Title: Increased brain volume from cereal, decreased brain volume from coffee -- shared genetic determinants and impacts on cognitive function, body mass index (BMI) and other metabolic measures: cohort study of UK Biobank participants

Authors: Jujiao Kang MSc [1,2,3†], Tianye Jia PhD [2,3,4†‡], Zeyu Jiao MSc [1,2,3], Chun Shen MSc [2,3], Chao Xie MSc [2,3], Wei Cheng PhD [2,3], Barbara J Sahakian DSc [2,3,5,6†‡], David Waxman PhD [2,3], Jianfeng Feng PhD [1,2,3,7,8†‡*]

Affiliations:

[1] Shanghai Center for Mathematical Sciences, Fudan University, Shanghai, China.

[2] Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China.

[3] Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence (Fudan University), Ministry of Education, China.

[4] Centre for Population Neuroscience and Precision Medicine (PONS), Institute of Psychiatry, Psychology & Neuroscience, SGDP Centre, King's College London, United Kingdom, SE5 8AF.

[5] Department of the Behavioural and Clinical Neuroscience Institute, University of Cambridge, Cambridge, United Kingdom.

[6] Department of Psychiatry, University of Cambridge School of Clinical Medicine, Cambridge, United Kingdom.

[7] Department of Computer Science, University of Warwick, Coventry, United Kingdom.

[8] School of Mathematical Sciences and Centre for Computational Systems Biology, Fudan University, Shanghai, China.

*Correspondence authors.

Jianfeng Feng

Institute of Science and Technology for Brain-inspired Intelligence, Fudan University, Shanghai, 200433, China.

Email: jianfeng64@gmail.com

† These authors contributed equally to this work.

‡ These authors contributed equally to this work.

Abstract

Objective: To explore how different diets may affect human brain development and if genetic and environmental factors play a part.

Design: Cohort study.

Setting: UK Biobank data were collected from 22 centres across the UK.

Participants: Only white British individuals free of Alzheimer's or dementia diseases were included in the study, where 336517 participants had quality-controlled genetic data, and 18879 participants had qualified brain MRI data.

Main outcome measures: Grey matter volume, intake of cereal and coffee, body mass index and blood cholesterol level.

Results: We investigated diet effects in the UK Biobank data and discovered anti-correlated brain-wide grey matter volume (GMV)-association patterns between coffee and cereal intake, coincidence with their anti-correlated genetic constructs. These genetic factors may further affect people's lifestyle habits and body/blood fat levels through the mediation of cereal/coffee intake, and the brain-wide expression pattern of gene CPLX3, a dedicated marker of subplate neurons that regulate cortical development and plasticity, may underlie the shared GMV-association patterns among the coffee/cereal intake and cognitive functions.

Conclusions: Our findings revealed that high-cereal and low-coffee diets shared similar brain and genetic constructs, leading to long-term beneficial associations regarding cognitive, BMI and other metabolic measures. This study has important implications for public health, especially during the pandemic, given the poorer outcomes of COVID-19 patients with greater BMIs.

Introduction

Increases in human brain volume, due to growth, begin at an early stage of embryonic development and continue until late adolescence[1]. After this, the brain experiences a persistent but slow decrease in size throughout adulthood[2]. Generally, development is tissue-specific but systematically organized across the brain[2 3] and may be susceptible to both genetic and environmental influences[2-6], as well as their interactions, e.g. through epigenetic modifications[7]. Diet is a common environmental factor that can influence the trajectory of brain size. For example, a lack of nutrients over an extended period of time causes both structural and functional damage to the brain[8], and improved diet quality is associated with larger brain volumes[9]. Furthermore, evidence suggested that ingested substances (both food and drink) in well-fed and healthy adults may also cause changes in brain size. For example, in a small-scale study, an increase in the size of the hippocampus was inferred to have occurred as an effect of both low and high coffee consumption[10].

While there are extensive studies of the degree to which different diets affect the body[11-14], there is an absence of systematic investigation into how different diets may affect the human brain in both the short and long term. Thus, it is not known if impacts of different diets on brain structures follow similar patterns, or whether different brain regions exhibit differential sensitivity to diet and other environmental factors. In addition, there is a lack of knowledge about whether genetic factors play any role in the sensitivity of the brain to environmental factors. In the present work, we provide a detailed analysis of brain-size changes that occur in healthy adults due to the ingestion of different common foods and drinks to investigate whether these influences from diets were

systematically organized across the brain, whether these diets influences have underlying genetic factors, and whether these genetic factors have further implications in people's daily activities, metabolism and cognitive functions.

Methods

Study participants

Study samples were from the UK Biobank study, a prospective epidemiological study that involves over 500,000 individuals in 22 centres across the UK[15]. Between 2006 and 2010, approximately 9.2 million mailed invitations to participate in the survey were sent to people in the National Health Service registry who were aged 40–69 years and living < 25 miles from a study centre. Participants were recruited to collect a range of questionnaires about detailed phenotypic information including diet, lifestyle, anthropometric and cognitive function assessments, biological samples, including blood and medical records obtained from the NHS registries. Since 2014, a subsample of the original population has been invited back to collect magnetic resonance imaging of body and brain, and questionnaires about diet, lifestyle, and cognitive function assessments.

In the current study, we used data collected at both recruitment and MRI scan. The original sample comprised 488289 individuals (56.54 ±8.09 years; 54.21% women). We included 431039 white British individuals and then excluded 810 individuals who were diagnosed with Alzheimer's or dementia defined by codes G30/F00 in the 10th edition of the International Classification of Diseases (ICD-10). Of the 430228 individuals, 336517 individuals had quality-controlled genetic data, and 18879 individuals had available brain MRI data. Table S1 summarized relevant demographic information. Behavioural and neuroimaging data collection and protocol are publicly available on [15 16]. All participants provided written informed consent to UK Biobank. Data access permission was granted under UKB application 19542 (PI Jianfeng Feng).

Assessment of the intake of cereal and coffee

Dietary information was obtained from the touchscreen questionnaire at the baseline and the MRI scan appointment. Cereal intake was the number of bowls of cereal the participants consumed per week. The types of cereal included bran cereal, biscuit cereal, oat cereal, muesli, and other types (e.g., cornflakes, Frosties). Coffee intake was the number of cups of coffee the participants drank per day. The types of coffee included decaffeinated coffee, instant coffee, ground coffee, other types of coffee. Detailed information can be found in supplementary materials.

Assessment of lifestyle

Lifestyle phenotypes included physical activity, sleep, smoking and alcohol and were obtained from the touchscreen questionnaire at the baseline appointment. Physical activities were assessed using MET (Metabolic Equivalent Task) scores derived based on International Physical Activity Questionnaire) of total physical activity (including walking, moderate, and vigorous activity) and usual walking pace. The time spent watching television was also included to reflect physical activity. Sleep data included information for sleep duration, morningness or eveningness type, insomnia symptoms, daytime dozing, getting up in morning, and nap during day. Smoking status included smoking history and the number of cigarettes currently smoked daily. Alcohol intake was examined using frequency and amounts of alcohol drinking. Detailed description can be found in supplementary materials.

Assessment of cognitive functions

Cognitive function performances were examined at the baseline and the MRI scan appointment. The cognitive tests included fluid intelligence score, reaction time, numeric memory, pairs matching, prospective memory, matrix pattern completion, symbol digit substitution and trail making. Detailed descriptions of procedures can be found in supplementary materials.

Assessment of body size and blood cholesterol

Body mass index was calculated from the participant's measured weight (kg)/height (m2). Cholesterol, high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, and triglycerides were measured in the blood sample collected at recruitment.

Assessment of the Alzheimer's disease risk

We used a proxy phenotype for Alzheimer's disease (AD) case-control status derived from the genetic risk index for AD based on parents' diagnoses as suggested in a previous study[17]. The proxy phenotype ranged approximately from 0 to 2, with values near zero when both parents were unaffected (lower for older parents and possible values below zero if both parents were over age 100) and values of two when both parents were affected.

COVID-19 test

COVID-19 test results data are linked to UK Biobank by Public Health England (PHE). Data were available for the period 16th March 2020 to 3rd August 2020. Data provided included specimen origin (hospital inpatient indicating severe COVID-19 vs. other settings). Detailed information is available on the website (http://biobank.ndph.ox.ac.uk/ukb/exinfo.cgi?src=COVID19_tests). To focus on the COVID-19, we excluded individuals passed away except for those who had positive test results. There were 13145 unique test results available, of which 1649 (12.54%) were positive; 10098 (76.82%) tests were conducted on inpatients; 1069 (639 had available data on BMI and diet) were inpatients and positive.

Structural MRI preprocessing

Detailed structural MRI data collection and acquisition procedures can be found in supplementary materials. All UK Biobank structural MRI data were preprocessed in the Statistical Parametric Mapping package[18] (SPM12) using the VBM8 toolbox with default settings, including the usage of high-dimensional spatial normalization with an already integrated Dartel template in Montreal Neurological Institute (MNI) space. All images were subjected to nonlinear modulations and corrected for each individual head size. Images were then smoothed with a 6 mm full-width at half-maximum Gaussian kernel with the resulting voxel size 1.5mm3. The estimated total intracranial volume (TIV) covariate, were calculated as the summation of the grey matter, white matter, and CSF volumes in native space. The automated anatomical labelling 3 (AAL3) atlas[19], which partitioned the brain into 166 regions of interest, was employed to obtain the total brain grey matter volume and region-wise grey matter volume. 18879 T1 images were successfully preprocessed, and the grey matter volumes of the AAL3 atlas of the discovery sample were extracted. The majority of participants were assessed in the Cheadle MRI site (84.49%) and the rest in the Newcastle site (15.51%).

Genetic data quality control

Detailed genotyping and quality control procedures of the UK Biobank can be found in supplementary materials or http://biobank.ctsu.ox.ac.uk/. In this study, we performed stringent QC standards by PLINK 1.90[20]. Single-nucleotide polymorphisms (SNPs) with call rates <95%, minor

allele frequency <0.1%, deviation from the Hardy–Weinberg equilibrium with p<1E-10 were excluded from the analysis. In addition, we selected subjects that were estimated to have recent British ancestry and have no more than ten putative third-degree relatives in the kinship table using the sample quality control information provided by UKB. For more details, we refer to the official document for genetic data of the UKB (http://www.ukbiobank.ac.uk/scientists-3/genetic-data/). After the quality control procedures, we obtained a total of 616,339 SNPs and 336517 participants.

Preprocessing of the Allen Human Brain Atlas data

We followed the AHBA preprocessing pipeline suggested by Arnatkevičiūtė et al.[21] and using the same pipeline as Shen et al.[22], including probe-to-gene re-annotation, data filtering, probe selection. In the next step, we separated the samples into the areas based on their MNI coordinates, using the automated anatomical labelling 3 (AAL3) atlas[19] and excluding the samples located outside of the grey matter defined by this atlas. To control for the inter-individual differences, we conducted two within-donor normalizations. The expression data were first normalized within-sample and across-gene and then normalized across samples. One gene failed the normalization and therefore was deleted, resulting in 15,408 genes. We used the mean expression of samples located in the brain region and the mean expression in the brain region of all subjects as the gene expression in each brain region defined by AAL3 atlas[19].

Linear regression analysis

We conducted linear regression analysis to test the pairwise associations between the diet phenotypes and the total and regional grey matter volume (GMV), respectively. The covariate variables included were age at imaging scan, sex, imaging sites (dummy variable), and total intracranial volume (TIV).

To further understand the biological insights of the shared variants of cereal intake and coffee intake, we performed linear regression analysis to examine the pairwise associations between the independent lead SNPs and other diets, lifestyle, and body/blood fat covarying age, sex, the top 40 genetic principal components. We also performed linear regression analysis to examine the pairwise associations between the cereal/coffee intakes, other diets, lifestyle, and cholesterol covarying age, sex.

Genome-wide association analysis and annotation of significant variants.

We performed genome-wide association analysis adjusting for age, sex, and the top 40 ancestry principal components using PLINK 1.90[20] to assess the association between phenotype and genotype on cereal intake and coffee intake separately. After association analysis, we employed the FUMA[23] online platform (version 1.3.6, http://fuma.ctglab.nl/) to define genomic risk loci. The GWAS summary statistics was submitted as input. FUMA identifies significant variants with P value less than 5E-8 that were largely independent of each other ($r^2 < 0.6$). Based on the clumping of the independent significant variants ($r^2 < 0.1$), independent lead variants were obtained.

Shared lead SNPs of cereal and coffee were mapped to genes based on cis-eQTL (p value≤ 0.05) in the brain using database GTEX[24] v8 with FUMA[23]. The eQTL mapping assigned SNPs to genes up to 1Mb apart.

Heritability and genetic correlation estimation

The LDSC software (https://github.com/bulik/ldsc) was employed to estimate the heritability of cereal intake and coffee intake as well as their genome-wide genetic correlation[25]. We used the

pre-calculated LD scores using 1000 Genomes European data. We used the overlap of summary statistics variants and HapMap variants as recommended[25].

Mediation analysis

Mediation effects were examined using Baron and Kenny's (1986)[26] causal steps approach. The causal steps approach involved four steps to establishing mediation. Firstly, a significant relation of the independent variable to the dependent variable is required in $Y = k_1 + \tau X + \varepsilon_1$ ($reject\ H_0: \tau = 0$). Secondly, a significant relation of the independent variable to the hypothesized mediating variable is required in $Z = k_2 + \alpha X + \varepsilon_2$ ($reject\ H_0: \alpha = 0$). Thirdly, the mediating variable must be significantly related to the dependent variable when both the independent variable and mediating variable are predictors of the dependent variable in $Y = k_3 + \tau' X + \beta Z + \varepsilon_3$ ($reject\ H_0: \beta = 0$) Fourthly, the coefficient relating the independent variable to the dependent variable must be larger (in absolute value) than the coefficient relating the independent variable to the dependent variable in the regression model with both the independent variable and the mediating variable predicting the dependent variable ($i.e.\ |\tau| > |\tau'|$). To further evaluate the p-value of the significant mediation identified by the above process, we performed 1000 times bootstrap of the individuals to obtain the distribution of the proportion of the mediation, i.e., $PM = (|\tau| - |\tau'|)/|\tau|$, under the alternative hypothesis. Thus, the PM was expected to be positive by definition, and the corresponding p-value could be calculated as the doubled chance of observing the PM less than zero during the 1000 bootstrap procedure. As no priory assumption about whether diet or lifestyle/ blood and body fat levels should serve as the mediator for their associations with the lead SNPs, we, therefore, identified the most likely mediator with an excess PM, i.e., the model showing higher PM, of which the significance level was again evaluated through a 1000-bootstrap process.

Pattern similarity analysis

We examined the similarity among the brain-wide GMV-association patterns of cereal/coffee intake and cognitive functions. Specifically, we first performed association analyses between region-wide GMV and each phenotype. Then, we calculated the Pearson correlation coefficient (similarity) between the GMV-association patterns of a pair of phenotypes of interest, of which the significance level was evaluated through 10000 times permutation that shuffled the individual's IDs of the GMV data at each iteration.

The similarity between brain-wide GMV-association pattern of a given phenotype and the brain-wide gene-expression pattern was also examined through their pattern correlation, of which the null distribution was established through 10000 times permutation that at each iteration, the pattern correlation was re-calculated with the GMV-association patterns been regenerated with shuffled IDs of the GMV data. The corresponding p-values were hence calculated as the chance of randomly getting a higher pattern correlation than the observed one in terms of their absolute value based on the established null distribution. The above permutation process was employed to ensure that the potential oversampling of brain regions will not inflate the false positive rate.

Patient and public involvement

We used anonymised data collected by UK Biobank study. No patients were involved in setting the research questions or the outcome measures.

Results

Association between grey matter volume and diets

We first investigated the relationship between grey matter volume (GMV) and 17 different diet phenotypes, which were both measured at the second visit (i.e. at follow-up) of participants to a research center[15]. We found that the *total grey matter volume of the brain* (TGMV) is affected by diet. Some dietary items were negatively correlated with the TGMV, thus decreasing consumption of these items had the tendency to increase the TGMV, while other items were positively correlated, and had the opposite tendency on the TGMV. With a statistically significant correlation (P<0.05 Bonferroni corrected), intake of coffee, water, processed meat, beef, lamb/mutton and pork were found to be negatively correlated with TGMV, while intake of cereal and dried fruit were positively correlated with TGMV (see Table S2). We note that predated measurements (i.e., baseline measurements) of cereal and coffee intake were also related to the follow-up values of TGMV (Table S3), and these remained significant even after controlling for the corresponding follow-up intakes (Table S4). This indicates a persistent, rather than a short-term connection between diet and GMV. Table S5 and S6, based on an alternative way of measuring TGMV, obtained similar results, confirming the methodological stability of the above findings.

Correlations were further investigated between 17 diet phenotypes and the volumes of 166 brain regions defined by the automated anatomical labelling 3 (AAL3) atlas[19]. A total of 454 statistically significant correlations (Bonferroni correction: P<0.05/166/17) were found, again mainly between GMV and intake of cereal, coffee, water, dried fruit, processed meat, beef, pork and lamb/mutton (Fig.1.A and Table S2). It is interesting to note that the GMV-association pattern of cereal intake highly resembles, although in the opposite direction, the GMV-association pattern of coffee intake (pattern correlation across the whole brain: r=-0.6177, $P_{perm<}$1E-04 based on 10000-permutation; Fig.1.B).

Genome-wide association studies for the intake of cereal and coffee

We conducted genome-wide association studies (GWAS) for the intake of both cereal (n=335696) and coffee (n=335068) at baseline and identified 21 and 45 independent lead genome-wide significant variants with P<5E-08 (i.e., the lead SNPs) respectively (Fig.2, Table S7&S8 and fig.S1). A linkage disequilibrium (LD) score regression[27] analysis indicates that both findings were free from systematically inflated false-positive rates, e.g., due to population stratification, with intercepts of 1.013 (cereal intake) and 1.005 (coffee intake), and the corresponding SNP-based heritabilities were estimated as 0.0652 (se=0.0038) and 0.0618 (se=0.007) respectively. Furthermore, we observed a significant negative genetic correlation between intake of cereal and coffee (r_g=-0.233, se=0.052, z-score=-4.49, P=7.1E-06), i.e., the alleles associated with higher cereal intake were likely to be in association with reduced coffee intake, which is in line with the above GWAS findings, where the three shared lead SNPs, i.e. rs2504706, rs4410790 and rs2472297, were found in associations with both cereal and coffee intake, again in opposite directions (Table S9).

The minor C-allele at rs2504706 was associated with a higher intake of cereal (regression coefficient 0.058, 95% confidence interval 0.042 to 0.073, $T_{df=334441}$=7.30, P=2.94E-13) and a lower intake of coffee (regression coefficient -0.034, 95% confidence interval -0.045 to -0.023, $T_{df=333816}$=-6.03, P=1.67E-09). The minor C-allele at rs4410790 was associated with lower intake of cereal (regression coefficient -0.038, 95% confidence interval -0.052 to -0.024, $T_{df=334331}$=-5.47, P=4.40E-08) and higher intake of coffee (regression coefficient 0.120, 95% confidence interval 0.110 to 0.130, $T_{df=333705}$=24.37, P=4.71E-131). The minor T-allele, at rs2472297, was associated

with lower intake of cereal (regression coefficient -0.059, 95% confidence interval -0.074 to -0.044, $T_{df=334951}$=-7.84, P=4.38E-15) and higher intake of coffee (regression coefficient 0.142, 95% confidence interval 0.131 to 0.152, $T_{df=334321}$=26.55, P=4.20E-155). While rs4410790 and rs2472297 have both been previously associated with coffee/caffeine consumption[28-30], caffeine metabolism[31], and alcohol consumption[32], this is the first study to identify an association with cereal intake. It is notable that SNPs rs4410790 (the C-allele) and rs2472297 (the T-allele) were also strongly associated with higher intake of tea (regression coefficient 0.111, 95% confidence interval 0.098 to 0.123, $T_{df=332509}$=17.04, P=4.58E-65 for rs4410790; regression coefficient 0.148, 95% confidence interval 0.134 to 0.162, $T_{df=333124}$=21.03, P=3.82E-98 for rs2472297, respectively) and lower intake of water (regression coefficient -0.075, 95% confidence interval -0.085 to -0.065, $T_{df=331879}$=-14.64, P=1.65E-48; regression coefficient -0.086, 95% confidence interval -0.097 to -0.076, $T_{df=332497}$=-15.62, P=5.53E-55, respectively) (Fig.3.A and Table S10), although both intakes were not observed with significant long term impacts on the TGMV (Table S4). This result is remarkable because there is a median to large anti-correlation between the intake of coffee and tea (r=-0.359, regression coefficient -0.472, 95% confidence interval -0.477 to -0.468, $T_{df=332711}$=-221.65, P<1.0E-256), which is likely due to the seesaw effect given the limited amount of beverages one may consume each day. Thus, individuals with both SNPs (i.e., C-allele of rs4410790 and T-allele of rs2472297) might generally prefer flavoured beverages to the water.

Association between genetic variants, diets and lifestyle

As both cereal and coffee intake, as well as their shared lead SNPs, were associated with different lifestyles, such as the frequency of physical activity (R=0.016, P=2.52E-17 for cereal and R=-0.011, P=3.23E-09 for coffee), being a morning/evening person (R=-0.040, P=2.57E-104 for cereal and R=0.032, P=3.05E-69 for coffee) and the frequency of alcohol use (R=-0.101, P<1.0E-256 for cereal and R=0.050, P=5.77E-184 for coffee) (Fig.3A and Table S11 & S12), we then investigated possible mediation roles of diet or/and lifestyles on their associations with SNPs. As no prior assumptions about whether diet or lifestyle should serve as the mediator for their associations with the lead SNPs, we evaluated the most likely mediator, based on the corresponding proportion of mediation (PM) that they are responsible for (see the Supplementary Material for more details). We found the following:

1) Both intake of cereal and coffee were likely to mediate the positive association of the frequency of alcohol intake with the T-allele of rs2472297 (PM=24.75%, $P_{bootstrap}$=5.47E-15 and PM=38.14%, $P_{bootstrap}$=1.37E-82 respectively; Table S13); these were superior to alternative mediation models with the frequency of alcohol intake as the mediator (excess PM>20% with $P_{bootstrap}$<0.002 for both alternative models, Fig.3.B and Table S13; see supplementary materials for detailed analyses);

2) The association between higher T-alleles of rs2472297 and less daytime sleeping was mediated by cereal intake (PM=1.98%, $P_{bootstrap}$=2.82E-6; Fig.3.B and Table S13), which was superior to the alternative mediation model with daytime sleeping as the mediator (excess PM=1.39% with $P_{bootstrap}$=0.018, Table S13; see supplementary materials for detailed analyses);

3) Both difficult in rising and less daytime sleeping were found to mediate the negative association of cereal intake with the C-allele of rs4410790, so did the alternative mediation models with the cereal intake as the mediator. However, neither group of mediation models was superior to the other (Table S13);

4) Interestingly, while individuals with rs2504706 (the C-allele) were more likely to be an 'evening person' and experience difficulties in rising, both lifestyle traits did not mediate the associations of

the SNP with higher cereal intake or lower coffee intake (nor did the alternative mediation models), which was mainly due to nonconcordant correlations, e.g., a positive correlation was observed between ease in rising and higher cereal intake while a negative one was expected (Fig.3.A and Table S9, S11 & S12).

Association between genetic variants, diets and metabolic measures

In addition to lifestyle, both cereal and coffee intake, as well as their shared lead SNPs, were also associated with blood (for example with total cholesterol, R=-0.066, P<1.0E-256 for cereal and R=0.045, P=1.89E-139 for coffee) and body fat levels (for example with the body mass index (BMI), R=-0.076, P<1.0E-256 for cereal and R=0.053, P=3.84E-206 for coffee) (Table S14, Table S15). Therefore, we further explored possible mediator roles of fat levels and the intake of cereal and coffee. We found the following:

1) Associations between rs4410790 (C-allele) and: an increased body mass index (BMI), triglycerides and decreased HDL cholesterol, were mediated by increased coffee intake (PM=35.31%, $P_{bootstrap}$=1.41E-81, PM=3.30%, $P_{bootstrap}$=1.04E-5, and PM=3.09%, $P_{bootstrap}$=2.28E-4, respectively), which were superior to the alternative mediation models with corresponding fat levels as mediators (excess PMs=34.48%, 3.10% and 2.95%, respectively; all corresponding $P_{bootstrap}$<0.002) (Table S16);

2) Associations between rs2472297 (T-allele) and higher body mass index (BMI), total cholesterol, and LDL cholesterol, were mediated by higher coffee intake (PM=27.71%, $P_{bootstrap}$=6.72E-83, PM=25.14%, $P_{bootstrap}$=6.82E-68 and PM=28.92%, $P_{bootstrap}$=4.65E-75, respectively), as well as by lower cereal intake, to a lesser extent (PM=11.46% for total cholesterol, $P_{bootstrap}$=6.51E-14 and PM=8.56% for LDL cholesterol, $P_{bootstrap}$=1.80E-13). The above models were superior to alternative mediation models with corresponding fat levels as mediators (for the coffee intake: excess PMs=26.66%, 24.38% and 28.13%, respectively, with all corresponding $P_{bootstrap}$<0.002; for the cereal intake: excess PMs=7.54% and 5.83%, respectively, with all corresponding $P_{bootstrap}$<0.05) (Table S16).

Related to the current COVID-19 pandemic, using the UK Biobank data we found that individuals who tested positive of COVID-19 (n=639, inpatients only) had higher BMIs (Cohen's D=0.27, t=6.72, P=1.86E-11) and lower cereal intake (Cohen's D=-0.09, t=-2.36, P=0.019) than the rest population (n=314982, either tested negative or not tested). This further highlights the importance of our finding for public health that cereal intake is associated with lower BMIs.

Associations between the GMV-association patterns of cognitive functions and the GMV-association patterns of the intake of cereal and coffee

To further characterize the negatively correlated brain-wide GMV-association patterns for cereal and coffee intakes, we further investigated if such similarities have any implications for cognitive functions, and we found that brain-wide GMV-association patterns of most cognitive functions were significantly correlated with those of both cereal and coffee intake, although in opposite directions, at both baseline and follow-up (GMV were measured at follow-up only). In particular, performance in tasks of matrix pattern completion, symbol digit substitution, and numeric and alphabet-numeric trail making showed similar brain-wide GMV-association patterns with both cereal (in positive correlation) and coffee (in negative correlation) intake at both baseline and follow-up ($|R|_{min}$=0.5945, all P_{FDR}<0.05; Fig.4 and Table S17&S18), while the fluid intelligence score only showed a similar brain-wide GMV association pattern with the cereal intake (at both

baseline and follow-up; R_{min}=0.62, all P_{FDR}<0.05; Table S17&S18). In line with the above findings, higher risk of Alzheimer's disease (estimated as the proxy-AD [17]), characterized by reduced cognitive functions, was associated with reduced cereal intake (R=-0.009, P=3.42E-6), as well as increased coffee intake to a much lesser extent (R=0.004, P=0.024), in contrast to previous findings of either protective[33] or non-significant[34] effect of high coffee intake on Alzheimer's disease.

Associations between the GMV-association patterns of the cereal/coffee intake and the gene-expression patterns of eQTL genes

We further investigated if the identified putative genetic variants may also contribute to observed similarities of brain-wide GMV association patterns between diet and cognitive function, through the expression of candidate genes. We first performed eQTL mapping of the 3 lead SNPs using software FUMA[23] and identified 31 candidate protein-coding genes (Table S19) that also have brain-wise gene expression information from the Allen Institute for Brain Science (AIBS)[35]. After also mapping to the AAL3 atlas, the brain-wide expression pattern for each candidate gene (i.e. the mean expression level across all AIBS individuals for each brain region) was then correlated with the brain-wide GMV association patterns for cereal and coffee intakes. While multiple candidate genes had their brain-wide expression pattern in significant correlation with brain-wide GMV associations patterns for the coffee intake (Table S19), only gene CPLX3 showed significant 'gene-expression vs GMV-association' pattern similarity with both intakes of cereal (R=0.47, P_{perm}=2.9E-3, $P_{FDR\text{-}corrected}$=0.033) and coffee (R=-0.44, P_{perm}=7.2E-3, $P_{FDR\text{-}corrected}$=0.046). It is of particular interest that the gene-expression of CPLX3 (a known prominent marker that is specific for subplate neurons that regulate cortical development and plasticity across the brain [36-38] and also respond to both light and electrical stimuli in retinal neurons [39 40]) also showed significant pattern correlations with almost all cognitive functions (i.e., R=0.42 for fluid intelligence, R=0.49 for numerical memory, R=0.44 for prospective memory, R=0.46 for matrix pattern completion, R=0.39 for symbol digit substitution, and R=0.44/R=0.55 for both trail making tasks; all corresponding $P_{FDR\text{-}corrected}$<0.05; Table S20).

Discussion

In the large-scale imaging/genetics analysis presented in this work, we have: (i) gained insights into long-term associations between brain-wide GMV and diets, especially the anti-correlated impacts from cereal and coffee intake, (ii) identified shared genetic constructs for both higher cereal and lower coffee intake, and explored the complex relationship among cereal/coffee intake, their genetics constructs, lifestyle, and body/blood fat level, (iii) revealed shared brain-wide GMV-association patterns between cognitive function and the intake of cereal and coffee and further showed that such similarity might be underlaid by the brain-wide expression patterns of gene CPLX3, a shared genetic determinant identified for the intake of cereal and coffee. These novel findings hence suggest the existence of a brain-wide systematic organization of GMV that is susceptible to both genetic and environmental influences, which may have further impacts on people's lifestyles, cognitive functions, and metabolic measures (e.g. BMI and blood cholesterol level).

Comparisons with other studies

Two lead SNPs shared by the intake of coffee and cereal, i.e. rs4410790 and rs2472297, have been previously associated with coffee/caffeine consumption[28-30], caffeine metabolism[31], and alcohol consumption[32]. However, this is the first study to identify their associations with cereal intake.

Moreover, while CPLX3, within the LD complexity around the lead SNP rs2472297 along with 90 other genes (based on R package "biomaRt"[41]), has previously been proposed as a candidate gene for both coffee consumption[42] and blood pressure[43], it is the first time that this gene has been linked with multiple cognitive functions, such as intelligence, as well as with cereal intake. Remarkably, the expression of CPLX3 is a highly specific marker of subplate neurons [36] that regulate cortical development and neuronal plasticity across the brain [37 38]. Specifically, while most subplate neurons were short-lived during the development of the brain, previous studies have shown that the Cplx3-positive subplate neurons could survive into adulthood in mice [36]. Therefore, our findings were not only congruent with the role of subplate neurons in the cortical development, but with further implications that these CPLX3-positive subplate neurons might mark the dynamic system of GMV in the brain that is susceptible to environmental factors. Such a hypothesis could be supported by previous findings that Cplx3 protein's regulation of exocytosis in mice retinal neurons could be altered by both light and electrical stimuli [39 40].

Limitations of the study

UK Biobank only includes participants aged 40 and above. Therefore, some of our results may not necessarily reflect the situation in the younger population, although the GWAS findings are highly consistent with previous literature across different age bands, which may help to alleviate such a concern. In addition, further studies are still needed to fully understand the molecular and metabolic pathways involved in the systematic alternation of GMVs in the brain, as well as its influences on cognitive function and metabolism.

Conclusion

Since high cereal diets, but low coffee diets, have long-term beneficial associations regarding the brain, cognition, BMI and other metabolic measures, this study has significant implications for public health. Our findings highlight the importance of a 'cereal' breakfast across the life span, but perhaps especially for children and adolescents whose brains are still in development and for reducing the risk of Alzheimer's disease and poor outcomes due to high BMIs in patients with COVID-19 [44 45].

References

1. Rapp PR, Bachevalier J. Cognitive development and aging. Fundamental neuroscience: Elsevier 2013:919-45.
2. Ziegler G, Dahnke R, Jäncke L, et al. Brain structural trajectories over the adult lifespan. *Human brain mapping* 2012;33(10):2377-89.
3. Tamnes CK, Østby Y. Morphometry and Development: Changes in Brain Structure from Birth to Adult Age. Brain Morphometry: Springer 2018:143-64.
4. Fjell AM, Grydeland H, Krogsrud SK, et al. Development and aging of cortical thickness correspond to genetic organization patterns. *Proceedings of the National Academy of Sciences* 2015;112(50):15462-67.
5. Satizabal CL, Adams HH, Hibar DP, et al. Genetic architecture of subcortical brain structures in 38,851 individuals. *Nature genetics* 2019;51(11):1624-36.
6. Zhao B, Ibrahim JG, Li Y, et al. Heritability of regional brain volumes in large-scale neuroimaging and genetic studies. *Cerebral Cortex* 2019;29(7):2904-14.
7. Jia T, Chu C, Liu Y, et al. Epigenome-wide meta-analysis of blood DNA methylation and its association with subcortical volumes: findings from the ENIGMA Epigenetics Working Group.

Molecular Psychiatry 2019:1-12.
8. Dewey KG, Begum K. Long‐term consequences of stunting in early life. *Maternal & child nutrition* 2011;7:5-18.
9. Croll PH, Voortman T, Ikram MA, et al. Better diet quality relates to larger brain tissue volumes: The Rotterdam Study. *Neurology* 2018;90(24):e2166-e73.
10. Perlaki G, Orsi G, Kovacs N, et al. Coffee consumption may influence hippocampal volume in young women. *Brain imaging and behavior* 2011;5(4):274-84.
11. Klerks M, Bernal MJ, Roman S, et al. Infant Cereals: Current Status, Challenges, and Future Opportunities for Whole Grains. *Nutrients* 2019;11(2):473. doi: 10.3390/nu11020473
12. Truswell A. Cereal grains and coronary heart disease. *European journal of clinical nutrition* 2002;56(1):1-14.
13. Poole R, Kennedy OJ, Roderick P, et al. Coffee consumption and health: umbrella review of meta-analyses of multiple health outcomes. *bmj* 2017;359
14. van Dam RM, Hu FB, Willett WC. Coffee, Caffeine, and Health. *New England Journal of Medicine* 2020;383(4):369-78.
15. Sudlow C, Gallacher J, Allen N, 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 medicine* 2015;12(3):e1001779.
16. Miller KL, Alfaro-Almagro F, Bangerter NK, et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature neuroscience* 2016;19(11):1523.
17. Jansen IE, Savage JE, Watanabe K, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature genetics* 2019;51(3):404-13.
18. Eickhoff SB, Stephan KE, Mohlberg H, et al. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 2005;25(4):1325-35.
19. Rolls ET, Huang C-C, Lin C-P, et al. Automated anatomical labelling atlas 3. *NeuroImage* 2019:116189.
20. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics* 2007;81(3):559-75.
21. Arnatkevičiūtė A, Fulcher BD, Fornito A. A practical guide to linking brain-wide gene expression and neuroimaging data. *Neuroimage* 2019;189:353-67.
22. Shen C, Luo Q, Chamberlain SR, et al. What is the Link between Attention-Deficit/Hyperactivity Disorder and Sleep Disturbance? A multimodal examination of longitudinal relationships and brain structure using large-scale population-based cohorts. *Biological Psychiatry* 2020
23. Watanabe K, Taskesen E, Van Bochoven A, et al. Functional mapping and annotation of genetic associations with FUMA. *Nature communications* 2017;8(1):1826.
24. Consortium G. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015;348(6235):648-60.
25. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nature genetics* 2015;47(11):1236.
26. Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology* 1986;51(6):1173.
27. Bulik-Sullivan BK, Loh P-R, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics* 2015;47(3):291.

28. Cornelis MC, Monda KL, Yu K, et al. Genome-wide meta-analysis identifies regions on 7p21 (AHR) and 15q24 (CYP1A2) as determinants of habitual caffeine consumption. *PLoS genetics* 2011;7(4):e1002033.
29. Cornelis MC, Byrne EM, Esko T, et al. Genome-wide meta-analysis identifies six novel loci associated with habitual coffee consumption. *Molecular psychiatry* 2015;20(5):647.
30. Sulem P, Gudbjartsson DF, Geller F, et al. Sequence variants at CYP1A1–CYP1A2 and AHR associate with coffee consumption. *Human molecular genetics* 2011;20(10):2071-77.
31. Cornelis MC, Kacprowski T, Menni C, et al. Genome-wide association study of caffeine metabolites provides new insights to caffeine metabolism and dietary caffeine-consumption behavior. *Human molecular genetics* 2016;25(24):5472-82.
32. Liu M, Jiang Y, Wedow R, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature genetics* 2019;51(2):237.
33. Poole R, Kennedy OJ, Roderick P, et al. Coffee consumption and health: umbrella review of meta-analyses of multiple health outcomes. *BMJ (Clinical research ed)* 2017;359:j5024. doi: 10.1136/bmj.j5024 [published Online First: 2017/11/24]
34. Larsson SC, Orsini N. Coffee Consumption and Risk of Dementia and Alzheimer's Disease: A Dose-Response Meta-Analysis of Prospective Studies. *Nutrients* 2018;10(10):1501. doi: 10.3390/nu10101501
35. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 2012;489(7416):391.
36. Viswanathan S, Sheikh A, Looger LL, et al. Molecularly Defined Subplate Neurons Project Both to Thalamocortical Recipient Layers and Thalamus. *Cerebral cortex (New York, NY : 1991)* 2017;27(10):4759-68. doi: 10.1093/cercor/bhw271
37. Kanold PO. Subplate neurons: crucial regulators of cortical development and plasticity. *Frontiers in neuroanatomy* 2009;3:16-16. doi: 10.3389/neuro.05.016.2009
38. Kanold PO, Shatz CJ. Subplate Neurons Regulate Maturation of Cortical Inhibition and Outcome of Ocular Dominance Plasticity. *Neuron* 2006;51(5):627-38. doi: https://doi.org/10.1016/j.neuron.2006.07.008
39. Mortensen LS, Park SJH, Ke J-B, et al. Complexin 3 Increases the Fidelity of Signaling in a Retinal Circuit by Regulating Exocytosis at Ribbon Synapses. *Cell Rep* 2016;15(10):2239-50. doi: 10.1016/j.celrep.2016.05.012 [published Online First: 05/26]
40. Babai N, Sendelbeck A, Regus-Leidig H, et al. Functional Roles of Complexin 3 and Complexin 4 at Mouse Photoreceptor Ribbon Synapses. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2016;36(25):6651-67. doi: 10.1523/JNEUROSCI.4335-15.2016
41. Durinck S, Spellman PT, Birney E, et al. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature protocols* 2009;4(8):1184.
42. Amin N, Byrne E, Johnson J, et al. Genome-wide association analysis of coffee drinking suggests association with CYP1A1/CYP1A2 and NRCAM. *Mol Psychiatr* 2012;17(11):1116-29. doi: 10.1038/mp.2011.101 [published Online First: 08/30]
43. Evangelou E, Warren HR, Mosen-Ansorena D, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nature Genetics* 2018;50(10):1412-25. doi: 10.1038/s41588-018-0205-x
44. Simonnet A, Chetboun M, Poissy J, et al. High Prevalence of Obesity in Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) Requiring Invasive Mechanical Ventilation. *Obesity (Silver Spring)* 2020;28(7):1195-99. doi: 10.1002/oby.22831 [published Online First:

06/10]
45. Dugail I, Amri E-Z, Vitale N. High prevalence for obesity in severe COVID-19: Possible links and perspectives towards patient stratification. *Biochimie* 2020:S0300-9084(20)30155-3. doi: 10.1016/j.biochi.2020.07.001

Acknowledgments
Funding: This work received support from the following sources: the National Key Research and Development Program of China (2018YFC1312900 and 2019YFA0709502), the National Natural Science Foundation of China (91630314 and 81801773), the 111 Project (B18015), The Key Project of Shanghai Science &Technology Innovation Plan(16JC1420402), the Shanghai Municipal Science and Technology Major Project (2018SHZDZX01) and Zhangjiang Lab. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author Contributions
Jujiao Kang and Tianye Jia contributed equally in writing the first draft of the manuscript and conducting data analysis.
Tianye Jia, Barbara J. Sahakian and Jianfeng Feng contributed equally in conception or design of the study and results interpretation.

Conception or Design of the Study: T.J., B.J.S. and J.F.
Manuscript Writing and Editing: J.K., T.J. and D.W. wrote the manuscript; B.J.S. and J.F. edited the first draft; all authors critically reviewed the manuscript
Imaging Data Preprocessing: J.K., Z.J. and W.C.
Visualisation: J.K., T.J. and C.X.
Data Analysis: J.K. conducted all the statistical analyses, under the instruction of T.J.
Results Interpretation: T.J., B.J.S. and J.F.
Supervision of the Study: T.J. and J.F.
Funding Acquisition: T.J. and J.F.

Competing Interests
All authors have completed the ICMJE uniform disclosure form and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work.

Data and Materials Availability
All UK Biobank data used in this work were obtained under Data Access Application 19542 and are available to eligible researchers through the UK Biobank (www.biobank.ac.uk). Gene expression data from the Allen Institute for Brain Science are freely available at https://human.brain-map.org/static/download. Custom code that supports the findings of this study is available from the corresponding author upon request.

Supplementary Materials

Supplementary Methods

Figures S1

Tables S1-S20

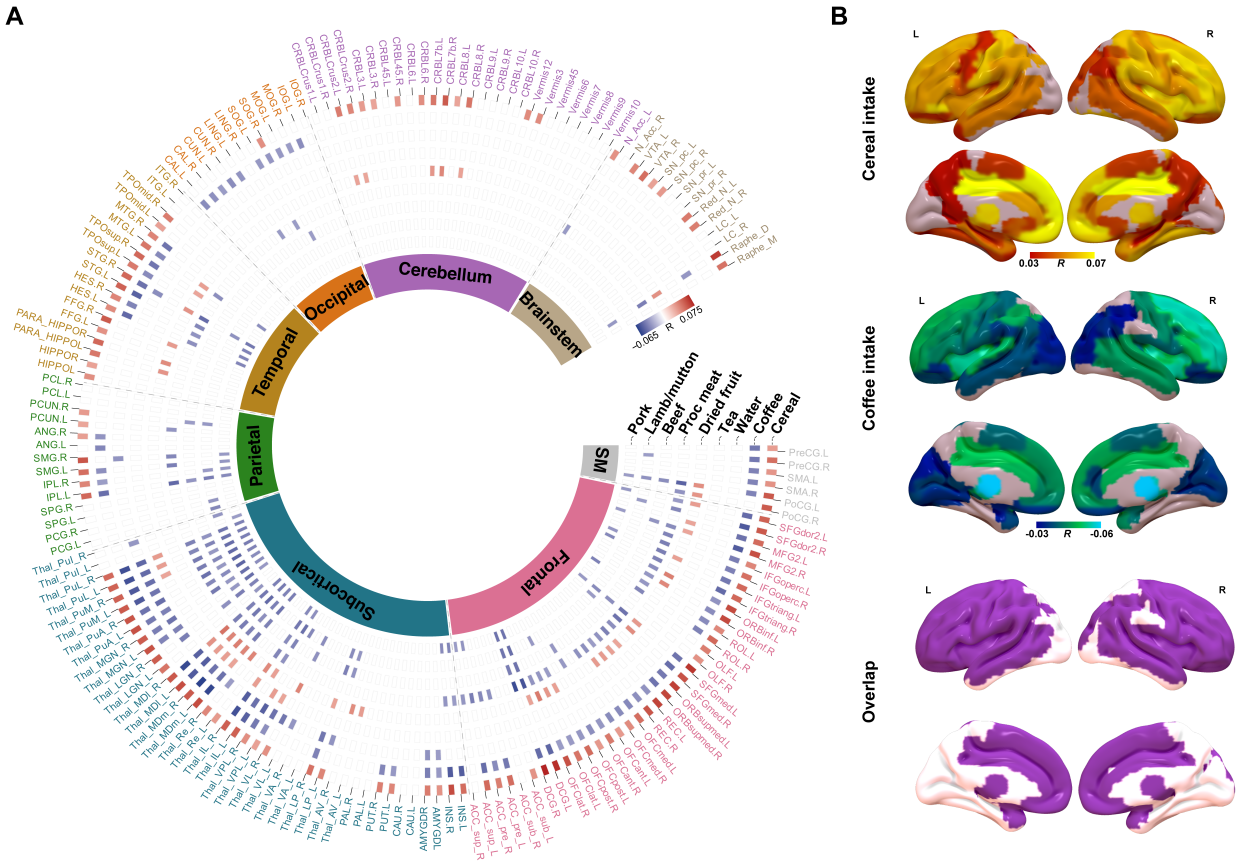

Fig. 1. Correlations between grey matter volume (GMV) and different daily diets. (A) Circular heatmap of correlations between GMVs of 166 brain regions from AAL3 (the outer layer) and different diets (along radius). As indicated by the colour bar, positive correlations were highlighted in red while negative correlations were highlighted in blue. The inner layer indicates the lobes that brain regions belong to. (B) Brain regions with significant correlations between their GMVs and the intake of cereal (upper) and coffee (middle), as well as the overlapped significant regions (bottom). SM: Sensorimotor.

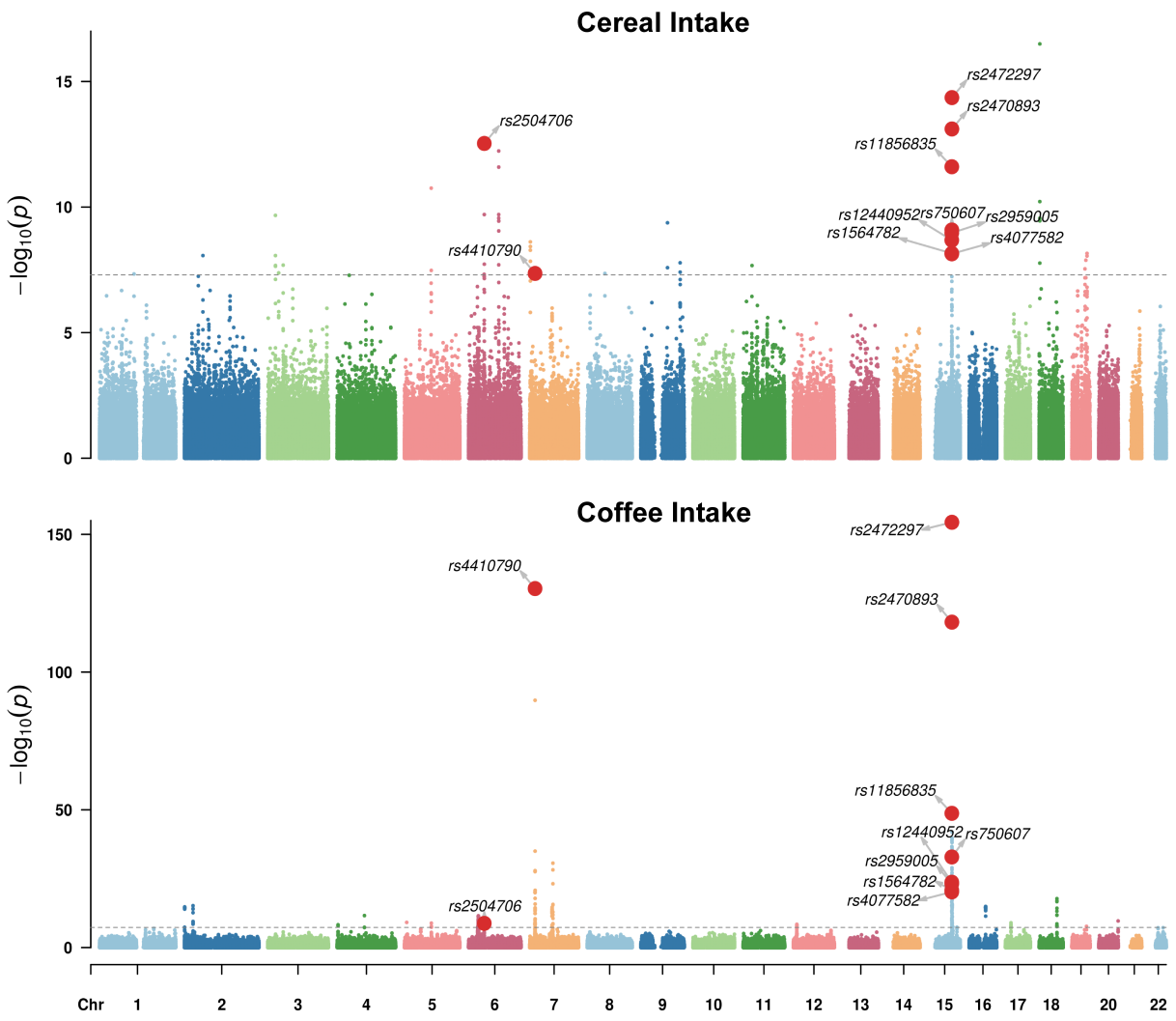

Fig. 2. Manhattan plots of the genome-wide association results for the intake of cereal (upper) and coffee (bottom). The gray line indicates the genome-wide significance level (i.e. P-value=5E-08). Variants with significant associations with both cereal and coffee intake were highlighted with red dots.

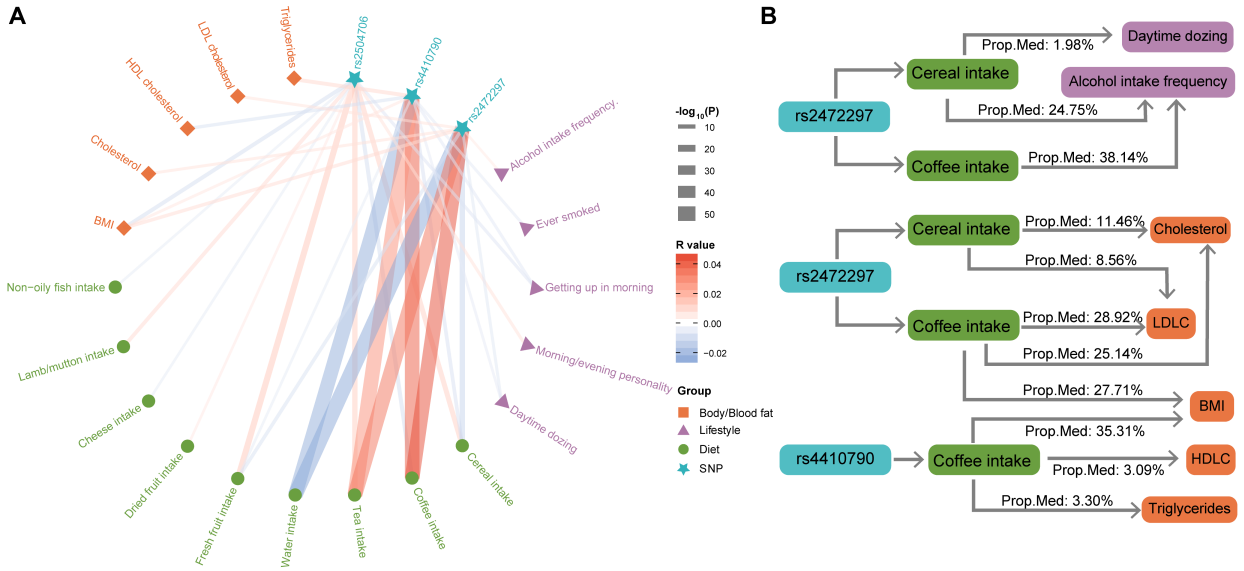

Fig. 3. Relationships between the lead SNPs and diets, lifestyle, and body/blood fat. (A) Associations between the three lead SNPs (of both cereal and coffee intake) and other diets, lifestyle, and body/blood fat levels. The colour of each line represents the correlation coefficient (positive in red and negative in blue), and the thickness of each line represents the -$\log_{10}$ P-value (capped at 50) of the corresponding correlation. (B) Proposed mediation models of genetic variants, body/blood fat levels, lifestyles, and the intake of cereal and coffee. Prop. Med: proportion of mediation. BMI: Body mass index, HDL: High-density lipoprotein, LDL: Low-density lipoprotein.

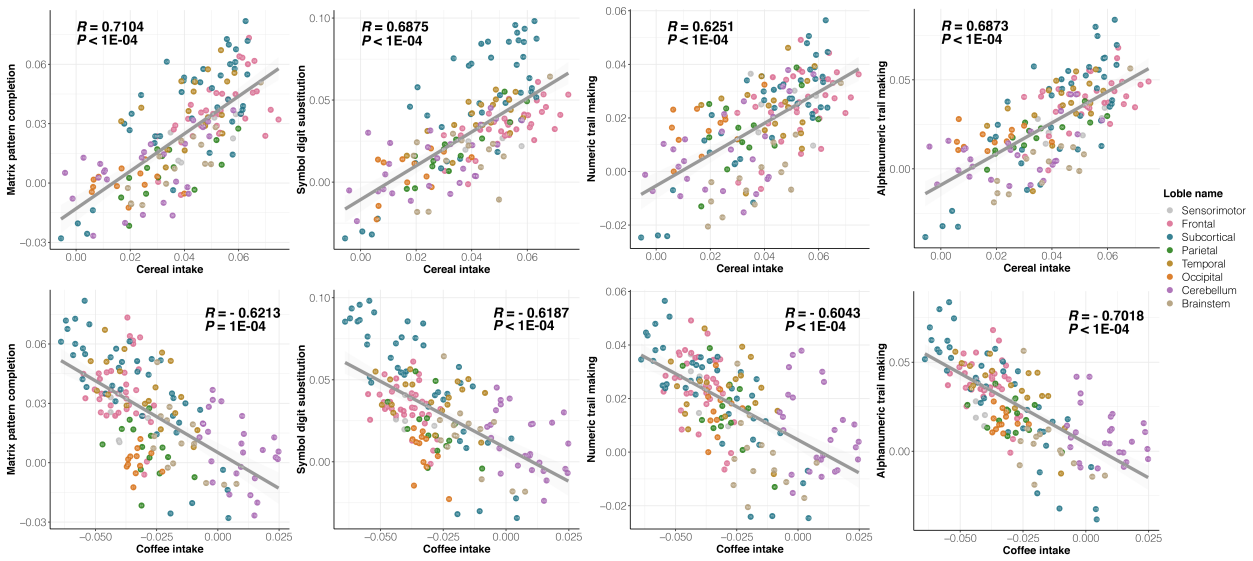

Fig. 4. Scatter plots of brain-wide GMV-association patterns of cognitive functions and the intake of cereal (upper) and coffee (bottom). Each dot represents one of 166 AAL3 brain regions, where the colours indicate at which lobes the brain regions were located.