Journal of Translational Medicine

RESEARCH

Open Access

Machine learning based metabolomic and genetic profiles for predicting multiple brain phenotypes

Xueli Zhang[1,5,12†], Yu Huang[1,2†], Shunming Liu[1], Shuo Ma[6], Min Li[5], Zhuoting Zhu[1,2,3], Wei Wang[7], Xiayin Zhang[1,2], Jiahao Liu[3], Shulin Tang[1], Yijun Hu[1], Zongyuan Ge[8], Honghua Yu[1*], Mingguang He[1,9,10,11*] and Xianwen Shang[1,2,3,4,9,10*]

Abstract

Background It is unclear regarding the association between metabolomic state/genetic risk score(GRS) and brain volumes and how much of variance of brain volumes is attributable to metabolomic state or GRS.

Methods Our analysis included 8635 participants (52.5% females) aged 40–70 years at baseline from the UK Biobank. Metabolomic profiles were assessed using nuclear magnetic resonance at baseline (between 2006 and 2010). Brain volumes were measured using magnetic resonance imaging between 2014 and 2019. Machine learning was used to generate metabolomic state and GRS for each of 21 brain phenotypes.

Results Individuals in the top 20% of metabolomic state had 2.4–35.7% larger volumes of 21 individual brain phenotypes compared to those in the bottom 20% while the corresponding number for GRS ranged from 1.5 to 32.8%. The proportion of variance of brain volumes (R [2]) explained by the corresponding metabolomic state ranged from 2.2 to 19.4%, and the corresponding number for GRS ranged from 0.8 to 8.7%. Metabolomic state provided no or minimal additional prediction values of brain volumes to age and sex while GRS provided moderate additional prediction values (ranging from 0.8 to 8.8%). No significant interplay between metabolomic state and GRS was observed, but the association between metabolomic state and some regional brain volumes was stronger in men or younger individuals. Individual metabolomic profiles including lipids and fatty acids were strong predictors of brain volumes.

Conclusions In conclusion, metabolomic state is strongly associated with multiple brain volumes but provides minimal additional prediction value of brain volumes to age + sex. Although GRS is a weaker contributor to brain

†Xueli Zhang and Yu Huang contributed equally to this work.

*Correspondence:
Honghua Yu
yuhonghua@gdph.org.cn
Mingguang He
mingguang.he@polyu.edu.hk
Xianwen Shang
andy243@126.com

Full list of author information is available at the end of the article

volumes than metabolomic state, it provides moderate additional prediction value of brain volumes to age + sex. Our findings suggest metabolomic state and GRS are important predictors for multiple brain phenotypes.

Key points

- Question: What is the relationship between metabolomic/genetic profiles and brain volumes?
- Finding: Metabolomic profiles show a robust association with various brain volumes, while genetic profiles have a less significant impact on brain volumes.
- Meaning: This study elucidated the landscape of the relationship between metabolomic/genetic profiles and brain volumes.

Keywords Metabolomic profiles, Metabolomic state, Genetic risk score, Brain phenotype, Prediction value, Moderation analysis

Introduction

The importance of brain structure on behaviors such as personality traits and intelligence and diseases especially psychosocial and neurological disorders has been investigated in an increasing number of studies [1, 2]. Non-pathological brain damage may be a prelude to dementia [3], and some other neurodegenerative disorders [4]. Brain atrophy is a public health challenge in the global ageing population as older age is the most important risk factor for brain atrophy [4, 5]. Although numerous chronic conditions and biomarkers have been linked to brain atrophy [5–8], these previous studies failed to identify determinants (besides age) with great contribution to brain atrophy. Therefore, it is imperative to identify new important determinants for brain atrophy.

Cohort studies have shown that metabolic disorders including diabetes, hypertension, and obesity were associated with greater brain atrophy [5, 6, 9, 10]. Data from three independent cohorts showed that higher glucose levels and lower small high-density lipoprotein (HDL) were associated with brain atrophy [11]. In contrast, a recent study of 9290 individuals from 15 populations investigated the association between individual metabolomic profiles and white matter hyperintensities (WMH) and found that lipids and amino acids might play an important role in WMH variance [12]. Buergel et al. applied machine learning to generate metabolomic state by involving 168 metabolomic profiles, which was strongly predictive of multiple common diseases [13]. However, the potential of metabolomic profiles as a whole state in the prediction of brain atrophy has not been investigated thus far.

Although genome-wide association studies (GWAS) of brain imaging phenotypes have been conducted based on the UK Biobank cohort [14], it is unknown regarding the proportion of variance in brain volumes explained by genetic risk score (GRS). Machine learning-based GRS has been demonstrated to have better performance in disease risk prediction [15]. A GRS has not been established for brain volume using machine learning techniques. It is unknown regarding the interaction between metabolomic state and GRS for brain volumes.

Using the UK Biobank, we aimed to examine associations of metabolomic states with brain volumes and to establish GRS for brain volumes using machine learning. We also tested whether metabolomic states and GRS provided additive values for the prediction of brain volumes compared with age + sex, multimorbidity and/or multiple risk factors. The interplay between metabolomic states and GRS for brain volumes was then examined.

Methods

Study population

The UK Biobank is a population-based cohort of more than 500,000 participants aged 40–70 years at enrolment [16]. Baseline data (2006–2010) was collected from 502,505 participants out of approximately 9.2 million invited people. These participants attended one of the 22 assessment centers throughout the UK. The study design has been shown in detail elsewhere [16].

The UK Biobank Study's ethical approval has been granted by the National Information Governance Board for Health and Social Care and the NHS North West Multicenter Research Ethics Committee (REC reference: 16/NW/0274). All participants provided informed consent through electronic signature at recruitment.

Brain magnetic resonance imaging

Brain magnetic resonance imaging (MRI) data collected between August 2014 and October 2019 were used in the analysis. Images were generated using a standard Siemens Skyra 3T scanner with a standard 32-channel radio-frequency receiver head coil [17, 18]. The T1- and T2-weighted scans with the Functional MRI of the Brain Software Library were used to estimate brain volumes. Total brain volume was the sum of the grey and white matter volumes. Volumes of total brain, grey matter, white matter, peripheral cortical grey matter, and ventricular cerebrospinal fluid were normalised for head size [17, 18], using the ratio-corrected method [17]. Considering the positively skewed distribution, WMH was

log-transformed in the analysis. Volumes of the hippocampus, thalamus, caudate, putamen, pallidum, amygdala, and accumbens on both right and left sides were calculated. Larger volumes of total and regional brain but smaller WMH load represent better brain health. Normalization of each brain phenotype was conducted to result in it being Gaussian distributed, with mean = 0 and SD = 1.

Metabolomic profiling

A high-throughput NMR-based metabolic biomarker profiling platform was used to measure metabolomic profiles [19]. EDTA plasma samples at baseline were collected from a randomly selected subset of 117,121 UK Biobank participants. Levels of 249 metabolic traits (168 concentrations and 81 ratios) including the lipoprotein lipids in 14 subclasses, fatty acids, amino acids, ketone bodies, and glycolysis were quantified in the study (Table S1). Quality control was performed to eliminate systemic and technical variance (https://biobank.ctsu.ox.ac.uk/crystal/label.cgi? id = 220). Sample collection, metabolomic quantification, and quality control were conducted according to protocols [19, 20]. Metabolite levels were normalized with mean = 0 and SD = 1.

Genetic data

BiLEVE Axiom array, or the UK Biobank Axiom array was used for genotyping by Affymetrix. Linear association tests on the samples were performed between each of the 17,103,079 genetic variants and each of the 21 brain phenotypes. We included these 21 brain phenotypes in the analysis because these brain volume measurements have been demonstrated to be reliable [21] and have been linked to behavioral and psychological symptoms of dementia in previous studies [22]. Single nucleotide polymorphisms (SNPs) with significant associations were then selected to create GRS for individual brain phenotypes. Bgenie software was used to conduct the GWAS and record the effect sizes (beta), standard errors, and −log10 (P-value) values for the associations. (R1.1)

Covariates

Geographic information on age, sex, ethnicity, education, and income was self-reported using a questionnaire. Townsend index of material deprivation was used to represent neighbourhood-level socioeconomic status. Lifestyle factors including diet, smoking, sleep duration, and alcohol consumption were assessed using a questionnaire on a touch-screen computer. A diet score was calculated based on seven commonly eaten food groups with a higher score representing a healthier diet [23]. Levels of physical activity during work and leisure time were assessed using a short form of the International Physical Activity Questionnaire. The use of medications for antihypertension, lipid-lowering, and glucose-lowering was self-reported.

A multimorbidity score for brain atrophy was computed based on the association between individual major chronic diseases and brain volumes [24]. The multimorbidity score has been shown to be a strong predictor of brain atrophy, independent of age. Body mass index (BMI) was calculated based on measured weight and height. Other chronic diseases were defined using self-reported data, interviews, or inpatient data. (R1.2)

Statistical analysis

Baseline characteristics were expressed as frequency (percentage) or means ± standard deviations (SDs). T-test for continuous variables and Chi-square test for categorical variables were used to test the difference of characteristics by sex.

Metabolomic state model development

For model development and testing, we split the dataset into 18 spatially separated partitions based on the location of the assessment center at enrolment. Of the 22 assessment centers, 17 with more than 150 available participants were divided into 17 separated partitions and 5 with < 150 available participants were combined as one separated partition.

The data was analyzed in 18-fold nested cross-validation, setting aside one of the spatially separated partitions as a test set. Among the remaining partitions, 50% were randomly selected as training data and the other 50% as validation data [25]. Within each of the 18 cross-validation loops, the individual test set was not involved in the model development and the validation data was used to validate the prediction performance. Metabolomic state score was created for each of the 18 test sets based on the corresponding obtained model. The score for all participants was then developed by aggregating predictions of 18 test sets. This analysis was conducted for each of the 21 brain phenotypes (Fig. 1).

Four machine learning models including general linear regression, random forest, gradient boosting machine, and deep learning were used to develop the prediction models. The one with the best prediction performance in the validation analysis was used to create metabolomic state.

Genetic risk score

The same method for metabolomic state development was used to develop a GRS for each of 21 brain phenotypes. SNPs with P-values $< 5\times10^{-5}$ as tested in the GWAS analysis were used to develop the GRS (Fig. 1).

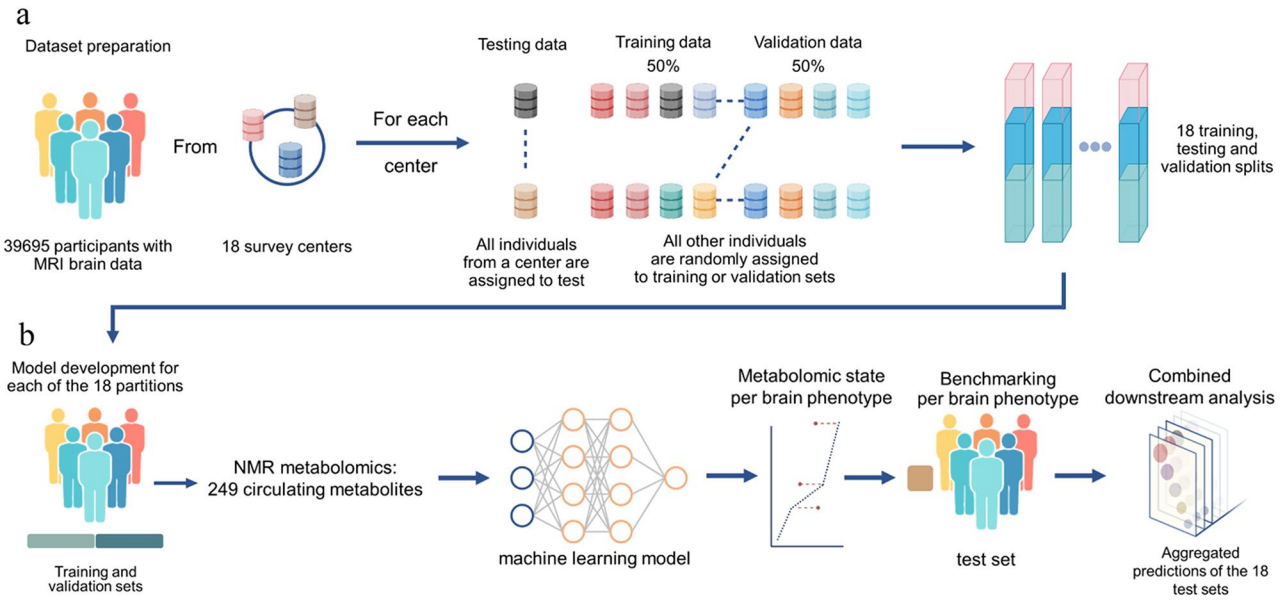

Fig. 1 Flowchart for the development of metabolomic state and genetic risk score for brain volumes using machine learning. **a**. Dataset preparation and model construction. **b**. Model development

Metabolomic state/Genetic risk score and brain volumes

The association between metabolomic state/GRS and brain volumes was examined using general linear regression models. The percentage of the total variance in brain volumes (R [2]) attributable to metabolomic state/GRS as well as age+sex, multimorbidity score, and Panel was analyzed. Panel predictors included age, sex, education, and vascular risk factors [6]. Additional prediction values in addition to these traditional predictor sets by metabolomic state/GRS were then calculated.

Associations between individual metabolomic profiles and brain volumes were also tested using general linear regression models. Benjamin-Hochberg's procedure was used to control the false discovery rate (FDR) at a 5% level for multiple comparisons [26].

Whether associations between metabolomic state/GRS and brain volumes were modified by age or sex was tested using general linear regression models. The interaction between metabolomic state and GRS for brain volumes was also tested.

Multiple imputations for missing data were conducted, and age, sex, and all covariates were included in the imputation models to create 10 imputed datasets.

Data analyses were conducted using SAS 9.4 for Windows (SAS Institute Inc.) and all P values were two-sided with statistical significance set at <0.05.

Results

Our study utilized the UK Biobank dataset to explore the correlations between metabolomic profiles and brain volumes, and to construct GRS for individual brain volumes using machine learning algorithms. We assessed the additional predictive value of metabolomic profiles and GRS in forecasting brain volumes, considering factors such as age, sex, multimorbidity, and other risk factors.

The cohort consisted of 8,635 participants, selected from the UK Biobank dataset. Metabolomic profiles were obtained through NMR scans during the baseline assessment period (2006–2010), and brain volumes were measured using magnetic resonance imaging scans from 2014 to 2019. Machine learning techniques were employed to derive metabolomic states and GRS for 21 distinct brain phenotypes. The study design is illustrated in Fig. 1 (R2.1).

Population selection

Individuals with no data on MRI assessments ($n=462{,}809$), or metabolomic profiles ($n=30635$), or those of non-European ancestry ($n=277$), or those with neurological disorders ($n=149$) were excluded from the analysis. We included 8635 participants (52.5% females) in the final analysis. They were aged 40–70 (mean±SD: 54.9±7.5) years when lifestyle, biomarkers, and metabolomic profiles were assessed at baseline and aged 44–81 (63.6±7.6) years when brain MRI assessments at a repeat visit were conducted.

Men were more likely to have higher household income and to be older and current smokers than women. Women had higher diet quality and lower BMI than men (Table 1).

Metabolomic state and brain volumes

The prediction performance for machine learning models is shown in Table S2. General linear regression had the

Table 1 Baseline characteristics in women and men

	Women	Men	*P*-value*
Age (years)	63.0 ± 7.4[†]	64.4 ± 7.6	< 0.0001
Education			0.0565
0–5 years	265 (5.8)	271 (6.6)	
6–12 years	2211 (48.8)	1858 (45.3)	
≥13 years	2054 (45.4)	1976 (48.1)	
Household income (pounds)			< 0.0001
<18,000	576 (12.7)	381 (9.3)	
18,000–30,999	1094 (24.2)	848 (20.7)	
31,000–51,999	1405 (31.0)	1283 (31.3)	
52,000-100,000	1154 (25.5)	1263 (30.8)	
>100,000	301 (6.6)	330 (8.0)	
Diet score[†,‡]	4.37 ± 1.31	3.69 ± 1.38	
Physical activity (MET-minutes/week)[†]	2420.2 ± 2058.6	2553.8 ± 2394.7	0.0053
Smoking			< 0.0001
Never	2927 (64.6)	2332 (56.8)	
Former	1380 (30.5)	1481 (36.1)	
Current	223 (4.9)	292 (7.1)	
Sleep duration (hours/day)[†]	7.20 ± 1.08	7.15 ± 0.97	0.0403
Alcohol consumption			< 0.0001
Never	124 (2.7)	61 (1.5)	
Previous	92 (2.0)	83 (2.0)	
Current	4314 (95.3)	3961 (96.5)	
BMI (kg/m^2)	26.11 ± 4.57	27.13 ± 3.84	< 0.0001

*T-test for continuous variables and Chi-square test for categorical variables were used to analyze the difference between women and men

[†]Data are means ± SDs. Others are frequency (percentage)

[‡]Diet score was computed based on seven commonly eaten food groups following recommendations on dietary priorities for cardiometabolic health with a higher score representing healthier diet quality

highest performance and was used to fit the metabolomic state.

The association of metabolomic state and brain volumes has been presented in Fig. 2.

As shown in Fig. 2a, the volume of total brain and regional areas increased with the increasing metabolomic state. The association between metabolomic state and grey matter volume was non-linear, whereas all other associations appeared to be linear.

Individuals in the top 20% of metabolomic state (53.8 (95% CI: 52.7, 54.8) ml) had 35.7% larger volume of ventricular cerebrospinal fluid compared with those in the bottom 20% (39.6 (95% CI: 38.8, 40.4) ml), indicating that the different states of metabolomic had significantly different brain volumes, which added evidence to the potential of metabolomic profiling in predicting brain health. For brain stem+4th ventricle, individuals in the top 20% of metabolomic state (24.1 (95% CI: 24.0, 24.3 ml) had 11.0% larger volume than those in the bottom 20% (21.7 (21.6, 21.9 ml). Larger volumes of total brain, grey matter, and other regional areas were also observed among individuals in the top 20% than in the bottom 20% of metabolomic state. Individuals in the top 20% metabolomic state (8.21 (95% CI: 8.16, 8.27 ml) had 5.3% higher load of WMH compared with those in the bottom 20% (7.79 (95% CI: 7.73, 7.86 ml) (Fig. 2b). (R1.3+2.1+2.2+R3.1)

Although the adjustment for age and sex substantially attenuated the association between metabolomic state and brain volumes, most of these associations remained significant (Figure S1).

Genetic risk score and brain volumes

As shown in Table S3, general linear regression had the highest prediction performance was used to generate GRS for individual brain volumes (R3.1).

A higher genetic risk score was associated with larger brain volumes. Individuals in the top 20% of GRS (54.2 (95% CI: (53.1, 55.3)) ml) had 32.8% larger volume of ventricular cerebrospinal fluid compared with those in the bottom 20% (40.8 (40.1, 41.6) ml). Individuals in the top 20% of GRS had larger volumes of left accumbens (13.0%), right accumbens (16.6%), brain stem+4th ventricle (9.8%), left caudate (10.8%), right caudate (10.6%), left putamen (10.3%), and right putamen (9.0%) compared to those in the bottom 20% (Fig. 3).

The adjustment for age and sex did not substantially change the association between metabolomic state and brain volumes (Figure S2).

Attribution of metabolomic state and clinical predictors to brain volumes

The application of several sets of predictors (metabolomic state, age+sex, multimorbidity score, GRS) was examined in the analysis. The largest variance of brain volumes was explained by age+sex, followed by metabolomic state and multimorbidity score (Fig. 4). Metabolomic state added minimally additional prediction value to age+sex but large additional prediction value to multimorbidity and GRS (Figure S3).

Attribution of GRS and clinical predictors to brain volumes

GRS explained 7.4% (95% CI: 3.4–10.5%), 5.4% (1.0-9.2%), and 2.6% (0.2–5.6%) of the variance of brain volumes of brain stem+4th ventricle, ventricular cerebrospinal fluid, and peripheral cortical grey matter, respectively. The percentage of the variance of volumes of thalamus, caudate, putamen, pallidum, and accumbens explained by GRS ranged from 3.1 to 8.7% (Fig. 5). Although the prediction values by GRS were lower than that by metabolomic state, GRS added higher prediction values to age+sex (Figure S4).

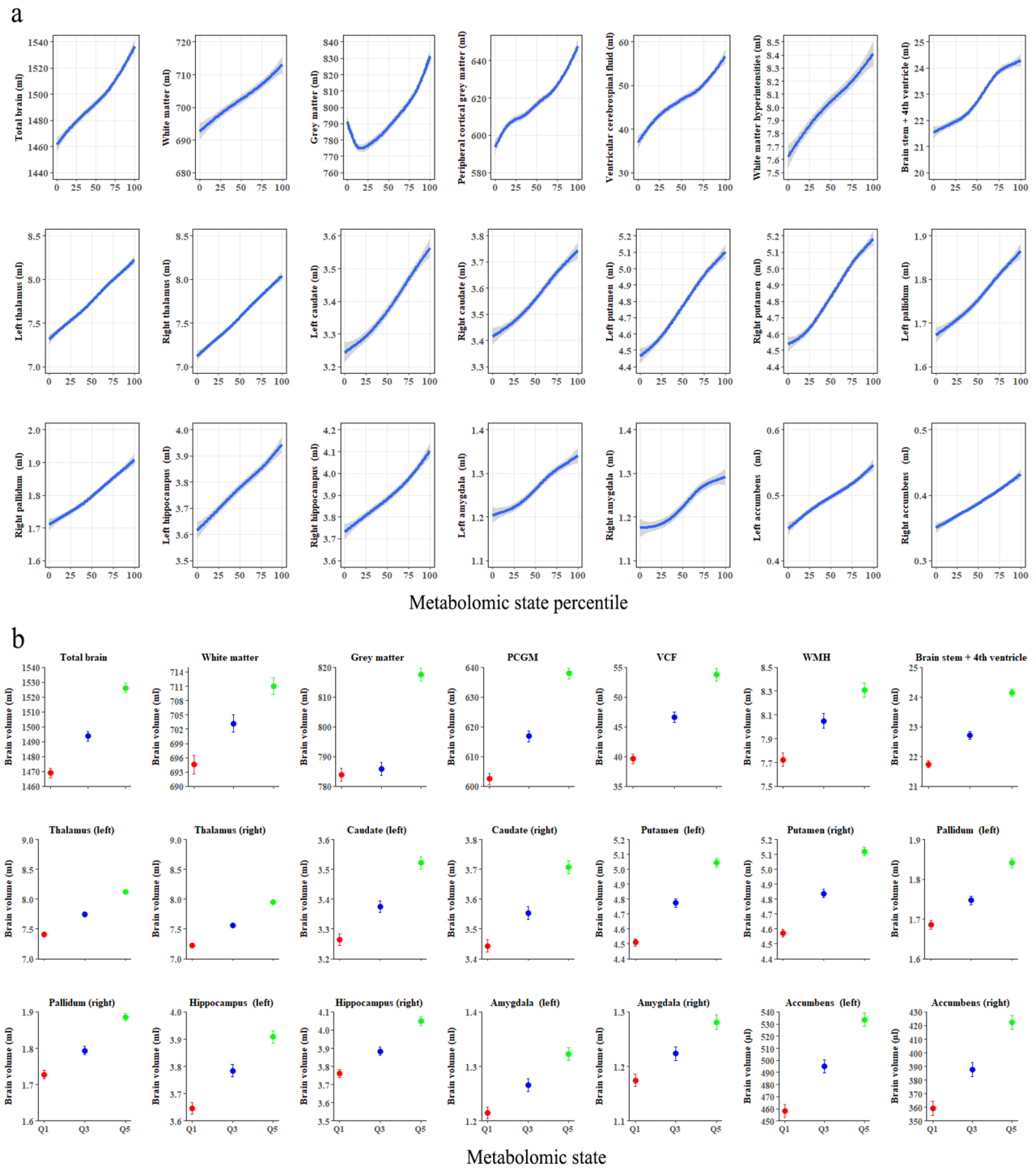

Fig. 2 Metabolomic state and brain phenotypes. **a**. Trends of 21 brain phenotypes by metabolomic state percentile. **b**. Brain volumes stratified by metabolomic state quantiles (red, bottom 20%; blue, median 20%; green, top 20%), with 95% CIs

Individual metabolomic profiles and brain volumes

After controlling for FDR, 139 individual metabolomic profiles were significantly associated with brain volumes (Fig. 6). When age and sex were adjusted for, this number was reduced to 51 (Figure S5). Lipids (medium VLDL, small LDL/HDL, large LDL/HDL, very large LDL/HDL), fatty acids (polyunsaturated fatty acids to total fatty acids ratio, saturated fatty acids to total fatty acids ratio), and amino acids (histidine) were important determinants for brain volumes.

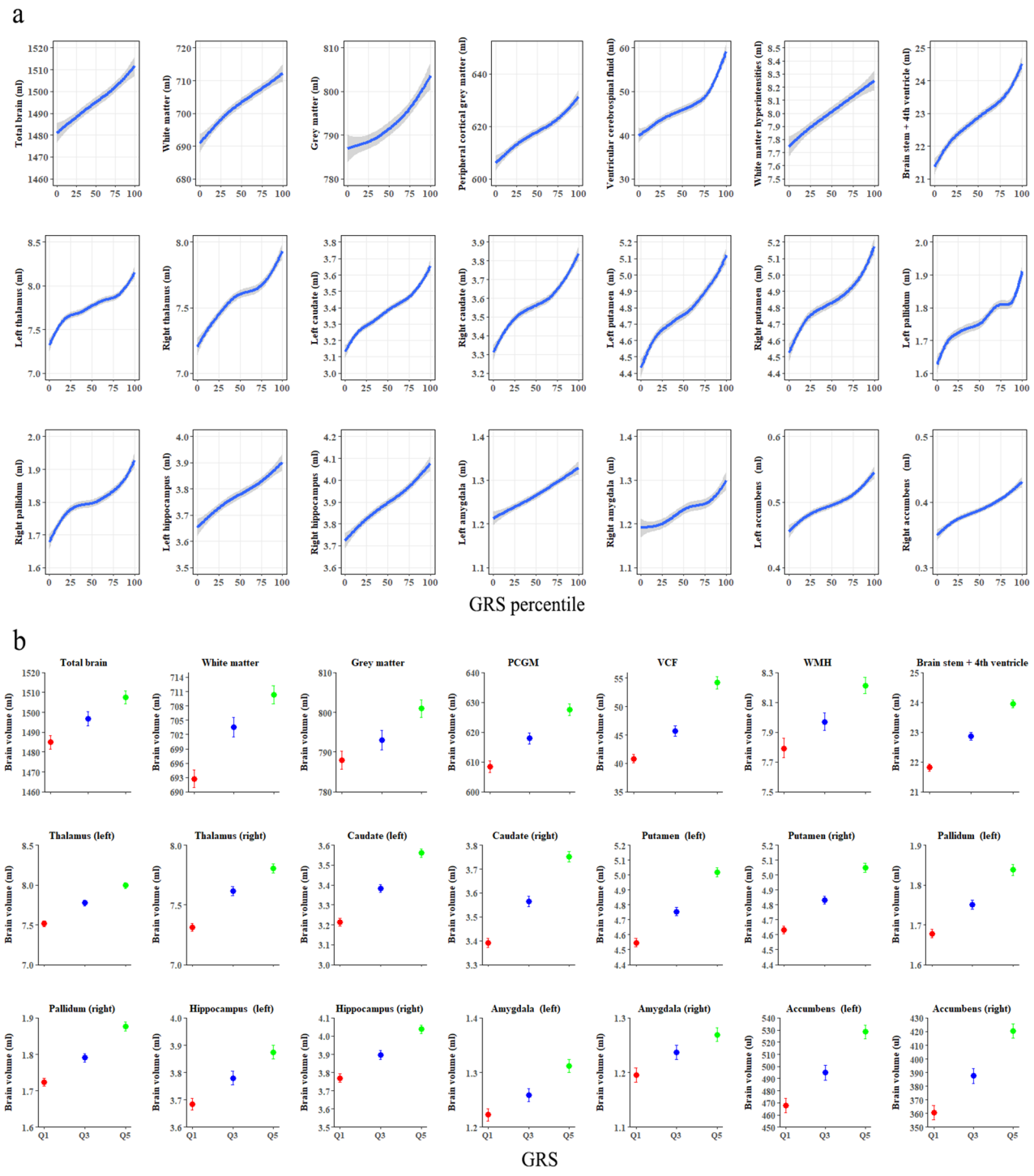

Fig. 3 Genetic risk score and brain phenotypes. **a**. Trend of 21 brain phenotypes by genetic risk score percentile. **b**. Brain volumes stratified by genetic risk score quantiles (red, bottom 20%; blue, median 20%; green, top 20%), with 95% CIs

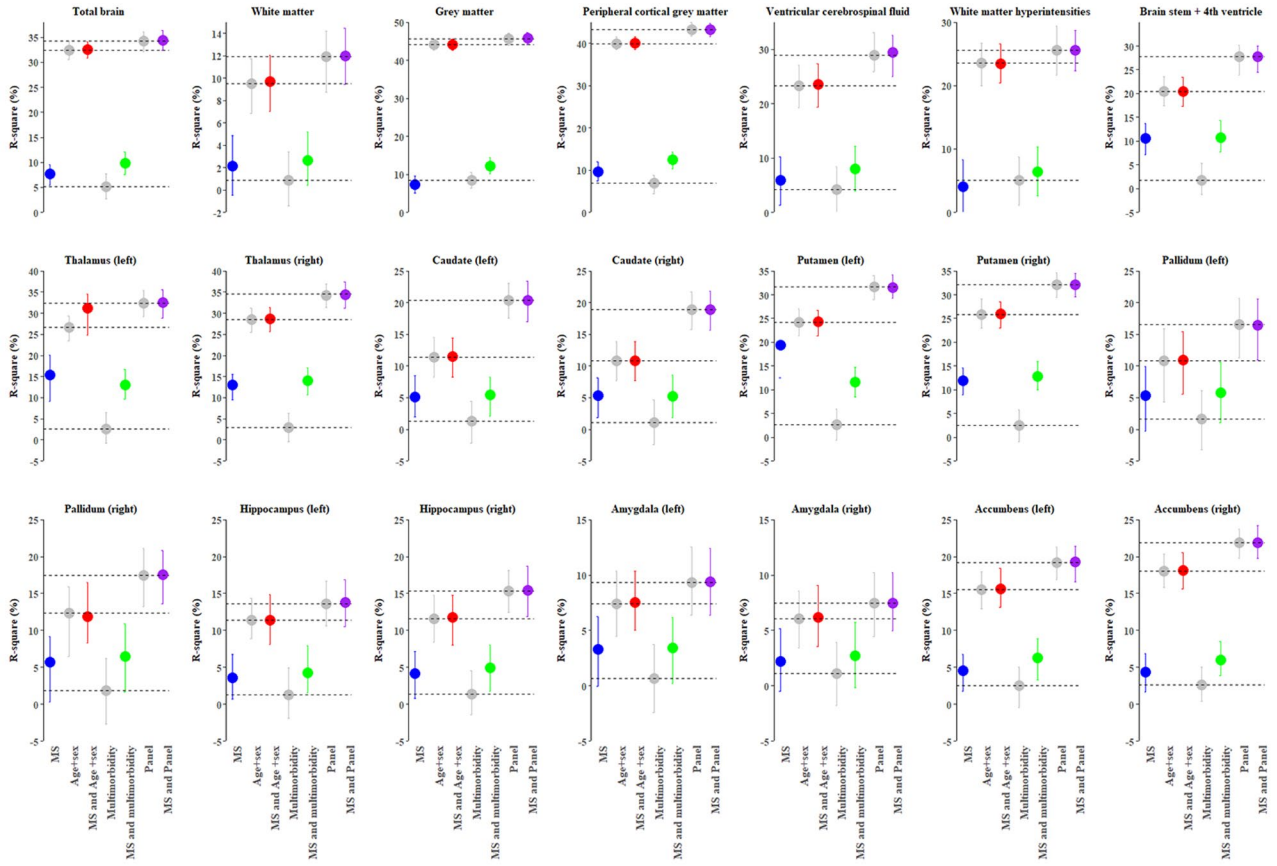

Fig. 4 The variance of brain phenotypes by metabolomic state only and its combination with age + sex, multimorbidity, and panel predictors. The percentage of the total variance in brain volumes (R^2) attributable to metabolomic state as well as age + sex, multimorbidity score, and Panel was estimated by bootstrapping with 500 iterations. MS refers to metabolomic state, and the panel refers to predictors including age, sex, education, and vascular risk factors. The dashed horizontal lines indicate the variance of brain volumes of age + sex, multimorbidity score, and Panel. The vertical lines indicate 95% CIs. For those R^2 with the lower 95% CIs < 0, the associations were reversed when the analysis was conducted among some bootstrapped samples compared with other samples. In the figure, colors were used to distinguish different items. Grey represents individual groups of predictors (age + sex, multimorbidity score, and Panel), while red, green, and purple indicate these groups plus the metabolomic state. (R2.3)

After controlling for FDR, 147 individual metabolomic profiles were significantly associated with brain volumes of thalamus, caudate, putamen, pallidum, hippocampus, amygdala, and accumbens (left/right sides) (Fig. 7). When age and sex were adjusted for, this number was reduced to 133 (Figure S6). Lipids (VLDL, LDL, and HDL) were important determinants for these brain volumes.

Moderation analysis

The association between metabolomic state and volumes of thalamus, pallidum, hippocampus, and accumbens at both sides was stronger in men than in women. The association between metabolomic state and volumes of thalamus and accumbens was stronger in younger than in older individuals. The association for total brain and WMH volumes did not differ between sexes or those younger and older participants (Figure S7). GRS was not a significant moderator for the association between metabolomic state and brain volumes.

The association between GRS and volumes of ventricular cerebrospinal fluid, thalamus, pallidum, hippocampus, and accumbens was stronger in men than in women. The association between GRS and volumes of white matter and ventricular cerebrospinal fluid was stronger among individuals aged ≥ 65 years than those aged < 65 years (Figure S8).

Discussion

This large cohort study of community-dwelling middle-aged and older adults demonstrated that both metabolomics and genetics played an important role in the variance of brain volumes. Metabolomic profiles were more predictive of brain volumes compared with traditional risk factors (multimorbidity and GRS), but these associations were substantially attenuated after adjustment for age and sex. These findings are consistent with previous studies demonstrating the importance of lipid metabolism on brain health, as most of the brain is

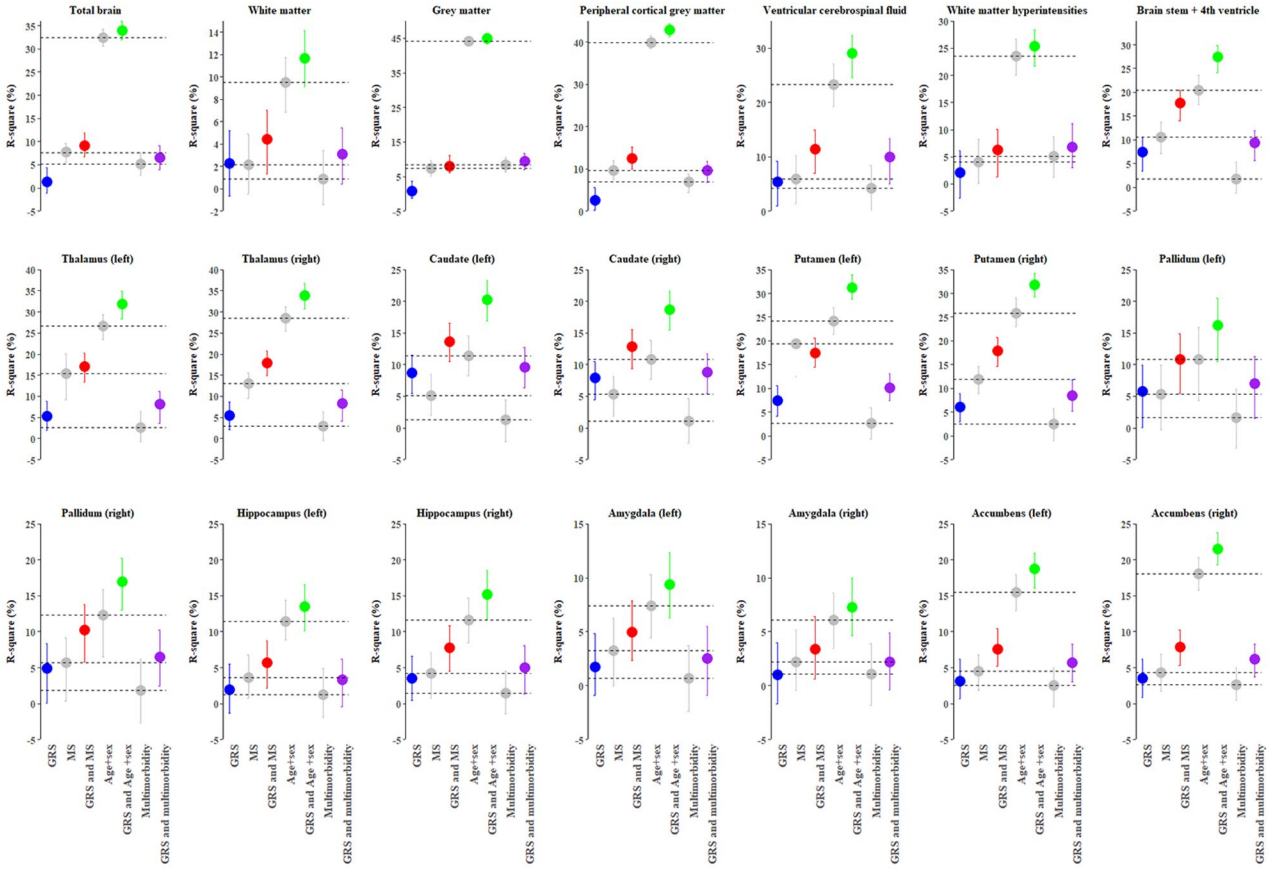

Fig. 5 The variance of brain phenotypes by genetic risk score only and its combination with age + sex, multimorbidity and panel predictors. The percentage of the total variance in brain volumes (R^2) attributable to genetic risk score as well as metabolomic state, age + sex, and multimorbidity score was estimated by bootstrapping with 500 iterations. The dashed horizontal lines indicate the variance of brain volumes of metabolomic state, age + sex, and multimorbidity score. The vertical lines indicate 95% CIs. For those R^2 with the lower 95% CIs < 0, the associations were reversed when the analysis was conducted among some bootstrapped samples compared with other samples

composed of lipids [27, 28]. Although GRS than metabolomic state was a smaller contributor to brain phenotypes, these associations were not significantly changed by age and sex. Metabolomic state provided minimal values and GRS provided moderate values to the prediction of brain volumes in addition to age and sex. No significant interplay between metabolomic state and GRS was observed, but the association between metabolomic state and some regional brain volumes was stronger in men or younger individuals. Individual metabolomic profiles including lipids and fatty acids were strong predictors of brain volumes. (R1.4)

Only several studies have investigated the association between multiple metabolomic profiles and brain volumes with inconsistent results. A pooled analysis of three independent cohorts (n=3962) showed that higher glucose levels and lower total cholesterol in small HDL, cholesterol esters in small HDL, and total lipids in small HDL levels were associated with smaller total brain volume whilst only glucose level was independently associated with WMH [11]. Another pooled analysis of 8 community-based cohorts demonstrated that lipids including lysophosphatidylcholines, hydroxysphingomyelins, low-density lipoprotein size, and composition and amino acids including hydroxyphenylpyruvate and glucuronate were significantly associated with WMH [12]. Consistently, our study demonstrated the importance of lipids and amino acids on brain atrophy. In addition, we found fatty acids were strongly associated with brain volumes. Several recent cohort studies have examined the association between patterns of metabolomic profiles and brain volumes. A study of 689 participants from the Alzheimer's Disease Neuroimaging Initiative cohort identified 9 principal components from 84 triglycerides and two consisting of long-chain, polyunsaturated fatty acid–containing triglycerides were significantly associated with hippocampal volume and entorhinal cortical thickness [29]. In the analysis of a subgroup of the UK Biobank cohort, some patterns of metabolic profiles (high triglycerides and liver enzymes or high BMI, CRP,

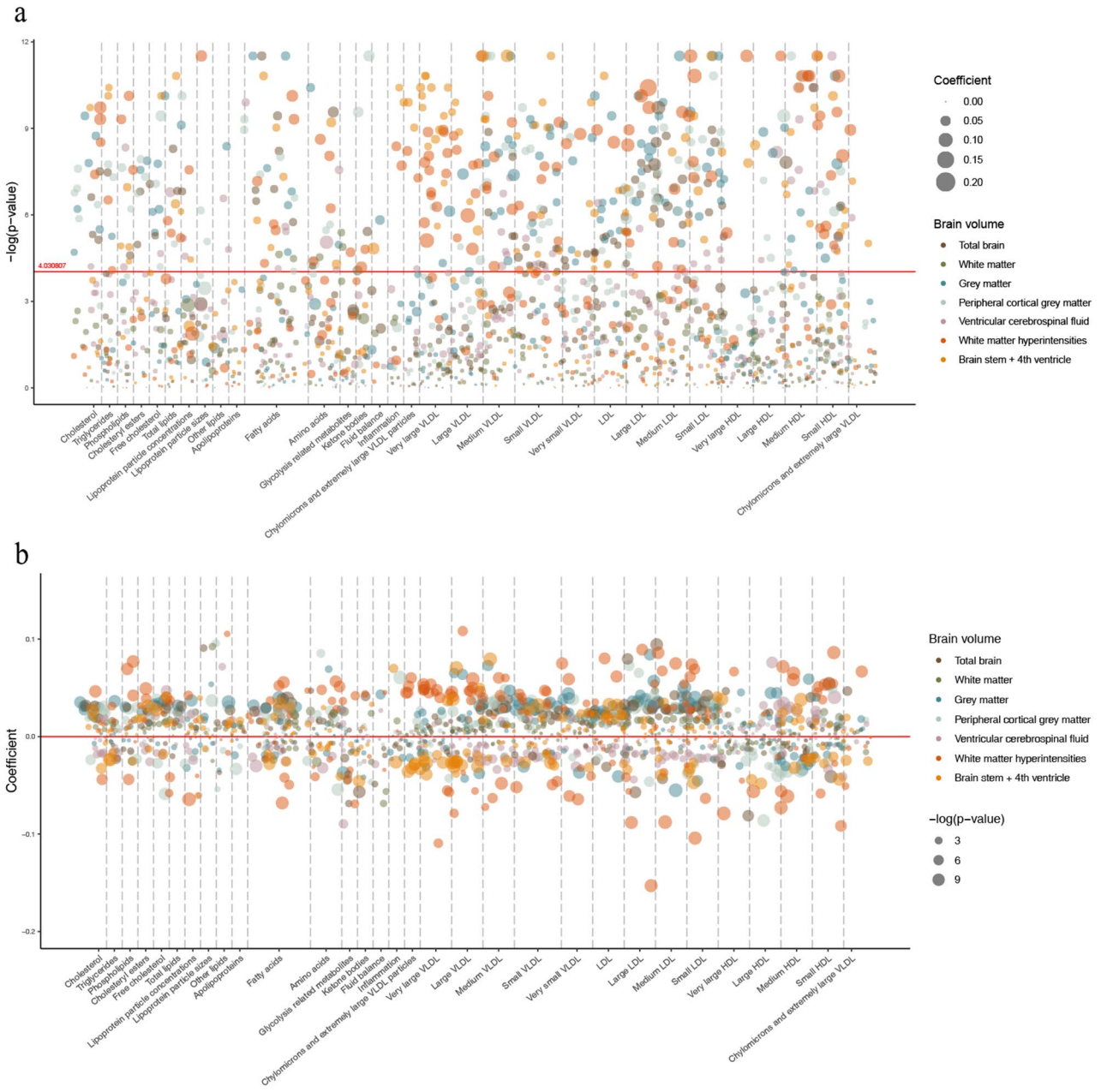

Fig. 6 Individual metabolomic profiles and total brain phenotypes. **a**. Manhattan plot shows the P values for correlations between total brain phenotypes and individual metabolomic profiles. The height of each point denotes the negative logarithm of the unadjusted correlation P value between one brain phenotype and one metabolomic profile. The area of the point denotes the absolute value of the Pearson's correlation coefficient. The color of the point denotes the brain phenotype. The false discovery rate for multiple comparisons ($\alpha = 0.05$) is shown as a red horizontal line. **b**. Manhattan plot shows the Pearson's correlation coefficients for correlations between total brain phenotypes and metabolomic profiles. The height of each point denotes the unadjusted correlation coefficient between one brain phenotype and one metabolomic profile. The area of the point denotes the negative logarithm of the age- and sex-adjusted correlation P value. The color of the point denotes the brain phenotype. The red horizontal line denotes the reference of coefficient as 0

and cystatin C) were associated with smaller grey matter and hippocampal volumes and higher WMH load [25]. The importance of metabolomic state developed based on 168 metabolomic profiles on the development of multiple common diseases has been highlighted in a previous study [13]. Our study showed that metabolomic state was a stronger predictor of brain volumes compared with conditional vascular risk score [6, 30, 31]. These previous studies focused on only several brain phenotypes including total brain, grey matter, WMH, and/or hippocampus.

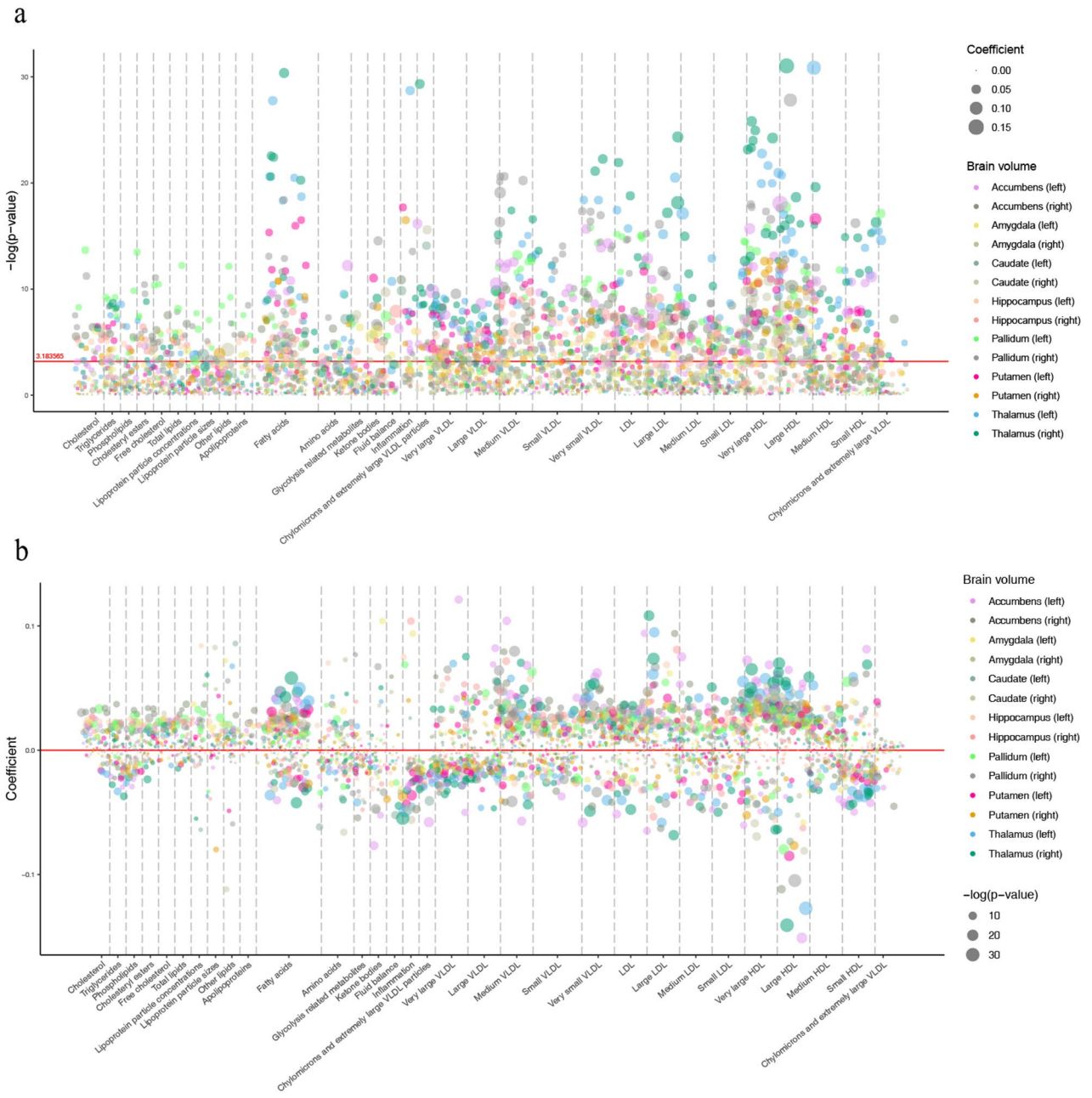

Fig. 7 Individual metabolomic profiles and regional brain phenotypes. **a**. Manhattan plot shows the P values for correlations between regional brain phenotypes and metabolomic profiles. The height of each point denotes the negative logarithm of the unadjusted correlation P value between one brain phenotype and one metabolomic profile. The area of the point denotes the absolute value of the Pearson's correlation coefficient. The color of the point denotes the brain phenotype. The false discovery rate for multiple comparisons ($\alpha = 0.05$) is shown as a red horizontal line. **b**. Manhattan plot shows the Pearson's correlation coefficients for correlations between regional brain phenotypes and metabolomic profiles. The height of each point denotes the unadjusted correlation coefficient between one brain phenotype and one metabolomic profile. The area of the point denotes the negative logarithm of the age- and sex-adjusted correlation P value. The color of the point denotes the brain phenotype. The red horizontal line denotes the reference of coefficient as 0

Our findings on multiple brain phenotypes including peripheral cortical grey matter, ventricular cerebrospinal fluid, brain stem+4th ventricle, thalamus, caudate, putamen, pallidum, amygdala, and accumbens need to be confirmed by further research.

Although the associations between metabolomic state and most brain volumes were independent of age and sex, these associations were largely attenuated after adjusting for age and sex in our study. Stratified analysis showed that the association between metabolomic

state and thalamus and accumbens volumes was more pronounced among younger than older individuals. We also found the association between metabolomic state and volumes of thalamus, pallidum, hippocampus, and accumbens was stronger in men than in women. This is consistent with a previous study showing that associations between levels of several individual metabolomic profiles including hydroxyphenylpyruvate, lysophosphatidylcholines, hydroxysphingomyelins, and diameter of low-density lipoprotein particles and WMH were stronger in men than in women [12]. Diet quality as a whole diet state seems to be more predictive of brain volumes and cognitive decline [32, 33]. This suggested the importance of examining the moderation of GRS, age and sex on the association between metabolomic state and brain volumes.

Although an increasing number of studies have conducted GWAS analysis on brain volumes [34–36], little is known regarding the contribution of GRS to multiple brain volumes. A recent cohort study has demonstrated that machine learning-based GRS could estimate an individual's predicted risk at desired error level [15]. We found GRS developed using machine learning (penalized models based on ridge regression and least absolute shrinkage and selection operator) was strongly predictive of brain volumes. In our study, the percentage of variation of caudate explained by GRS was around 8%, and the number for ventricular cerebrospinal fluid was greater than 5%. Although the prediction value of GRS was smaller than that of metabolomic state, GRS was more predictive of brain volumes than metabolomic state independent of age and sex. In further analysis, GRS was more predictive of ventricular cerebrospinal fluid, thalamus, caudate, putamen, pallidum, hippocampus, and accumbens in men than in women suggesting GRS for these brain MRI phenotypes may need to be created separately for men and women. GRS was a stronger predictor for volumes of white matter, ventricular cerebrospinal fluid, and pallidum in older than younger individuals. As an accelerated decrease in brain volumes is observed in adults beyond age 60 years [4, 5], larger variation of brain volumes in older adults may result in a stronger association between GRS and brain volumes.

To our knowledge, this is the first large cohort study to investigate the association between metabolomic state/GRS and multiple brain volumes. Several potential limitations need to be considered in our study. Firstly, because of the observational design of our study, causal relationships cannot be inferred based on the findings. Secondly, data on both MRI and metabolomic profiles were collected from only a small proportion of the UK cohort, from which the findings might not be applied to the whole UK population. Finally, the analysis of metabolomic and genetic profiles was conducted among individuals of European ancestry, which may reduce the generalizability of our findings to other ethnic groups.

In conclusion, metabolomic state is strongly associated with multiple brain volumes but provides minimal additional prediction value of brain volumes to age + sex. Although GRS is a weaker contributor to brain phenotypes than metabolomic state, it provides significant additional prediction value of brain volumes to age + sex. The predominant rationale behind this phenomenon lies in the lipid-rich composition of the brain. Metabolomic state is more predictive of some brain phenotypes in men than in women and in younger than in older individuals. Our findings suggest metabolomic state and GRS are important predictors for multiple brain phenotypes (R1.5).

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-024-05868-3.

Supplementary Material 1

Acknowledgements

This research was conducted using the UK Biobank resource. We thank the participants of the UK Biobank.

Author contributions

XS, MH conceived and designed the study. ZZ, WW performed data curation. XS conducted data analysis and drafted the initial manuscript. XS, XLZ, YH, SL, SM, LK, ZZ, WW, XYZ, JL, ST, YH, ZG, HY, and MH made a critical revision to the manuscript for important intellectual content. All authors read the manuscript and approved the final draft.

Funding

XLZ receives the National Natural Science Foundation of China under grant number 32200545, the GDPH Supporting Fund for Talent Program under grant numbers KJ012020633 and KJ012019530 from Guangdong Provincial People's Hospital, the Guangdong Provincial Medical Research Fund (A2024494), and the Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application under grant number 2022B1212010011. XYZ receives the National Natural Science Foundation of China (82171075, 82301260, 82271125), China Postdoctoral Science Foundation (2024T170185), the Science and Technology Program of Guangzhou, China (20220610092), the launch fund of Guangdong Provincial People's Hospital for NSFC (8217040546, 8220040257, 8217040449, 8227040339), Personalized Medical Incubator Project, The fund for Precison Medicine Research and Industry Development in SIMQ (2023-31), Guangdong Basic and Applied Basic Research Foundation (2023B1515120028). ZZ receives support from the National Natural Science Foundation of China (82101173), the Research Foundation of Medical Science and Technology of Guangdong Province (B2021237). HY receives support from the National Natural Science Foundation of China (81870663, 82171075), the Outstanding Young Talent Trainee Program of Guangdong Provincial People's Hospital (KJ012019087), Guangdong Provincial People's Hospital Scientific Research Funds for Leading Medical Talents and Distinguished Young Scholars in Guangdong Province (KJ012019457), Talent Introduction Fund of Guangdong Provincial People's Hospital (Y012018145). MH receives support from the High-level Talent Flexible Introduction Fund of Guangdong Provincial People's Hospital (No. KJ012019530), the Global STEM Professorship Scheme (P0046113), Research Matching Grant Scheme (P0048181), and PolyU - Rohto Centre of Research Excellence for Eye Care (P0046333). The sponsor or funding organization had no role in the design or conduct of this research. The sponsor or funding organization had no role in the design, conduct, analysis, or reporting of this study. The funding sources did not participate in the design and conduct of the study; collection, management, analysis, and interpretation of the data;

preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Data availability
Data are available in a public, open access repository (https://www.ukbiobank.ac.uk/).

Declarations

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Ophthalmology, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou 510080, China
[2]Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou 510080, China
[3]Centre for Eye Research Australia, Melbourne, VIC 3002, Australia
[4]Department of Medicine (Royal Melbourne Hospital), University of Melbourne, Melbourne, VIC 3050, Australia
[5]Medical Research Institute, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, China
[6]Medical Big Data Center, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences, Southern Medical University), Guangzhou, China
[7]State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou 510060, China
[8]Monash e-Research Center, Faculty of Engineering, Airdoc Research, Nvidia AI Technology Research Center, Monash University, Melbourne, VIC 3800, Australia
[9]School of Optometry, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China
[10]Research Centre for SHARP Vision (RCSV), The Hong Kong Polytechnic University, Kowloon, Hong Kong, China
[11]Centre for Eye and Vision Research (CEVR), 17W Hong Kong, Science Park, Hong Kong, China
[12]Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application, Guangzhou, China

Received: 8 July 2024 / Accepted: 9 November 2024
Published online: 03 December 2024

References

1. Thompson PM, et al. ENIGMA and global neuroscience: a decade of large-scale studies of the brain in health and disease across more than 40 countries. Translational Psychiatry. 2020;10:100.
2. Genon S, Eickhoff SB, Kharabian S. Linking interindividual variability in brain structure to behaviour. Nat Rev Neurosci. 2022;23:307–18.
3. Lombardi G et al., *Structural magnetic resonance imaging for the early diagnosis of dementia due to Alzheimer's disease in people with mild cognitive impairment. The Cochrane database of systematic reviews 3, Cd009628 (2020).*
4. Hou Y, et al. Ageing as a risk factor for neurodegenerative disease. Nat Rev Neurol. 2019;15:565–81.
5. Shang X, et al. Association of a wide range of individual chronic diseases and their multimorbidity with brain volumes in the UK Biobank: a cross-sectional study. EClinicalMedicine. 2022;47:101413.
6. Cox SR, et al. Associations between vascular risk factors and brain MRI indices in UK Biobank. Eur Heart J. 2019;40:2290–300.
7. Dekkers IA, Jansen PR, Lamb HJ, Obesity. Brain volume, and White Matter Microstructure at MRI: a cross-sectional UK Biobank Study. Radiology. 2019;291:763–71.
8. Song R, et al. Associations between Cardiovascular Risk, Structural Brain Changes, and Cognitive decline. J Am Coll Cardiol. 2020;75:2525–34.
9. Shang X, et al. The Association of Age at diagnosis of Hypertension with Brain structure and Incident Dementia in the UK Biobank. Hypertension. 2021;78:1463–74.
10. Shang X et al. Association of type 1 diabetes and age at diagnosis of type 2 diabetes with brain volume and risk of dementia in the UK Biobank: a prospective cohort study of community-dwelling participants. Diabet Med. 2023;40(2):e14966.
11. de Leeuw FA, et al. Circulating metabolites are associated with brain atrophy and white matter hyperintensities. Alzheimer's Dement J Alzheimer's Assoc. 2021;17:205–14.
12. Sliz E, et al. Circulating Metabolome and White Matter hyperintensities in Women and men. Circulation. 2022;145:1040–52.
13. Buergel T, et al. Metabolomic profiles predict individual multidisease outcomes. Nat Med. 2022;28:2309–20.
14. Smith SM, et al. An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. Nat Neurosci. 2021;24:737–45.
15. Sun J, et al. Translating polygenic risk scores for clinical use by estimating the confidence bounds of risk prediction. Nat Commun. 2021;12:5276.
16. Sudlow C, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12:e1001779.
17. Smith SM, et al. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. NeuroImage. 2002;17:479–89.
18. Alfaro-Almagro F, et al. Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. NeuroImage. 2018;166:400–24.
19. Würtz P, et al. Quantitative serum nuclear magnetic resonance metabolomics in large-scale epidemiology: a primer on -Omic technologies. Am J Epidemiol. 2017;186:1084–96.
20. Soininen P, Kangas AJ, Würtz P, Suna T, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. Circ Cardiovasc Genet. 2015;8:192–206.
21. Maclaren J, Han Z, Vos SB, Fischbein N, Bammer R. Reliability of brain volume measurements: a test-retest dataset. Sci Data. 2014;1:140037.
22. Boublay N, et al. Brain volume predicts behavioral and psychological symptoms in Alzheimer's Disease. J Alzheimers Dis. 2020;73:1343–53.
23. Lourida I, et al. Association of Lifestyle and genetic risk with incidence of Dementia. JAMA. 2019;322:430–7.
24. Shang X, et al. Association of a wide range of chronic diseases and apolipoprotein E4 genotype with subsequent risk of dementia in community-dwelling adults: a retrospective cohort study. EClinicalMedicine. 2022;45:101335.
25. Lumsden AL, Mulugeta A, Mäkinen VP, Hyppönen E. Metabolic profile-based subgroups can identify differences in brain volumes and brain iron deposition. Diabetes Obes Metab. 2023;25(1):121–131.
26. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser. 1995;57:289–300.
27. Kawade N, Yamanaka K. Novel insights into brain lipid metabolism in Alzheimer's disease: oligodendrocytes and white matter abnormalities. FEBS Open Bio. 2024;14:194–216.
28. Yoon JH, et al. Brain lipidomics: from functional landscape to clinical significance. Sci Adv. 2022;8:eadc9317.
29. Bernath MM et al. Serum triglycerides in Alzheimer disease: relation to neuroimaging and CSF biomarkers. Neurology 94, e2088-e2098 (2020).
30. Jeerakathil T, et al. Stroke risk profile predicts white matter hyperintensity volume: the Framingham Study. Stroke. 2004;35:1857–61.
31. Scharf EL, et al. Cardiometabolic Health and Longitudinal Progression of White Matter Hyperintensity: the Mayo Clinic Study of Aging. Stroke. 2019;50:3037–44.
32. Valls-Pedret C, et al. Mediterranean Diet and Age-Related Cognitive decline: a Randomized Clinical Trial. JAMA Intern Med. 2015;175:1094–103.
33. Ballarini T, et al. Mediterranean Diet, Alzheimer Disease biomarkers and brain atrophy in Old Age. Neurology. 2021;96:e2920–2932.
34. Zhao B et al. Common genetic variation influencing human white matter microstructure. Science 372(2021).
35. Elliott LT, et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. Nature. 2018;562:210–6.
36. Vuoksimaa E, et al. Heritability of white matter microstructure in late middle age: a twin study of tract-based fractional anisotropy and absolute diffusivity indices. Hum Brain Mapp. 2017;38:2026–36.

Publisher's note