likelihood for significant associations of PGSs. Instead, the distinct PLS components suggest that different PGSs are related to distinct brain measures.
We note several limitations to our study. First, the predictive power of PGSs and the significance of the genetic correlations (LDSC regression) depend on the sample size of the discovery GWAS summary statistics. Thus, the statistical power differs substantially between different traits, which may influence the comparisons of association patterns. Second, our results point to genetic associations between different traits and brain measures, but the nature of the association is not yet known. We cannot distinguish genetic pleiotropy versus causal relationships from mental health and cognitive traits to brain structure and function or vice versa. The general assumption is that the brain is the mediator through which genes exert their effects on behavior. But it is also possible that genetic variants simultaneously influence brain measures and behaviour (i.e. pleiotropy) and there are also genetic influences on the environment and continuous interactions between the brain, behavior and the environment. The relation between genes and the brain may therefore also be influenced by genetic effects on the environment that subsequently shape the brain(54,55). Third, the polygenic scores may contain some environmental effects, because genetic effects may possibly act via the environment to influence the brain. These environmental effects can include the influence of parents who provide the offspring with their genome as well as with their rearing environment(56), which can result in an inflation of estimated genetic effects. Gene-environment correlations can also occur on a regional level: the educational attainment polygenic score, for example, has been shown to influence where people live (individuals with a higher polygenic score are more likely to migrate to a richer neighborhood), which may have an effect on mental health and cognitive outcomes(22). Fourth, all participants in UK biobank are 40+ years old and most of them are healthy, so whether our findings can be generalized to younger population as well as individuals with mental disorders needs further verification. Fifth, the functional measures were derived from resting-state fMRI, and the genetic associations may be different for task-based fMRI that can probe specific brain functions.
In summary, this study provides evidence for distinct association patterns of genetic predispositions to mental health and cognition with brain structure and function, with genetic predispositions to ADHD, smoking initiation, and cognitive traits having stronger associations with brain structure than with brain function, whereas genetic predispositions to most other psychiatric disorders having stronger associations with brain function than with brain structure. These results bring us closer to understanding the neurobiological basis of these traits.
Acknowledgments and Disclosures

This research has been conducted using data from UK Biobank under application numbers 40310 and 30091. S.L. is supported by the China Scholarship Council. A.A. & K.J.H.V. are supported by the Foundation Volksbond Rotterdam.

A previous version of this article was published as a preprint on medRxiv:

https://www.medrxiv.org/content/10.1101/2021.03.07.21252728v1.

The authors declare no biomedical financial interests or potential conflicts of interest.
References


schizophrenia and bipolar disorder. https://doi.org/10.1007/s11682-018-9879-z


https://doi.org/10.1177/1745691619867107

https://doi.org/10.1126/science.aan6877
Table 1. Correlations of 14 PGSs for mental health and cognitive traits with three global structural brain measures

<table>
<thead>
<tr>
<th>Traits</th>
<th>Total cortical surface area (CSA)</th>
<th>Average cortical thickness (CT)</th>
<th>Intra-cranial volume (ICV)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
<td>p</td>
<td>p_FDR</td>
</tr>
<tr>
<td>ADHD</td>
<td>-0.0587</td>
<td>1.68E-29</td>
<td>2.36E-28</td>
</tr>
<tr>
<td>Alzheimer’s disease</td>
<td>-0.022</td>
<td>2.41E-05</td>
<td>1.01E-04</td>
</tr>
<tr>
<td>ASD</td>
<td>0.0035</td>
<td>0.498</td>
<td>0.675</td>
</tr>
<tr>
<td>Bipolar disorder</td>
<td>0.0136</td>
<td>9.33E-03</td>
<td>2.24E-02</td>
</tr>
<tr>
<td>Eating disorder</td>
<td>0.0081</td>
<td>0.119</td>
<td>0.2</td>
</tr>
<tr>
<td>Major depression</td>
<td>-0.0196</td>
<td>1.74E-04</td>
<td>6.08E-04</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>-0.0092</td>
<td>7.76E-02</td>
<td>0.136</td>
</tr>
<tr>
<td>Cigarettes per day</td>
<td>-0.0183</td>
<td>4.40E-04</td>
<td>1.32E-03</td>
</tr>
<tr>
<td>Smoking initiation</td>
<td>-0.0378</td>
<td>4.22E-13</td>
<td>3.55E-12</td>
</tr>
<tr>
<td>Smoking cessation</td>
<td>0.0217</td>
<td>3.07E-05</td>
<td>1.17E-04</td>
</tr>
<tr>
<td>Caffeine consumption</td>
<td>-0.003</td>
<td>0.56924</td>
<td>0.703</td>
</tr>
<tr>
<td>Alcoholic drinks per week</td>
<td>0.0191</td>
<td>2.48E-04</td>
<td>8.02E-04</td>
</tr>
<tr>
<td>Cognitive ability</td>
<td>0.0289</td>
<td>2.97E-08</td>
<td>1.56E-07</td>
</tr>
<tr>
<td>Educational attainment</td>
<td>0.0835</td>
<td>5.43E-58</td>
<td>1.14E-56</td>
</tr>
</tbody>
</table>

r = Pearson’s correlation, p = p-value; p_FDR = p-value after FDR multiple correction. Bold values indicate significant associations after FDR multiple correction testing (p_FDR < 0.0167).
**Figure 1.** Flow chart of data-driven multivariate statistical analysis. A, input features of PLS regression. $w_i$ indicated the weight of i-th input feature, which was served as the importance of relative importance for describing the multimodal relationship. $T_i$ and $U_i$ respectively indicated the i-th genotype and phenotype PLS component transformed from genetic and imaging data. B, the dataset was randomly divided into two groups for validation in which PLS analyses were performed independently. This procedure was repeated 5000 times. The averaged results from the 5000 iterations identified the robust and consistently selected PGSs and brain imaging measures that received the largest weights in both groups for each iteration. C, for each repetition, the absolute weights of 14 PGSs were sorted, and the top 5 PGSs were selected in two random groups. The selected PGSs were labeled one or zero otherwise. The resulting binary arrays independently estimated in the two groups were then merged, such that only replicated PGSs receiving the largest PLS weights were selected. D, for each repetition, the absolute weights of brain measures were sorted, and the top 10% brain measures are selected in two random groups. Similarly, only replicated brain measures receiving the largest weights were selected.

**Figure 2.** Associations between the 14 PGSs and the CSA and CT of 66 cortical regions, corrected for total CSA and average CT, respectively. The horizontal axis shows the different brain regions and the vertical axis shows the various traits. The size and color of each box indicate the magnitude of the association. Blue squares indicate positive associations, and red squares indicate negative associations. A, the PGS-CSA correlation map across the left hemisphere. B, the PGS-CSA correlation map across the right hemisphere. C, the PGS-CT correlation map across the left hemisphere. D, the PGS-CT correlation map across the right hemisphere.

**Figure 3.** Significant associations between the 14 PGSs and functional measures. The nodes in each circle show the 21 resting-state networks. Asterisks indicate significant associations with network amplitude (temporal fluctuation) of the 21 networks after FDR multiple testing correction. The lines between the nodes indicate significant FCs between the 21 functional networks after FDR multiple testing correction. The color of each line and node indicates the magnitude of the associations. Blue squares indicate positive associations, and red squares indicate negative associations.
**Figure 4.** First two components of the PLS model. A, the mean weights across 5000 repetitions for 14 PGSs in the first genotype component. B, the mean weights across 5000 repetitions for the CSA of cortical regions and volumes of subcortical regions in the first phenotype component. C, probabilities of the 14 PGSs of being selected across 5000 repetitions in the first genotype component. D, probabilities of regional CSA and subcortical volumes of being selected across 5000 repetitions in the first phenotype component. E, the mean weights across 5000 repetitions for 14 PGSs in the second genotype component. F, the mean weights across 5000 repetitions for amplitude of 21 functional networks and 210 FCs between these networks in the second phenotype component. G, probabilities of the 14 PGSs of being selected across 5000 repetitions in the second genotype component. H, probabilities of amplitude of 21 functional networks and 210 FCs between these networks of being selected across 5000 repetitions in the second phenotype component. In the bar plots, orange bars indicate PGSs with negative PLS weights, and green bars indicate PGSs with positive PLS weights.
**A**

PLS regression

- X: PGSs for mental health and cog
- Structural and functional measures

**B**

Randomly sampling the whole data in two nonoverlapping groups (split-half) and running PLS model independently

**C**

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Genotype" /></td>
<td></td>
</tr>
</tbody>
</table>

- Absolute PLS weights associated with 14 PGSs
- Select top 5 PGSs
- Overlapping results

**D**

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image2" alt="Phenotype" /></td>
<td></td>
</tr>
</tbody>
</table>

- Absolute PLS weights associated with brain structural and functional measures
- Select top 10% measures
- Overlapping results
The associations between PGs and CSA of left cortical regions

The associations between PGs and CSA of right cortical regions

The associations between PGs and CT of left cortical regions

The associations between PGs and CT of right cortical regions