nature mental health

Article

https://doi.org/10.1038/s44220-024-00226-0

Associations of dietary patterns with brain health from behavioral, neuroimaging, biochemical and genetic analyses

Received: 31 August 2023

Accepted: 5 March 2024

Published online: 1 April 2024

Check for updates

Ruohan Zhang[1,10], Bei Zhang[2,3,10], Chun Shen[2,3,10], Barbara J. Sahakian[2,4,5], Zeyu Li[2,3], Wei Zhang[2,3], Yujie Zhao[2,3], Yuzhu Li[2,3], Jianfeng Feng[1,2,3,6,7,8] ✉ & Wei Cheng[2,3,6,9] ✉

Food preferences significantly influence dietary choices, yet understanding natural dietary patterns in populations remains limited. Here we identifiy four dietary subtypes by applying data-driven approaches to food-liking data from 181,990 UK Biobank participants: 'starch-free or reduced-starch' (subtype 1), 'vegetarian' (subtype 2), 'high protein and low fiber' (subtype 3) and 'balanced' (subtype 4). These subtypes varied in diverse brain health domains. The individuals with a balanced diet demonstrated better mental health and superior cognitive functions relative to other three subtypes. Compared with subtype 4, subtype 3 displayed lower gray matter volumes in regions such as the postcentral gyrus, while subtype 2 showed higher volumes in thalamus and precuneus. Genome-wide association analyses identified 16 genes different between subtype 3 and subtype 4, enriched in biological processes related to mental health and cognition. These findings provide new insights into naturally developed dietary patterns, highlighting the importance of a balanced diet for brain health.

Food-liking is a complex trait that reflects the hedonic response to food for individuals[1] and is considered to be the most influential factor driving food choices and intake[2]. With an abundance of food choices available worldwide, people naturally develop diverse dietary patterns. Recently, growing evidence has highlighted that the profound impact of dietary patterns on health, including chronic medical diseases, such as cardiovascular disease[3], type 2 diabetes[4], metabolic syndrome[5] and cancer[6], as well as mental health and/or cognitive impairments[7–10], such as major depression disorders and anxiety. Understanding how diet preferences affect health, especially brain health, is critical for developing targeted dietary interventions to promote the consumption of nourishing foods and improve the landscape of brain health.

Previous evidence has demonstrated a strong link between diet and both cognitive functions and mental health. For example, a systematic review focusing on various dietary intake patterns and cognitive functions revealed associations such as increased consumption of simple carbohydrates (for example, sugars) being linked to decreased overall cognitive performance, while saturated fatty acids were associated with reduced memory and learning. Conversely, protein intake was found to potentially enhance executive function and working memory[7]. Furthermore, unhealthy diets have been implicated as a risk factor for a wide range of psychiatric disorders, including major depression disorders[11–13], anxiety[14], bipolar disorder[15,16], stroke[17], sleep problems[18,19] and Alzheimer's disease[20]. For instance, individuals with a 'Western dietary

[1]Department of Computer Science, University of Warwick, Coventry, UK. [2]Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China. [3]Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence at Fudan University, Ministry of Education, Shanghai, China. [4]Department of Psychiatry, University of Cambridge, Cambridge, UK. [5]Behavioural and Clinical Neuroscience Institute, University of Cambridge, Cambridge, UK. [6]MOE Frontiers Center for Brain Science, Fudan University, Shanghai, China. [7]Zhangjiang Fudan International Innovation Center, Shanghai, China. [8]School of Data Science, Fudan University, Shanghai, China. [9]Department of Neurology at Huashan Hospital, Fudan University, Shanghai, China. [10]These authors contributed equally: Ruohan Zhang, Bei Zhang, Chun Shen. ✉e-mail: jianfeng64@gmail.com; wcheng@fudan.edu.cn

pattern' (who preferred high sweet and fatty food but not plant-based food) showed a higher incidence of depression[11–13] relative to those following a balanced diet (including a balanced amount of vegetables, fruits, cereals, nuts, seeds, pulses, moderate dairy, eggs and fish)[15,21–23].

The linked diet and cognition and mental health might be related to alterations in molecular biomarkers as well as changes in brain structure and functions. Nutrition research suggests that the relationship between dietary patterns and mental disorders (or cognitive functions) could be potentially mediated by the gut–brain axis. Specific dietary patterns, such as the 'Western diet' have the potential to disrupt the balance of gut microbiota, leading to inflammation and oxidative stress, which can impair cognitive function and increase the risk of mental disorders[24–27]. Additionally, neuroimaging studies have revealed associations between dietary patterns and functions and structures in brain regions[28–31], emphasizing the intricate relationship between diet and brain health. For instance, higher adherence to the 'Mediterranean-type diet' (characterized by high consumption of fruit, vegetables, legumes and cereals, with olive oil as the primary source of fat and a low intake of red meat and poultry) that typically linked with reduced risk of Parkinson's disease and Alzheimer's disease was associated with lower reduction of total brain volume over a 3 year period[29], as well as with larger cortical thickness in key brain regions, such as the entorhinal cortex, posterior cingulate cortex, orbitofrontal cortex and inferior and middle temporal gyrus[31]. While previous research has established strong links between dietary and various domains of brain health, the complex relationships and regulation mechanisms underlying different domains of brain health remain poorly understood.

Moreover, based on the quantities, variety or combination of different foods and beverages in diets and the frequency with which they are habitually consumed, several traditional dietary patterns have emerged[32], such as the 'Western dietary pattern' and the 'Mediterranean dietary pattern', as described above, as well as the 'prudent dietary pattern' (characterized by a high intake of vegetables, fruit, legumes, whole grains and fish and other seafood)[33] and the 'vegetarian/plant-based dietary pattern' (a dietary pattern that excludes meat, meat-derived foods and, to different extents, other animal products)[34]. While extensive research has explored the links between these dietary patterns and brain health, findings across studies are not consistently aligned. For example, some studies associated the vegetarian dietary pattern with higher depression and anxiety[35,36], while others found the opposite effect[37,38] or no effect[39,40]. This variation may be attributed to limited sample sizes and different scopes and criteria used for defining dietary patterns. For instance, differences may arise from considerations such as whether individuals consuming dairy products are categorized within the 'vegetarian/plant-based dietary pattern'[34,41]. In addition, these studies tend to focus on specific populations adhering to a single dietary pattern, leaving a critical gap in understanding the relationship between dietary patterns and brain health in other populations. Thus, a universally recognized and reliable dietary pattern classification system within a large-scale population is warranted.

In this Article, to narrow these gaps, the current study leverages the large-scale dataset from the UK Biobank and employs data-driven approaches to identify naturally developed dietary patterns and their associations with cognitive function, mental health, blood and metabolic biomarkers, brain imaging and genomics. Specifically, we first utilized food-liking data from the UK Biobank participants and applied principal component analysis (PCA) and hierarchical clustering techniques to develop subtypes of food-liking. Subsequently, through one-way analysis of covariance (ANCOVA), we assessed differences in various brain health domains among these subtypes, including mental health, cognitive functions, blood and metabolism biomarkers, and brain magnetic resonance imaging (MRI) traits. We also examined differences among these subtypes by analyzing longitudinal data on mental disorders via Cox proportional hazards models. Third, structural equation models (SEMs) were employed to explore the relationships between dietary patterns and different aspects of brain health. Fourth, genome-wide association analysis (GWAS) and gene expression and enrichment analysis were conducted to investigate the genetic underpinnings of distinct subtypes of food-liking and potential biological pathways. This study pioneeringly represents the large-scale exploration of food preferences and their comprehensive associations with brain health. By exploring these intricate connections, our research lays the groundwork for further investigations and potential interventions that can significantly impact human health on a global scale, underscoring the importance of understanding the intricate relationship between diet and brain health.

Results

Distinct food-liking profiles of the four subtypes

A total of 181,990 participants (mean age 70.7 ± 7.7 years and 57.08% female) from the UK Biobank were included in the identification of food-liking subtypes. Supplementary Fig. 1 provides a general schema of the current study. First, 140 food and beverage items were classified into ten food categories, and PCA was performed separately for each category. Using this approach, we obtained a total of 83 principal components, which were used as input for hierarchical clustering. The dendrogram of the clustering results showed that participants could be grouped into four distinct food-liking subtypes (Fig. 1a), with proportions of 18.09%, 5.54%, 19.39% and 56.98% for subtypes 1 to 4, respectively. The demographic characteristics of the four subtypes were summarized in Supplementary Table 3.

To characterize the food preferences of the four subtypes, we generated a radar chart to visualize the liking scores of the ten food categories for the four subtypes (Fig. 1b). Subtype 1 showed a higher preference for fruits, vegetables and protein foods but a lower preference for starches, which is consistent with a 'starch-free or reduced-starch dietary pattern'. Subtype 2 displayed a stronger preference for fruits and vegetables, while showing a lower preference for protein foods, which is similar to a 'vegetarian dietary pattern'. Subtype 3 exhibited a greater preference for snacks and protein foods but a lower preference for fruits and vegetables, resembling the 'high protein and low fiber dietary pattern'. Finally, subtype 4 showed balanced preferences across all food categories, which can be regarded as a 'balanced dietary pattern'. To further validate the suitability of clustering into four subtypes, we utilized the silhouette criterion[42] to determine the optimal number of clusters. Our analysis encompassed cluster numbers ranging from two to seven, as visualized in Supplementary Fig. 2a. The results indicated that the most suitable numbers of food-liking subtypes did not exceed four. In addition to the four subtypes shown in Fig. 1b, we also examined a radar chart depicting three subtypes (Supplementary Fig. 2b). It is noteworthy that one of these three subtypes is a combination of two of the four subtypes (subtype 1 and subtype 2) displayed in Fig. 1b, while the other two subtypes closely resemble two subtypes from Fig. 1b. Furthermore, the radar chart of the four subtypes (Fig. 1b) exhibited distinct food-liking characteristics, indicating an intriguing and meaningful dimension to our exploration of dietary patterns within a large population.

Additionally, we assessed the robustness of our findings in the context of data imputation by utilizing nonimputed data from 72,419 participants for the identification of food-liking subtypes. The radar chart depicting the four subtypes identified among the 72,419 participants without imputation closely mirrored the one generated from the imputed data of 181,990 participants (Supplementary Fig. 3a). This consistency indicated the robustness of our findings with data imputation. Moreover, the food preference characteristics of the four subtypes, as determined using PCA with explained variance ratios of 70% and 90% (Supplementary Fig. 3b,c), both exhibited a strong resemblance to the subtypes identified at variance ratios of 80% (Fig. 1b). This finding indicated the robustness of the explained

https://doi.org/10.1038/s44220-024-00226-0

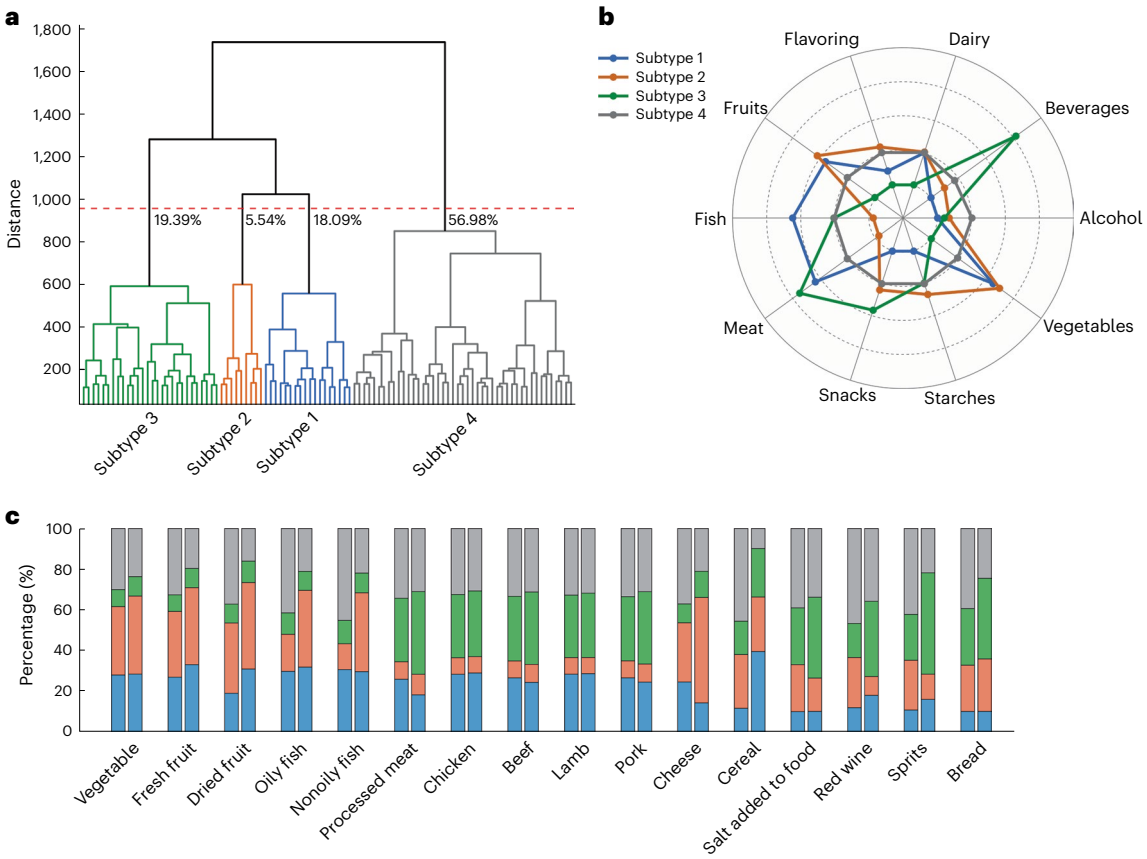


Fig. 1 | Food-liking profiles of the four subtypes. a, The dendrogram resulting from the hierarchical clustering of food preference data from 181,990 participants, revealing four distinct subtypes. The red dashed line indicates the delineation of four subtypes. **b**, The radar chart depicting the preference scores of the ten food categories for each subtype. **c**, Comparisons between food-liking and food-consumption traits using relative scores of the four subtypes. The selected foods cover a range of categories analyzed in this study. The food-liking measures are shown to be closely related to food consumption. The four identified subtypes include subtype 1, 'starch-free or reduced-starch dietary pattern', subtype 2, 'vegetarian dietary pattern', subtype 3, 'high protein and low fiber dietary pattern' and subtype 4 'balanced dietary pattern'.

variance of the derived components and validated the reliability of hierarchical clustering results based on PCA components.

To investigate the potential corresponding relationship between food liking and food consumption, we further calculated the average scores for specific food traits within each subtype, covering various food items. These food traits encompassed a variety of categories, including vegetables, fruits, several types of meat and alcohol, as well as cereal and bread. The results revealed a consistent alignment between the relative scores of the food-liking and food-consumption traits across all four subtypes. This congruence in scores indicated a robust relationship between individual preferences for certain foods and their actual consumption patterns (Fig. 1c).

Subtype-specific mental health and cognitive function

Before one-way ANCOVA analyses, we conducted Levene's tests to confirm that the data satisfied the assumption of the equality of variances ($P > 0.05$). After adjusting for covariates and applying the Bonferroni correction, one-way ANCOVAs with the factor of subtype revealed significant main effects on seven mental health measures, including anxiety symptoms ($F = 41.5$ and $P = 8.9 \times 10^{-27}$), depressive symptoms ($F = 71.4$ and $P = 3.9 \times 10^{-46}$), mental distress ($F = 62.1$ and $P = 4.0 \times 10^{-40}$), psychotic experience ($F = 17.4$ and $P = 2.6 \times 10^{-11}$), self-harm ($F = 116.8$ and $P = 1.6 \times 10^{-75}$), trauma ($F = 155.4$ and $P = 1.6 \times 10^{-100}$) and well-being ($F = 256.8$ and $P = 3.8 \times 10^{-166}$). Figure 2a depicts the subtype-specific patterns of mental health measures. By visual inspection, subtype 4 scored the lowest in most mental health measures and the highest in well-being, indicating better mental health conditions. Subtype 2 and subtype 3 had relatively higher scores in some mental health measures, such as anxiety and depressive symptoms, and a relatively lower level of well-being (Fig. 2a).

In addition, similar analysis was conducted for four cognitive functions, which also exhibited significant main effects on four subtypes, including fluid intelligence ($F = 15.0$ and $P = 8.9 \times 10^{-10}$), pairs matching ($F = 6.6$ and $P = 2.0 \times 10^{-4}$), reaction time ($F = 20.1$ and $P = 5.2 \times 10^{-13}$) and symbol–digit substitution ($F = 18.6$ and $P = 4.9 \times 10^{-12}$). Specifically, subtype 4 had the second-highest correct number of symbol digit matches and the lowest reaction time. Subtype 3 showed the highest correct number of symbol digit matches and the second-lowest reaction time (Fig. 2a).

To further investigate the differences among four subtypes in the risks of mental disorders, we employed Cox proportional hazards regression models, with subtype 4 as the reference group. The Cox model results showed significant differences in the risks of four mental disorders among the four subtypes after false discovery rate (FDR) corrections (adjusted P value <0.05), particularly in anxiety, depression, eating disorder and stroke (Fig. 2b). The P values for Schoenfeld's global test of the Cox models for anxiety and depression were both 0.2, indicating that the proportional hazards assumption was met. The Cox model for stroke satisfied the proportional hazards assumption after stratification by age (68 years). Additionally, the Cox model for eating disorder, when stratified by body mass index (BMI, ≥25 kg m^{-2}), also met the proportional hazards assumption. Specifically, when compared

https://doi.org/10.1038/s44220-024-00226-0

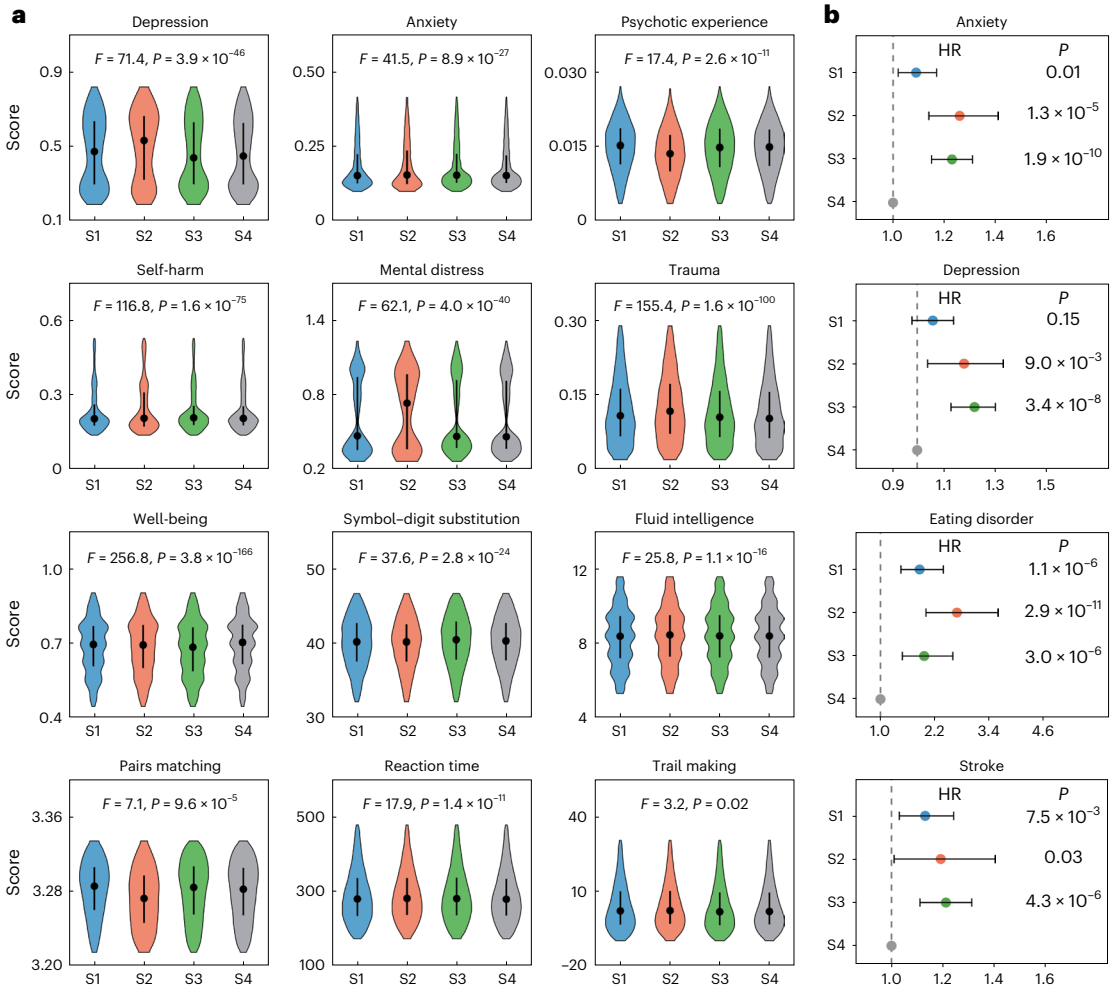


Fig. 2 | Subtype-specific patterns of mental health measures, cognitive function and mental disorder risk. a, The phenotypic differences in mental health (for all mental health symptoms, $n = 118,616$) and cognitive function, including the symbol–digit substitution test ($n = 93,325$), fluid intelligence ($n = 96,742$), pairs matching ($n = 93,394$), reaction time ($n = 179,740$) and trail making ($n = 82,375$), among the four subtypes: subtype 1 (S1), 'starch-free or reduced-starch dietary pattern', subtype 2 (S2), 'vegetarian dietary pattern', subtype 3 (S3), 'high protein and low fiber dietary pattern' and subtype 4 (S4), 'balanced dietary pattern'. These differences were determined by ANCOVA analyses after Bonferroni correction ($\alpha = 0.05$). The analysis controlled for several covariates, including age, BMI, education qualificationsf and Townsend deprivation index. For the cognitive tests (including reaction time and trail making), a lower score indicates better cognitive functions. For the other cognitive tests, a higher score indicates better cognitive performance. The data are presented using a violin plot (median point; upper and lower quartiles). **b**, Forest plots depicting Cox proportional hazards models for the risks of mental disorders, including anxiety ($n = 152,014$), depression ($n = 145,350$), eating disorder ($n = 165,158$) and stroke ($n = 152,730$), with subtype 4 as the reference group. The results are presented using HRs and their corresponding 95% CI. The significance of coefficients in the Cox models was evaluated using the Wald test (two-tailed P value). The analyses were adjusted for confounding factors, including sex, age, BMI, education qualifications and Townsend index, with FDR corrections for multiple comparisons (adjusted P value <0.05). The horizontal gray dashed line represents no effect (i.e., hazard ratio = 1). The gray dot represents S4 as the reference in the Cox models.

with subtype 4, both subtype 2 and subtype 3 exhibited a higher risk for depression, with hazard ratios (HRs) of 1.18 (95% confidence interval (CI), 1.04–1.33 and adjusted $P = 0.03$) and 1.22 (95% CI 1.13–1.30 and adjusted $P = 3.8 \times 10^{-7}$), respectively. However, no significant difference in HRs for this mental disorder was found between subtype 1 and subtype 4. Furthermore, subtype 1 and subtype 3 displayed higher risks than subtype 4 for stroke, with HRs of 1.13 (95% CI 1.03–1.24 and adjusted $P = 0.03$) and 1.21 (95% CI 1.11–1.31 and adjusted $P = 2.3 \times 10^{-5}$), respectively. However, no significant difference in HRs for this mental disorder was observed between subtype 2 and subtype 4. Additionally, all three subtypes, when compared with subtype 4, exhibited higher risks for anxiety (subtype 1, HR 1.09, 95% CI 1.0–1.17 and adjusted $P = 0.03$; subtype 2, HR 1.26, 95% CI 1.14–1.41 and adjusted $P = 6.2 \times 10^{-5}$; subtype 3, HR 1.23, 95% CI 1.15–1.31 and adjusted $P = 3.2 \times 10^{-9}$) and eating disorder, with subtype 2 showing particularly significant risk (subtype 1, HR 1.86, 95% CI 1.45–2.38 and adjusted $P = 9.1 \times 10^{-6}$; subtype 2, HR 2.68, 95% CI 2.00–3.58 and adjusted $P = 9.7 \times 10^{-10}$; subtype 3, HR 1.96, 95% CI 1.48–2.59 and adjusted $P = 2.0 \times 10^{-5}$).

Distinctive blood and metabolic biomarker across subtypes

The one-way ANCOVA analyses revealed that 167 of 229 blood and metabolomic biomarkers (32 blood biomarkers and 135 metabolomic biomarkers) were significantly different between the four subtypes after Bonferroni correction ($P < 0.05/229$) (Fig. 3a). The data satisfied the assumption of the equality of variances and the results were adjusted for covariates. The top 10% of significant biomarkers included the following categories: fatty acids (for example, docosahexaenoic acid ($F = 312.5$ and $P = 1.1 \times 10^{-200}$) omega-3 fatty acids ($F = 232.4$ and $P = 1.4 \times 10^{-149}$) and omega-6 fatty acids ($F = 18.4$ and $P = 6.1 \times 10^{-12}$)), amino acids (for example, glycine ($F = 122.1$ and $P = 1.0 \times 10^{-80}$)),

https://doi.org/10.1038/s44220-024-00226-0

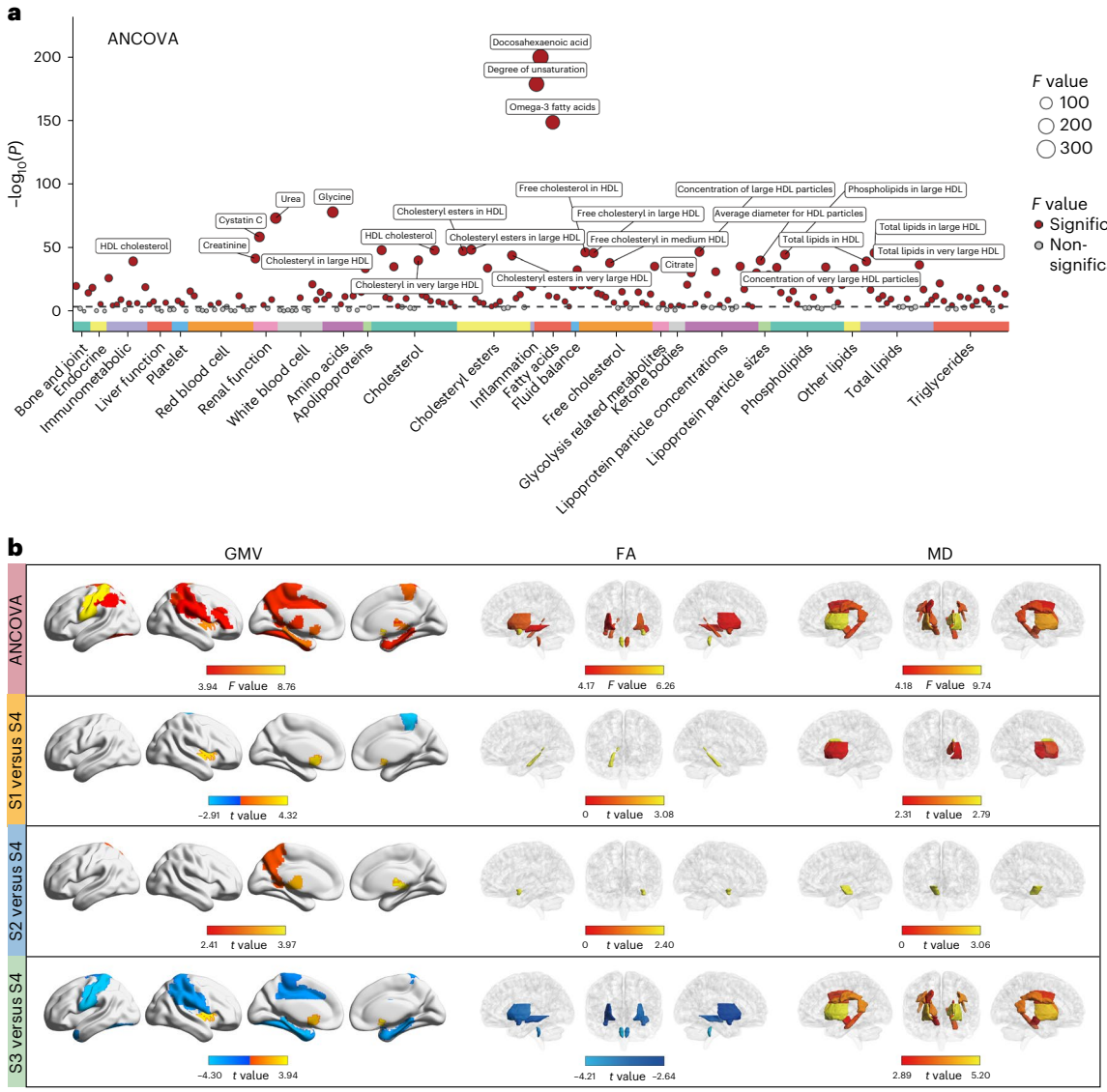


Fig. 3 | Differences in blood and metabolic biomarkers, as well as brain morphology and white matter integrity across the four subtypes. **a**, Manhattan plot of the one-way ANCOVA analyses (*F*-tests) for 24 categories of blood and metabolic biomarkers. The height of each point represents the negative logarithm of the *P* value of the *F*-test, with the color bar indicating the different biomarker categories. The black dashed line represents the Bonferroni threshold for multiple comparisons ($\alpha = 0.05$), and the top 15% of biomarkers that exhibited significant differences after Bonferroni correction ($P < 0.05/229$) are labeled using text annotations. The analysis was adjusted for covariates, including age, BMI, education qualifications and Townsend deprivation index. **b**, Brain regions that significantly differ in GMV, mean FA and MD among the four subtypes identified in the one-way ANCOVA analyses (*F*-tests). Multiple comparisons were corrected using FDR correction (adjusted *P* value <0.05). The analyses were adjusted for age, BMI, education qualifications, Townsend deprivation index, scanning sites and intracranial volume (the latter only for the analysis of GMV). The results of post hoc tests (two-tailed *t*-tests) on GMV, mean FA, and mean MD comparing subtype 1 (S1, 'starch-free or reduced-starch dietary pattern'), subtype 2 (S2, 'vegetarian dietary pattern') and subtype 3 (S3, 'high protein and low fiber dietary pattern') against subtype 4 (S4, 'balanced dietary pattern'), with FDR correction for multiple comparisons (adjusted *P* value <0.05). The same covariates were regressed out in the post hoc tests as in the ANCOVAs.

renal function (for example, urea ($F = 114.8$ and $P = 5.2 \times 10^{-74}$)), cholesteryl esters (for example, cholesteryl esters in large high-density lipoprotein (HDL) ($F = 76.1$ and $P = 4.1 \times 10^{-49}$)), cholesterol (for example, cholesterol in large HDL ($F = 75.5$ and $P = 1.0 \times 10^{-48}$)), lipoprotein particle concentrations (for example, concentration of large HDL particles ($F = 73.5$ and $P = 2.0 \times 10^{-47}$)), free cholesterol (for example, free cholesterol in HDL ($F = 72.9$ and $P = 4.9 \times 10^{-47}$)), total lipids (for example, total lipids in large HDL ($F = 72.2$ and $P = 1.5 \times 10^{-46}$)) and phospholipids (for example, phospholipids in large HDL ($F = 70.1$ and $P = 3.4 \times 10^{-45}$)).

Post hoc analysis further revealed that 127 of 167 blood and metabolomic biomarkers were significantly different between subtype 3 and subtype 4 after Bonferroni correction ($P < 0.05/(167 \times 3)$) (Supplementary Fig. 4c), with most of them being lower in subtype 3. The top 10% of significant biomarkers included fatty acids (for example, docosahexaenoic acid ($t = -25.7$, Cohen's $d = -0.3$ and $P = 3.7 \times 10^{-144}$) and omega-3 fatty acids ($t = -21.3$, Cohen's $d = -0.3$ and $P = 6.4 \times 10^{-100}$)), cholesteryl esters (for example, cholesteryl esters in HDL ($t = -14.2$, Cohen's $d = -0.2$ and $P = 6.2 \times 10^{-46}$)) and cholesterol (for example, HDL cholesterol, $t = -14.3$, Cohen's $d = -0.2$ and $P = 1.7 \times 10^{-46}$)).

Compared with subtype 4, subtype 1 also showed significantly different in 49 blood and metabolomic biomarkers (Supplementary Fig. 4a), with most of them being higher in subtype 1 (Supplementary Fig. 4a), such as fatty acids (for example, degree of unsaturation ($t = 8.7$, Cohen's $d = 0.1$ and $P = 4.0 \times 10^{-18}$) and docosahexaenoic acid

https://doi.org/10.1038/s44220-024-00226-0

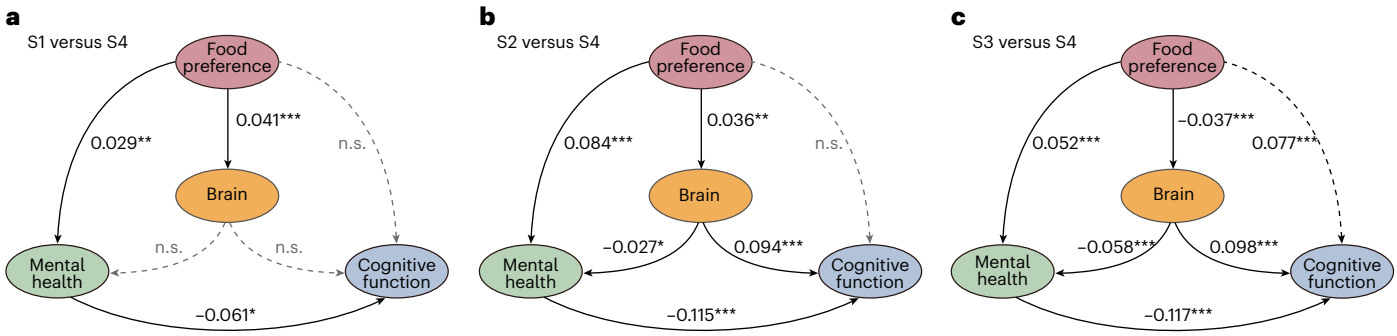


Fig. 4 | Directional associations among food preference, mental health, cognitive function and brain MRI trait. a, The results of a structural equation model comparing subtype 1 (S1, 'starch-free or reduced-starch dietary pattern') to subtype 4 (S4, 'balanced dietary pattern'). The analysis revealed that food preference was significantly associated with mental health ($\beta = 0.029$ and $P_{adj} = 0.009$) and brain MRI trait ($\beta = 0.041$ abd $P_{adj} = 6.1 \times 10^{-5}$). Mental health significantly predicted cognitive function ($\beta = -0.061$ and $P_{adj} = 0.03$). **b**,**c**, The structural equation model for subtype 2 (S2, 'vegetarian dietary pattern') (**b**) and subtype 3 (S3, 'high protein and low fiber dietary pattern') (**c**) versus subtype 4, respectively. All associations were in the expected direction, and all paths in the model for subtype 3 versus subtype 4 were significant. Wald tests were utilized to derive the two-sided *P* values of the standardized coefficients adjusted for multiple comparisons (FDR correction). The significance levels of the standardized coefficients are indicated by $^{*}P_{adj} < 0.05$, $^{**}P_{adj} < 0.01$ and $^{***}P_{adj} < 0.001$. n.s., not significant.

($t = 6.3$, Cohen's $d = 0.1$ and $P = 3.8 \times 10^{-10}$)). Some biomarkers showed lower levels in subtype 1, such as phospholipids (for example, phospholipids in small HDL ($t = -5.8$, Cohen's $d = -0.1$ and $P = 8.9 \times 10^{-9}$)), fatty acids (for example, saturated fatty acid (Cohen's $d = -0.1$, $t = -5.3$ and $P = 9.9 \times 10^{-8}$)) and total lipids (for example, total lipids in small HDL ($t = -5.5$, Cohen's $d = -0.1$ and $P = 3.2 \times 10^{-8}$)).

The results comparing of subtype 2 with subtype 4 were slightly different from those of comparing subtype 3 (or subtype 1) and subtype 4 (Supplementary Fig. 4b). After Bonferroni correction, 72 of the 167 blood and metabolomic biomarkers were found to be significantly different between subtype 2 and subtype 4 ($P < 0.05/(167 \times 3)$). The top 10% of significant biomarkers were renal function (for example, urea ($t = -18.7$, Cohen's $d = -0.4$ and $P = 2.5 \times 10^{-77}$)), amino acids (for example, glycine ($t = 18.9$, Cohen's $d = 0.4$ and $P = 3.1 \times 10^{-79}$)) and fatty acids (for example, omega-3 fatty acids ($t = -14.4$, Cohen's $d = -0.3$ and $P = 8.7 \times 10^{-47}$)). For complete statistical results of the analyses of blood and metabolic biomarkers, please refer to Supplementary Tables 6 and 7.

Differences in neuroimaging phenotypes across subtypes

In the one-way ANCOVA analyses, 23 of the 94 brain regions of the AAL2 atlas were significantly different among the four subtypes after applying FDR correction for multiple comparisons (adjusted *P* value <0.05) (Fig. 3b). The assumption of the equality of variances were satisfied for conducting one-way ANCOVAs ($P > 0.05$). These regions included the postcentral gyrus, caudate, putamen, parahippocampal gyrus and so on. Post hoc analysis further revealed significant differences in 16 out of the 23 brain regions between subtype 3 and subtype 4 after FDR correction (adjusted *P* value <0.05), among which 11 brain regions showed significantly lower in subtype 3, such as postcentral gyrus, parahippocampal gyrus, and inferior parietal gyrus (Fig. 3b). Additionally, subtype 1 showed significant differences compared with subtype 4 in seven regions, including the putamen, caudate, pallidum and paracentral lobule (Fig. 3b). Only four brain regions (thalamus, precuneus and paracentral lobule) were found to be significantly different between subtype 2 and subtype 4 after FDR correction (adjusted *P* value <0.05), which exhibited higher in subtype 2 (Fig. 3b). The complete statistical results for the analyses of the gray matter volume (GMV) are provided in Supplementary Tables 8 and 9.

Moreover, we performed analogous analyses on the diffusion tensor imaging measures of fractional anisotropy (FA) and mean diffusivity (MD) for the 48 white matter tracts within the John's Hopkins University (JHU) ICBM-DTI-81 white-matter labels atlas. Our results revealed, for the FA measurements, eight brain regions of interest (ROIs) differed significantly across the four subtypes after FDR correction with an adjusted *P* value <0.05 (Fig. 3b). These ROIs included the medial lemniscus, uncinate fasciculus and external capsule and so on. Post hoc analyses revealed significant differences in FA measures of seven ROIs between subtype 3 and subtype 4 after FDR correction (adjusted *P* value <0.05), with all these ROIs exhibiting lower FA values in subtype 3. Additionally, the cingulum hippocampus was the only region that showed significant differences between subtype 1 and subtype 4, with higher FA values in subtype 1. The uncinate fasciculus was the only region that exhibited significant differences between subtype 2 and subtype 4, with higher FA values in subtype 2 (Fig. 3b).

In terms of the MD measures, we found that 11 ROIs were significantly different across four subtypes, including external capsule, anterior limb of the internal capsule, superior fronto-occipital fasciculus and so on (Fig. 3b). Post hoc analysis of comparison of subtype 3 and subtype 4 mirrored these brain regions after FDR correction (adjusted *P* value <0.05), with higher MD values in all of these ROIs in subtype 3. Also, three ROIs showed significant differences between subtype 1 and subtype 4, such as the superior fronto-occipital fasciculus, anterior limb of the internal capsule and external capsule (Fig. 3b), with higher MD values in subtype 1. Only the cerebral peduncle was found to be significantly different between subtype 2 and subtype 4, with higher MD values in subtype 2. Complete statistical results for the ANCOVA analyses and post hoc tests of FA and MD measures can be found in Supplementary Tables 10–13.

Polygenic risk scores (PRSs) for mental disorders across subtypes

After adjusting covariates and applying the Bonferroni correction ($P < 0.05/8$), one-way ANCOVA analyses on eight PRSs of mental disorders revealed significant main effects across four subtypes (Fig. 5d), including a PRS for Alzheimer's disease ($F = 8.2$ and $P = 1.9 \times 10^{-5}$), ischemic stroke ($F = 8.3$ and $P = 1.7 \times 10^{-5}$), Parkinson's disease ($F = 6.7$ and $P = 1.7 \times 10^{-4}$), cardiovascular disease ($F = 6.0$ and $P = 4.2 \times 10^{-4}$), bipolar disorder ($F = 34.1$ and $P = 4.8 \times 10^{-22}$), schizophrenia ($F = 72.4$ and $P = 8.7 \times 10^{-47}$), depression ($F = 11.5$ and $P = 1.5 \times 10^{-7}$) and suicide attempt ($F = 6.4$ and $P = 2.5 \times 10^{-4}$). Levene's tests confirmed that the data satisfied the assumption of the equality of variances ($P > 0.05$). Specifically, subtype 2 showed a higher genetic predisposition for several mental disorders, including Alzheimer's disease, Parkinson's disease, bipolar disorder, schizophrenia and suicide attempts, than other subtypes, mirroring the comparisons on mental health measures

('Differences in neuroimaging phenotypes across subtypes' section). In addition, subtype 3 presented a high genetic susceptibility to ischemic stroke, whereas subtype 4 showed relatively lower genetic risks for most mental disorders, which was consistent with the results on mental health measures ('Differences in neuroimaging phenotypes across subtypes' section).

Complex interplay of food preferences and other phenotypes

To examine the complex relationships among food preference, mental health, cognitive function and brain MRI features, we constructed three SEMs with subtype 4 as the reference group. In the model that compared the food preference of subtype 3 with that of subtype 4, we selected those latent variables that were significantly different between subtype 3 and subtype 4. Specifically, the mental health measures encompassed anxiety symptoms ($\beta = 0.63$ and $P < 0.001$), depressive symptoms ($\beta = 0.73$ and $P < 0.001$), self-harm ($\beta = 0.54$ and $P < 0.001$), trauma ($\beta = 0.55$ and $P < 0.001$) and well-being ($\beta = -0.57$ and $P < 0.001$). The cognitive function was characterized by fluid intelligence, reaction time and symbol–digit substitution ($\beta = 0.38$ and -0.20 and 0.67, respectively; $P < 0.001$). The brain MRI traits included the GMV of the top ten brain regions, the mean FA of all seven white matter tracts and the mean MD of the top ten white matter tracts. Figure 4c depicts the directional association results. The food preference was significantly associated with mental health measurements ($\beta = 0.052$ and $P_{adj} = 5.5 \times 10^{-6}$), brain MRI traits ($\beta = -0.037$ and $P_{adj} = 4.6 \times 10^{-4}$) and cognitive function ($\beta = 0.077$ and $P_{adj} = 3.5 \times 10^{-8}$). Additionally, brain MRI traits and mental health were significant predictors for cognitive functions ($\beta = 0.098$ and $P_{adj} = 9.2 \times 10^{-11}$ and $\beta = -0.117$ and $P_{adj} = 1.1 \times 10^{-12}$, respectively). The brain MRI traits significantly predicted mental health ($\beta = -0.058$ and $P_{adj} = 1.5 \times 10^{-6}$). The root mean square error of approximation (RMSEA) of this SEM model was 0.1.

The SEM model with comparison of subtype 1 and subtype 4 (Fig. 4a) showed that food preference was significantly associated with mental health ($\beta = 0.029$ and $P_{adj} = 0.009$) and brain MRI trait ($\beta = 0.041$ and $P_{adj} = 6.1 \times 10^{-5}$). Mental health was also a significant predictor for cognitive function ($\beta = -0.061$ and $P_{adj} = 0.03$). The RMSEA of this model was 0.1. In addition, the SEM model with comparison of subtype 2 and subtype 4 showed that food preference was significantly associated with mental health ($\beta = 0.084$ and $P_{adj} = 3.7 \times 10^{-12}$) and brain MRI trait ($\beta = 0.036$ and $P_{adj} = 0.002$). The brain MRI traits significantly predicted mental health ($\beta = -0.027$ and $P_{adj} = 0.03$). Both brain MRI trait and mental health significantly predicted cognitive function ($\beta = 0.094$ and $P_{adj} = 1.7 \times 10^{-8}$ and $\beta = -0.115$ and $P_{adj} = 1.5 \times 10^{-10}$, respectively). The RMSEA of this model was 0.05, which indicated a good fit (Fig. 4b).

All observed associations in these three path models were in the expected direction, with most paths being significant after FDR correction. The loadings of the latent variables in these SEM models can be found in Supplementary Table 14.

GWAS for four subtypes

To explore the genetic underpinnings of distinct subtypes of food-liking, we performed three case–control GWAS analysis, with subtype 4 as the reference group. As depicted in Fig. 5a, the GWAS-identified 1,266 single-nucleotide polymorphisms (SNPs) that were significantly different between subtype 3 and subtype 4 ($P < 5 \times 10^{-8}$). These SNPs were mostly located on chromosomes 2, 3, 13 and 17, such as rs36164224 (chromosome 2 (odds ratio (OR) of 1.07 and $P = 8.6 \times 10^{-10}$)), rs62250502 (chromosome 3 (OR of 0.93 and $P = 1.3 \times 10^{-11}$)), rs3124402 (chromosome 13 (OR of 0.92 and $P = 2.5 \times 10^{-11}$)) and rs2532387 (chromosome 17 (OR of 1.07 and $P = 1.2 \times 10^{-8}$)). Additionally, we found that two SNPs were significantly different between subtype 1 and subtype 4 ($P < 5 \times 10^{-8}$), namely rs2622068 and rs11939395, located on chromosome 4. Furthermore, no SNPs were observed to be significantly different between subtype 2 and subtype 4. The summarized GWAS results for subtype 3 versus subtype 4 were provided in Supplementary Data Table 1.

Distinct gene expression and enrichment across subtypes

To provide further biological insights into the GWAS results, the identified 1,266 SNPs differed between subtype 3 and subtype 4 ($P < 5 \times 10^{-8}$) were mapped to 16 genes using SNP2GENE function in FUMA. Gene expression analysis based on the Genotype-Tissue Expression (GTEx v8 54 tissue types) dataset revealed a cluster of genes, including *MAPT*, *MVB12B* and *NSF*, which exhibited high expression in several brain tissues, such as the anterior cingulate cortex (BA24), frontal cortex (BA9), amygdala and hippocampus and so on (Fig. 5b). Moreover, the *CADM2*, *CRHR1*, *MEIS1*, *PLEKHM1* and *KANSL1* genes also showed high expression in the cerebellar hemisphere and cerebellum (Fig. 5b).

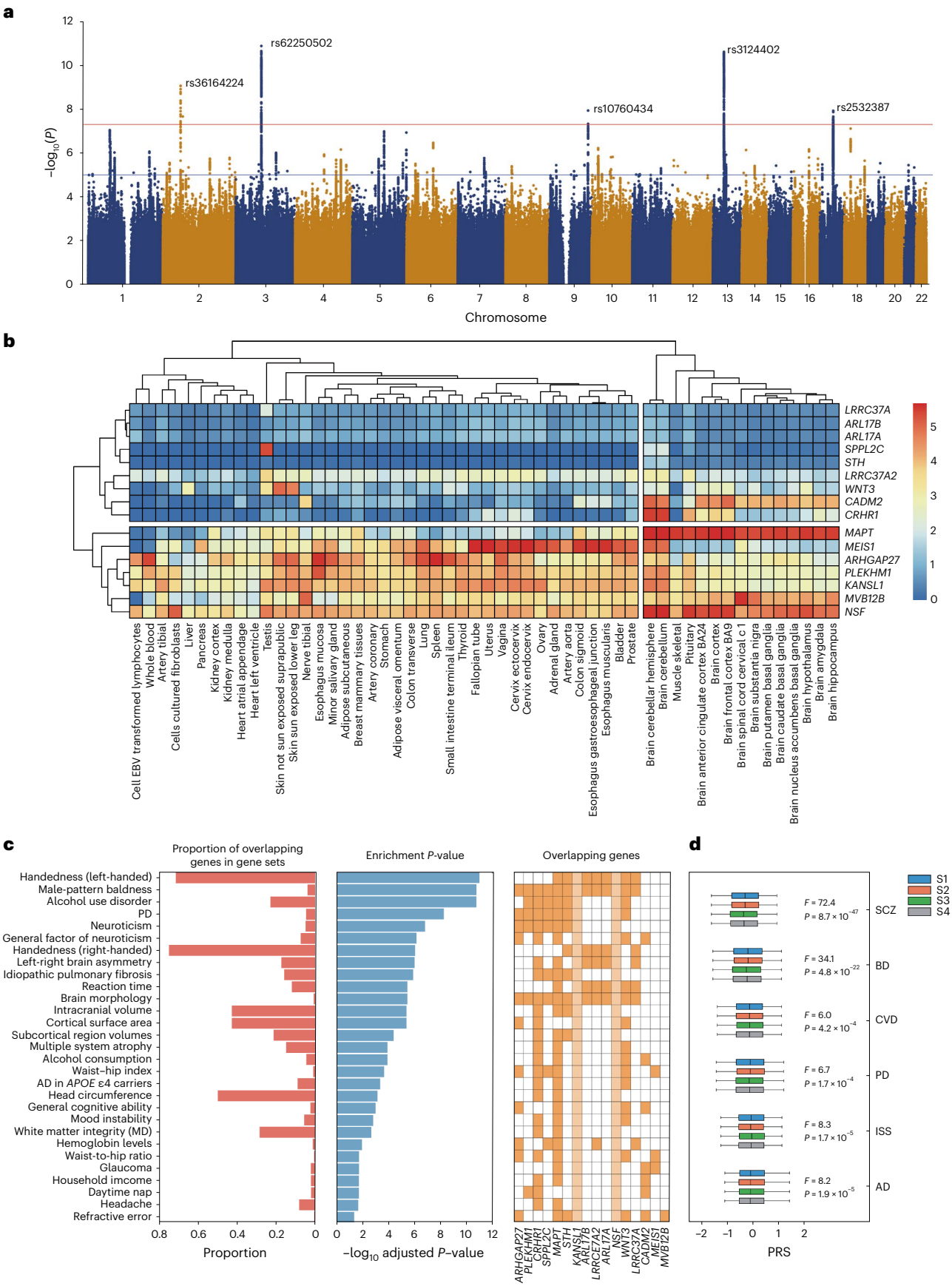
Furthermore, after Benjamini–Hochberg FDR corrections (adjusted *P* value <0.05), the identified 16 genes were found to converge on specific biological processes associated with mental health, cognitive functions and brain tissues, particularly within the context of the gene sets derived from the GWAS catalog (Fig. 5c). For instance, the *MAPT*, *STH*, *ARL17B*, *LRRC37A*, *LRRC37A2*, *ARL17A* and *WNT3* genes were most prominently enriched for handedness ($P_{adj} = 1.0 \times 10^{-11}$), whereas some genes were enriched for mental disorders, such as alcohol use disorder (*PLEKHM1*, *CRHR1*, *SPPL2C*, *MAPT*, *STH* and *WNT3*; $P_{adj} = 1.8 \times 10^{-11}$), Parkinson's disease (*ARHGAP27*, *PLEKHM1*, *CRHR1*, *SPPL2C*, *MAPT*, *STH* and *WNT3*; $P_{adj} = 5.5 \times 10^{-9}$) and Alzheimer's disease in *APOE* ε4 carriers (*CRHR1*, *MAPT* and *WNT3*; $P_{adj} = 4.3 \times 10^{-4}$). Additionally, some genes were enriched for cognitive functions, such as reaction time (*MAPT*, *ARL17B*, *LRRC37A*, *LRRC37A2*, *ARL17A* and *WNT3*; $P_{adj} = 3.5 \times 10^{-6}$). Moreover, some genes converged on brain tissues, such as brain morphology (*ARHGAP27, PLEKHM1, CRHR1, SPPL2C, MAPT, STH, ARL17B, LRRC37A, LRRC37A2, ARL17A* and *WNT3*; $P_{adj} = 3.5 \times 10^{-6}$), intracranial volume (*CRHR1, MAPT* and *STH*; $P_{adj} = 4.0 \times 10^{-6}$), cortical surface area (*ARHGAP27*, *CRHR1* and *WNT3*; $P_{adj} = 4.0 \times 10^{-6}$) and subcortical region volumes (*CRHR1*, *MAPT* and *STH*; $P_{adj} = 3.9 \times 10^{-5}$).

Discussion

In this study, we investigated naturally developed dietary patterns based on food-liking data from a large UK Biobank cohort ($n = 181{,}990$). Remarkably, this study represents a large-scale exploration of food preferences and their important implications for brain health. By employing data-driven approaches, we achieved a reliable and robust classification of dietary patterns. Our analyses identified four distinct dietary

Fig. 5 | GWAS-identified genetic variants, distinctive gene expression patterns and enriched functions between subtype 3 ('high protein and low fiber dietary pattern') and subtype 4 ('balanced dietary pattern'). **a**, Manhattan plot for the case–control GWAS analysis comparing subtypes 3 (cases, $n = 35{,}178$) and subtype 4 (controls, $n = 103{,}474$). Logistic regression analysis was performed, adjusting for sex, age, BMI, the top 10 ancestry principal components and genotype measurement batch. The red and blue horizontal lines indicate the conventional genome-wide significance thresholds of $P < 5 \times 10^{-8}$ and $P < 1 \times 10^{-5}$, respectively. **b**, Heatmap for gene expression analysis based on the GTEx (v8 54 tissue types) dataset. EBV, Epstein-Barr virus **c**, Associated biological functions from the GWAS catalog using the identified genes in GWAS. A multiple test correction was conducted using the Benjamini–Hochberg FDR with an adjusted *P* value cutoff of 0.05 and a minimum of two overlapped genes. **d**, The diverse PRSs for mental disorders ($n = 176{,}465$) and associated conditions across the four subtypes (subtype 1 (S1), subtype 2 (S2), subtype 3 (S3) and subtype 4 (S4)), as determined by ANCOVA analyses (*F*-tests) following Bonferroni correction ($\alpha = 0.05$). The analysis was adjusted for covariates, including age, BMI, education qualifications, Townsend deprivation index and PRS genetic principal components. The data are presented using a box plot (center line, median; box limits, upper and lower quartiles; and whiskers, 1.5 × interquartile range). SCZ, schizophrenia; BD, bipolar disease; CVD, cardiovascular disease; PD, Parkinson's disease; ISS, ischaemic stroke; AD, Alzheimer's disease.

https://doi.org/10.1038/s44220-024-00226-0



subtypes, each characterized by a unique dietary profile: subtype 1 ('starch-free or reduced-starch dietary pattern'), subtype 2 ('vegetarian dietary pattern'), subtype 3 ('high protein and low fiber dietary pattern') and subtype 4 ('balanced dietary pattern').

First, the current study provides a comprehensive understanding of the associations between data-driven dietary patterns and brain health, blood and metabolism and genetics. Our study has shed light on a coherent mediated pathway linking food preferences, brain MRI traits, cognition and mental health via structural equation modeling. A noteworthy finding of our study is the potential impact of food preferences on brain structure. We observed that individuals with specific food preferences displayed distinct patterns of brain MRI traits. These differential brain structural patterns may play an important role in shaping cognitive function and mental health outcomes[43,44]. The plasticity and adaptability of the brain, influenced by dietary choices, can lead to structural changes that influence cognitive functions and mental health[43,44]. Moreover, our results suggest a directional relationship between mental health and cognitive function. Mental health not only impacts cognitive abilities but is also influenced by brain structure. The intricate interplay between these factors underscores the importance of considering mental health as a crucial determinant in understanding brain health and cognitive performance[45].

Second, we revealed significant differences in mental health and cognitive function across four subtypes. Individuals in subtype 2, who consumed more vegetables and fruits, exhibited relatively higher levels of mental health scores, such as anxiety symptoms, depressive symptoms, mental distress, psychotic experience, self-harm and trauma and a relatively lower well-being score. The association between vegetarian (or vegan) diets and mental health in previous literature have been found to be controversial. Some investigations have reported positive associations of vegetarian and vegan diets with diverse mental health[35,36,46–51], while other studies found an inverse association[37,38,52,53] or no associations[39,40]. The conflicting findings can be attributed to differences in study designs (for example, cross-sectional, retrospective and randomized controlled trial), variations in how vegetarian and vegan diets were defined (with some of the studies including the consumption of fish or chicken also as vegetarian), discrepancies in the duration of adopting such diets, variations in the timing and methods used to assess mental health[52] and the unique characteristics of the groups studied (that is, biological sex). It should be noted that our observational study cannot draw a causal conclusion that vegetarianism leads to mental health problems. Particularly, our genetic analyses showed that individuals adopting the vegetarian dietary pattern exhibited higher PRSs in mental health, so it is possible that the worsened mental health conditions in subtype 2 may be indirectly influenced by the heightened genetic susceptibility. Further investigations are imperative in this regard to establish a causal conclusion in the future.

Subtype 3, which followed an unhealthy 'high protein and low fiber dietary pattern', had lower well-being scores than other subtypes. This finding was consistent with previous research that linked dietary quality with well-being[54,55] demonstrated that exposure to fast food images potentially impacting well-being[56]. In contrast, subtype 4, which followed a balanced and healthy dietary pattern, had less mental health problems and a higher well-being score than other subtypes, suggesting that a balanced intake of various food categories may be associated with better mental health[57,58]. Note that the Cox proportional hazards regression models further support the one-way ANCOVAs. Compared with subtype 4, subtype 3 had higher risks for anxiety, depression and stroke. Individuals, such as in subtype 3, who exhibit a higher intake of fatty meat may experience elevated stress levels and a higher risk of mental disorders, as reported in previous studies on the relationship between diet, stress and mental health[59]. Such effects could be attributed to an upsurge in the release of inflammatory factors and the permeation of gut flora through the intestinal wall, which is caused by high-fat foods. Our findings on blood and metabolic biomarkers revealed that higher levels of C-reactive protein and white blood cell count in subtype 3, compared with subtype 4, further support this point. These mechanisms may amplify the susceptibility to stress and depression by modifying signaling pathways leading to the brain[60,61]. This was also confirmed in previous studies which showed that an unbalanced diet may associate with a higher risk of mental disorders[15,62], and meat-eaters may have a higher risk for stroke[63].

Interestingly, the PRSs for various mental disorders mirrored the pattern. Subtype 2 displayed a heightened genetic susceptibility to a range of mental disorders, including Alzheimer's disease, Parkinson's disease, bipolar disorder, schizophrenia and suicide attempt, compared with other subtypes, while subtype 4 demonstrated relatively lower PRS risks for most mental disorders and related conditions. These results provide additional insights into the elevated risks of mental disorders for subtype 2 from a genetic standpoint. In other words, the higher mental health scores (for example, self-harm) as well as the lower cognitive performances scores observed in subtype 2 individuals might potentially be linked to their elevated genetic susceptibility to mental disorders (for example, PRSs for suicide attempts). In contrast, the relatively higher mental health symptoms in subtype 3 individuals, particularly as revealed in the Cox analysis, might be more strongly associated with their dietary habits rather than the genetic risks. Additionally, subtype 4 had the shortest reaction time, which may be attributed to their balanced dietary pattern[7].

Third, the associations between dietary patterns and brain morphology and white matter integrity are evident. Specifically, compared with subtype 4, subtype 3 had significantly lower GMV in 11 brain regions, including the postcentral gyrus, parahippocampal gyrus, inferior parietal gyrus, middle temporal gyrus, middle cingulate gyrus and so on. These findings were consistent with previous research that linked a 'healthier' diet (that is, rich in vegetables, vitamins, antioxidants and omega-3 polyunsaturated fatty acids or fish intake) with higher total GMV and the volumes of the hippocampus, cingulate gyrus, entorhinal cortex, temporal lobe and parietal lobe[28,29,64,65]. In contrast, diets high in saturated and trans fats and protein were associated with smaller GMV[66]. Our study also revealed significant differences in mean FA and MD of white matter tracts between subtype 3 and subtype 4. Higher MD and lower FA values are typically indicative of impaired fiber integrity due to increased diffusion and loss of coherence on preferred movement direction[67]. Subtype 4 showed higher FA than subtype 3 in seven brain regions, including the medial lemniscus, external capsule, uncinate fasciculus and so on. This findings further complemented the conclusions of previous studies, which suggested that a health-aware diet was associated with improved global white matter connectivity, as indicated by higher FA values[68]. Furthermore, compared with subtype 4, subtype 3 showed higher MD in 11 ROIs, including the external capsule, anterior limb of the internal capsule, hippocampal cingulum and so on. Overall, these regions were involved in emotional, motivational and cognitive and memory functions, as well as sensory and motor systems[69–71]. For example, the anterior limb of the internal capsule carries fibers from prefrontal cortical regions was associated with emotion, motivation, cognition and decision making[69,72], while the hippocampal cingulum is a major pathway connecting the cingulate gyrus to the hippocampal formation and is related to learning and memory functions[73]. In addition, subtype 2 had higher MD in the cerebral peduncle than subtype 4. The cerebral peduncle was associated with motor and sensory functions[74,75]. However, there is no conclusive evidence to establish the correlation between vegetarian diets and motor or sensory deficits.

Fourth, the blood and metabolic biomarkers examined in this study appear to be sensitive indicators of the impact of dietary patterns on the body. Our findings indicated that the four subtypes have significantly different levels of several biomarkers, such as omega-3 and omega-6 fatty acids[76,77], which are important components involved in serotonin synthesis. Additionally, subtype 4, characterized by a diet

that is generally considered to be healthier, exhibited higher levels of certain biomarkers than subtype 3, such as the degree of unsaturation in a fatty acid, HDL cholesterol, total lipids in HDL and so on. These results was consistent with previous investigations, which suggested that a balanced diet pattern was associated with higher HDL cholesterol levels in the elderly population[78]. Furthermore, in comparison with subtype 4, subtype 2 and subtype 3 both exhibited significantly lower levels of certain crucial fatty acids such as omega-3 fatty acids, which potentially be attributed to the lack of fish consumption in the latter subtype.

Finally, the case–control GWAS-identified 1,266 SNPs that differed between subtype 3 and subtype 4, and these were subsequently mapped to 16 genes. Our gene expression analysis pinpointed a cluster of genes, including notable candidates such as *MAPT*, *MVB12B* and *NSF*, which exhibited elevated expression across multiple brain tissues, encompassing regions such as the anterior cingulate cortex (BA24), frontal cortex (BA9), amygdala and hippocampus, among others. Intriguingly, many of these brain regions with potentially high gene expression overlap with the findings from our comparison of two subtypes based on GMV. This convergence supports the hypothesis that these genes play a crucial role in brain structure and may modulate the impact of dietary patterns on brain health[28,29,64,65]. Moreover, the identified genes were also found to be enriched in specific biological processes related to mental disorders and cognition, such as Parkinson's disease, Alzheimer's disease in *APOE* ε4 carriers and cognitive performance (reaction time). This further substantiates the potential link between dietary patterns and brain function and brain health[79,80]. Furthermore, we performed supplementary GWAS and PRS analyses to explore the predictive potential of genetics on brain MRI data and mental health within each dietary pattern. The GWAS analyses involved participants without MRI data, comparing other subtypes to subtype 4. Subsequently, PRSs were computed using genetic data from participants with MRI data, at *P* value thresholds of 0.01, 0.05 and 0.5. Similar analyses were conducted for mental health symptoms. However, after Bonferroni correction, no significant correlations were observed between brain MRI data and PRSs (Supplementary Fig. 5), or between mental health symptoms and PRSs (Supplementary Fig. 6). Integrating these supplementary analyses on PRSs related to dietary patterns with the evidence from GWAS, gene expression and functional enrichment analyses in our current study, we suggest a potential association between diet-related genes, brain function and mental health, but genes demonstrate a limited capacity to directly predict brain MRI data and mental health symptoms.

There are several highlights and substantial contributions of this study that are worth discussion. A key strength of the current research lies in its pioneering application of data-driven methods to analyze food preference data and identify naturally developed dietary patterns within a large-scale population. Previous studies have often utilized predefined dietary patterns, such as the Mediterranean or Western diet based on self-reported surveys[20,81–83]. However, the adopted definition and application of dietary patterns were not consistent across studies. In contrast, our study established a reliable classification system for dietary patterns by using a data-driven approach without prior assumptions and definitions. The identified dietary patterns reflect the usual eating habits in normal life, which can lead to meaningful investigations of their associations with brain structure and health. Another key strength of our study lies in the integration of multiple-dimensional data with a large sample size (including mental health measures, cognitive function, blood and metabolic biomarkers and genomics), which provides insights into the association between naturally developed dietary patterns by food preferences and brain health.

Our study has several implications for future research and clinical practice. First, our findings underscore the importance of considering dietary factors when examining brain structure, cognitive and mental health outcomes. Future studies could explore the mechanisms underlying these relationships, such as the potential role of specific nutrients or dietary patterns. Second, our study highlights the potential utility of using food preferences as a marker for identifying individuals at risk of cognitive impairment and mental health problems, which could be useful in developing targeted interventions and personalized dietary recommendations to promote brain health. The importance of our study lie in its pioneering exploration of food preferences and their profound impact on the brain, cognition, mental health and overall well-being. To the best of our knowledge, this is the large-scale investigation of its kind, representing a novel advancement in our understanding of the intricate relationship between diet and various aspects of human health. These findings also carry practical implications for educational practices. Early-age education in schools aimed at promoting healthy food preferences can play a vital role in fostering good brain health, cognition and overall well-being throughout the life of an individual. By nurturing healthy dietary habits from an early stage, we have the potential to positively impact public health and empower individuals to lead healthier and high-quality lives.

However, our study also has several limitations. First, it is important to note that the dietary patterns identified in the current study were based on data related to food-liking rather than actual food consumption. While our results have shown that food-liking measures were closely related to food consumption, subtle differences may exist and could influence the observed relationships. Second, participants included in this study are primarily healthy individuals from the UK Biobank. Given that the UK Biobank is known to have a 'healthy volunteer' selection bias[84], our results may not be entirely generalizable to other populations. Third, we observed demographic differences (for example, age, BMI, Townsend deprivation index and education qualifications) between respondents and nonrespondents to the food-liking questionnaire in the entire UK Biobank population. These disparities may arise from the substantial UKB sample, which may involve the selection criteria. Notably, despite the observed demographic disparities, our study demonstrates that this largest-scale food-liking dataset can effectively unveil robust and reliable food-liking subtypes. Fourth, omega-3 and omega-6 fatty acids[76,77], along with tryptophan[85], constitute key components in serotonin synthesis. While our findings revealed significant differences in omega-3 and omega-6 fatty acids among the four subtypes, a potential limitation is that we did not account for the actual levels of tryptophan, as well as a lack of detailed information on involvements of omega-3 and omega-6 fatty acids in these dietary patterns. Considering the pivotal role of serotonin in mood regulation and its substantial impact on overall mental health, future research should incorporate these aspects for a more comprehensive understanding. Finally, we used simplified measures to assess mental health factors, including well-being. Though brief measures are pragmatic for large-scale studies and have demonstrated reliability and validity in previous research[86–88], validation of our findings employing well-designed scales is needed in further investigations.

Our study highlights that the dietary patterns of elderly individuals may have significant associations with their mental health, cognitive functions, blood and metabolic biomarkers and brain imaging. A 'healthier' diet with balanced preferences in various food categories is associated with better mental health status, higher levels of cognitive functions and fewer risks of mental disorders. Our findings also indicate that there are genetic associations underlying these dietary patterns, implying that specific genes may be significant in regulating brain function and promoting mental health. Overall, our study provides systematic insights into understanding naturally developed dietary patterns in elderly individuals and underscores the associations between a balanced diet and brain health. The implications of these findings highlight the potential advantages of early-age education on diet, which could promote healthy food preferences and cultivate long-term brain health across the lifespan. Future research is needed to fully comprehend the potential long-term associations between

these dietary patterns and brain structure and health, particularly in adolescent and middle-aged populations.

Methods

Study population

This study used data from the UK Biobank under project 19542. The UK Biobank study was approved by the National Information Governance Board for Health and Social Care and the North West Multi-Centre Research Ethics Committee (ref. no. 11/NW/0382). All participants provided written informed consent. The risks of participants experiencing harm due to their involvement were minimal, and the UK Biobank is equipped with insurance to offer compensation for any instances of negligence resulting in harm during participation. The UK Biobank recruited more than 500,000 people aged 37–73 years from the United Kingdom between 2006 and 2010 (ref. 89). The data consisted of a wide range of phenotypic information and biological samples, including demographic characteristics, mental health, cognitive function, blood assays, multimodal neuroimaging and so on. In this study, we only included participants who completed the food-liking questionnaire and provided valid responses, resulting in a total of 181,990 participants (mean age 70.7 ± 7.7 years and 57.08% female).

Food-liking phenotypes

Food-liking data was gathered via an online questionnaire (https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/foodpref.pdf) from 182,176 participants. The questionnaire comprises 150 items that assessed both sensory attributes (for example, bitter and sweet) and food preferences (for example, fruit, vegetables and meat), as well as nonfood items, such as preferences for health-related activities (for example, physical activity and watching television). Liking is measured using a 9-point hedonic scale, where 1 represents 'extremely dislike' and 9 represents 'extremely like.' This widely used scale has good statistical properties, discrimination between different points and linearity[90]. The UK Biobank provided two response options, 'never tried' or 'do not wish to answer,' in addition to the 9-point scale. For our analyses, we included a subset of 140 items related to food and beverages, and classified these items into ten internally reliable categories, based on a classification system of a food preference questionnaire utilized in previous research[91], including: alcohol, beverages, dairy, flavorings, fruits, fish, meat, snacks, starches and vegetables. The use of these classification criteria serves to underscore our primary research objective, which is to explore the intricate association between different food categories and brain health. The individual items of each category are listed in Supplementary Table 1. We excluded 186 participants (0.1%) who responded 'never tried' or 'do not wish to answer' on more than 30% of the 140 food and beverage items, resulting in a final sample of 181,990 participants. Missing values (that is, 'never tried' or 'do not wish to answer') in the food-liking data were imputed using k-Nearest Neighbors[92] with the 'KNNImputer' function of the scikit-learn module in Python. The default settings were used, except for the number of neighboring samples (set as 7). Furthermore, to assess the robustness of our findings in the context of data imputation, we also utilized nonimputed data from 72,419 participants for the identification of food-liking subtypes. The demographic characteristics of the 181,990 participants, stratified by three age groups, are summarized in Table 1. In addition, we conducted a statistical analysis (*t*-test) to compare demographic characteristics between individuals who completed the food-liking questionnaire and those who did not within the entire UK Biobank population (Supplementary Table 2).

Identification of subtypes based on food-liking phenotypes

To identify data-driven food-liking subtypes, we first normalized the phenotypes using a *z*-score transformation. Next, to enhance comparability across food categories and reduce the dimensionality of the data[93], we performed a PCA for each food category. Specifically, the number of components was determined by adding the explained variance of each component until the total explained variance reached at least 80% (ref. 94). We then used the resulting principal components of the ten food categories as input for hierarchical clustering[95]. The hierarchical clustering used Euclidean distance and inner squared distance (minimum variance algorithm) for computing the distance between clusters. The clustering results were visualized using a dendrogram. Based on the dendrogram, we found that the population could be grouped into four distinct food-liking subtypes. To assess the stability of the variance explained by the obtained components and validate the reliability of the PCA results, we further examined the explained variance ratios at 70% and 90%.

Furthermore, to characterize the food preferences of the four subtypes, we calculated the average liking score for each food category across participants of each subtype. Given the variations in the score range across the four subtypes for different food categories, we normalized the liking score of each subtype to a range of 1 to 4 within each food category. To facilitate the comparison of food preferences among subtypes, we further standardized the liking score of each food category within each subtype by dividing it by the sum of the liking scores of all food categories, yielding a relative liking score for each food category within each subtype.

Comparisons between food-liking and food-consumption traits

To examine the relationship between food-liking measures and dietary habits, we adopted an approach that accounts for potentially corresponding relationships between the food frequency questionnaire (category 100052) and the food-liking questionnaire. Specifically, we selected food items that were common (for example, fruit, beef, lamb and so on) or those that had corresponding and similar items between

Table 1 | Demographic characteristics of the 181,990 UK Biobank participants, stratified by age groups

Characteristics	Ages 53–64 years (*n*=43,945, 24.15%)	Ages 65–76 years (*n*=90,724, 49.85%)	Ages 77–87 years (*n*=43,945, 26.00%)
Age, mean (standard deviation (s.d.))	60.0 (2.8)	71.1 (3.5)	79.8 (2.3)
Female	26,520 (60.35%)	52,666 (58.05%)	24,690 (52.18%)
BMI, mean (s.d.), $kg\,m^{-2}$	26.4 (4.8)	26.9 (4.6)	26.9 (4.2)
Townsend deprivation index, mean (s.d.), points	−1.2 (3.0)	−1.8 (2.8)	−2.0 (2.6)
Education qualifications			
College or University degree	21,987 (50.03%)	41,437 (45.67%)	17,967 (37.97%)
A levels and/or AS levels or equivalent	6,646 (15.12%)	12,448 (13.72%)	5,275 (11.15%)
O levels and/or GCSEs or equivalent	8,766 (19.95%)	17,116 (18.87%)	10,281 (21.73%)
CSEs or equivalent	2,782 (6.33%)	3,545 (3.91%)	703 (1.49%)
NVQ, HND, HNC or equivalent	1,684 (3.83%)	4,792 (5.28%)	2,789 (5.89%)
Other professional qualifications, for example, nursing or teaching	1,063 (2.42%)	4,549 (5.01%)	3,490 (7.38%)
None of the above	1,017 (2.31%)	6,837 (7.54%)	6,816 (14.40%)

both questionnaires (for example, chicken, cheese, bread and so on). To quantify this relationship, we calculated the average scores for both food-liking and food consumption associated with the selected food traits within each subtype. We visualized this quantitative relationship by generating a stacked bar plot to display these comparable scores of selective food items. This approach enabled us to gain a better understanding of the potential associations between food liking and food consumption.

Mental health assessment

The UK Biobank issued an online mental health self-assessment questionnaire (MHQ) in 2016. The questionnaire aimed to comprehensively evaluate self-reported symptoms of mental health and associated major environmental factors (https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=136). Note that the MHQ in the UK Biobank is a composite questionnaire, incorporating previously existing and validated measures, which is based, in part, on the World Health Organization's Composite International Diagnostic Interview—Short Form[96], alongside complementary tools that have been widely used in mental health research and have established validity and reliability[97,98]. The World Health Organization's Composite International Diagnostic Interview—Short Form forms the basis of many other major research studies, including those contributing to the work of the international Psychiatric Genomics Consortium. The self-reported diagnosed mental disorder rates in the MHQ align with population estimates from the Health Survey England[97]. For more details regarding the rationale and procedure for administration of the MHQ, refer to the UK Biobank website (http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=22). In this study, we analyzed eight mental health symptoms, including anxiety symptoms, depressive symptoms, mania symptoms, mental distress, psychotic experience, self-harm, trauma and well-being. The quantitative measures of these mental health symptoms were obtained by calculating an average score of the items used to assess each mental health symptom. The sample size of the data utilized in this study was from 118,616 participants. Specifically, the scores of items in one subcategory of the MHQ were firstly adjusted to the same direction, with higher values indicating more symptoms of mental disorder (with the exception of well-being, where a higher value indicated better well-being). Next, each item was normalized into a range of (0,1) using the MATLAB function 'mapminmax', and then the items within each category were averaged to generate an overall measure for each mental health symptom. The items used for the assessment of each mental health symptom were summarized in Supplementary Table 4.

Cognitive assessment

Several of the cognitive function tests administered via touchscreen during the initial assessment visit were reimplemented as web-based questionnaires, and the participants were invited to complete them remotely (https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=116). Six cognitive function tests were analyzed in the current study, including fluid intelligence, trail making, symbol–digit substitution, pairs matching, reaction time and numeric memory. These tests showed moderate correlations (ranging from 0.33 to 0.64, all $P < 0.001$) with their respective reference test(s) that was judged to be assessing the same cognitive capability or domain, suggesting substantial concurrent validity and test–retest reliability[99]. For instance, slower response on the reaction time test in the UK Biobank was associated with slower responses on Deary–Liewald reaction time test simple reaction time ($r = 0.52$ and $P < 0.001$). Additional information on the validity of cognitive tests can be found in Fawns-Ritchie et al.[99]. Additionally, the cognitive assessment data utilized in this analysis were from a substantial sample size of 179,740 participants. Summary information and sample size for these cognitive function tests are provided in Supplementary Table 5.

Blood and metabolic biomarkers

Blood biochemistry data (category 17518) was collected from ~480,000 participants during their recruitment visits between 2006 and 2010. The detailed procedures for quality control of blood biochemistry data can be found in Supplementary Information, as well as the open-source document (https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/biomarker_issues.pdf). Blood count data (category 100081) were also collected from the same number of participants during their first visit, using Beckman Coulter LH750 instruments to analyze samples collected in 4 ml EDTA vacutainers. Additional information about the hematology analysis is provided at (https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/haematology.pdf). We categorized the 30 blood biochemistry biomarkers into 'liver function', 'renal function', 'endocrine', 'immunometabolism' and 'bone and joint', while the 31 blood cell counts were classified into 'white blood cell', 'red blood cell' and 'platelet'.

The metabolic biomarkers were measured using a high-throughput nuclear magnetic resonance (NMR)-based metabolic biomarker profiling platform from randomly selected EDTA plasma samples collected during the first assessment, which included ~120,000 participants. Further details on the processing and quality control of NMR metabolic biomarkers in the UK Biobank were available in Supplementary Information and Julkunen et al.[100]. The NMR metabolomics (category 220) provided 249 metabolic biomarkers, of which 168 were directly measured and 81 were ratios of these. For this study, only the 168 directly measured metabolic biomarkers were used and categorized into 'amino acids', 'apolipoproteins', 'lipoprotein particle sizes', 'lipoprotein particle concentrations', 'fatty acids', 'triglycerides', 'phospholipids', 'cholesteryl esters', 'free cholesterol', 'cholesterol', 'other lipids', 'total lipids', 'ketone bodies', 'glycolysis-related metabolites', 'fluid balance' and 'inflammation.' The dataset for blood and metabolic biomarkers utilized in this study was from 42,665 participants. Supplementary Table 6 provides details of the category and sample size of these blood and metabolic biomarkers.

Brain MRI traits

The UK Biobank collected multimodal neuroimaging from ~40,000 participants using a standard Siemens Skyra 3T running VD13A SP4, with a standard Siemens 32-channel head coil. The details of the image acquisition are provided on the UK Biobank website in the form of a protocol (http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=2367). The UK Biobank conducted all the quality checking and data preprocessing procedures. The details of the acquisition protocols, image processing pipeline, image data files and imaging-derived phenotypes of brain structure and function are available on the UK Biobank website (http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=1977) and elsewhere[89,101]. We briefly describe the quality control steps for structural and diffusion MRI in Supplementary Information.

Structural MRI

This study utilized quality-controlled T1-weighted neuroimaging data obtained from structural MRI to investigate GMV in 32,715 participants. The T1 data were preprocessed with the Statistical Parametric Mapping software version 12 (https://www.fil.ion.ucl.ac.uk/spm/) using the CAT12 toolbox (https://neuro-jena.github.io/cat/) with default settings. The preprocessing involved high-dimensional spatial normalization with an integrated Dartel template in the Montreal Neurological Institute space, followed by nonlinear modulations and correction for the head size of each individual. Following these procedures, gray matter images (voxel size $1.5 \times 1.5 \times 1.5$ mm^3) were obtained for all participants. The AAL2 atlas with 94 cortical brain regions[102] was used to extract imaging-derived phenotypes referred to as atlas regional GMV. Intracranial volume was included as a covariate in the statistical analyses of GMV.

Diffusion MRI

The diffusion MRI data in the UK Biobank was obtained with two *b*-values (b = 1,000 and 2,000 s mm^{-2}) at a spatial resolution of 2 mm using a multiband acceleration factor of three, which allows for the acquisition of three slices simultaneously. For each diffusion-weighted shell, 50 distinct diffusion-encoding directions are obtained, resulting in a total of 100 directions across the two *b*-values. A standard (monopolar) Stejskal–Tanner pulse sequence is used for diffusion preparation, enabling a shorter echo time (TE of 92 ms) and higher signal-to-noise ratio compared with a twice-refocused (bipolar) sequence, although stronger eddy current distortions are introduced. The Eddy tool was used to correct for static field distortion, motion and eddy current distortions[103,104]. The diffusion MRI data were corrected for distortions, eddy currents and head motion and then modeled using FMRIB's Diffusion Toolbox for diffusion modeling and tractography analysis[105,106]. The neurite orientation dispersion and density imaging modeling were conducted using accelerated microstructure imaging via the Convex Optimization tool[107]. White matter pathways are aligned cross-subject for extracting image-derived phenotypes using tract-based spatial statistics[108,109], in which a high-dimensional warp maps a standard-space white matter skeleton to each participant, followed by defining ROIs as the intersection of the skeleton with standard-space masks for 48 tracts[110]. Definitions of tract regions and names can be found in the JHU ICBM-DTI-81 white-matter labels atlas described at http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases. The diffusion tensor imaging measures included FA, which reflects the directionality of diffusion, and MD, which measures overall diffusivity[111]. Statistical analyses were performed on the mean FA and MD of white matter tracts within the JHU ICBM-DTI-81 white-matter labels atlas (category 134) from 31,195 participants.

PRSs for mental disorders

The UK Biobank has released optimized PRSs for 28 diseases and 25 quantitative traits. The PRS scores were generated using a Bayesian approach applied to meta-analyzed summary statistics GWAS data. A principal component-based ancestry centering step was applied to center the score distributions around zero across all ancestries, and the score distributions were also standardized to have approximately unit variance within ancestry groups. More details about the initial PRS release are available at https://biobank.ndph.ox.ac.uk/showcase/refer.cgi?id=5202. For the current analyses, we used the standard PRS of several mental disorders, including Alzheimer's disease, bipolar disorder, Parkinson's disease and schizophrenia, as well as ischemic stroke and cardiovascular disease. The standard PRS set was calculated for all UK Biobank individuals, and the data were obtained entirely from external GWAS data. The dataset of PRSs used in this study was from 176,465 participants. The predictive performance of PRS for these six diseases is shown in Thompson et al.[112].

Additionally, we computed PRSs for depression and suicide attempt based on existing GWAS summary statistics[113,114]. To accomplish this, we utilized PRSice-2, in which details can be found at http://www.prsice.info, to estimate the PRS for each participant, after clumping the SNPs with an r^2 threshold of 0.1 and a physical distance of 250 kb, resulting in only the most strongly associated SNP being retained. The PRSs for each mental disorder were then calculated using a threshold of 0.05. The data for the PRSs of depression and suicide used in this study were both from 126,895 participants.

Statistical analyses

One-way ANCOVA and post hoc tests. To compare measures of interest among the four subtypes, we conducted one-way ANCOVA analyses[115] on mental health symptoms, cognitive functions, blood count and NMR metabolic biomarkers, brain MRI traits and PRSs of mental disorders. A Levene's test[116] was conducted to assess the equality of variances before one-way ANCOVAs. Additionally, post hoc tests[117] using two sample *t*-tests were performed to examine the differences between subtypes 1, 2, or 3 and subtype 4. To ensure the accuracy of our results, we included standard covariates of no interest, such as sex, age, BMI, education qualifications, Townsend deprivation index[118] and scanning sites (the last only applied to brain MRI traits). Additional covariates were also included to account for potential confounding factors, including the intracranial volume (regressed out in the analyses of GMV) and the PRS genetic principal components (regressed out in the analyses of PRSs of mental disorders). To correct for multiple comparisons, we used Bonferroni corrections in the analyses of mental health symptoms, cognitive functions, blood count and NMR metabolic biomarkers and PRSs of mental disorders. For the analyses of brain MRI traits, we used FDR corrections to correct for multiple comparisons.

Cox proportional hazards models. To assess the differences in survival rates of several common mental disorders among the four subtypes, Cox proportional hazards models were employed in this study, with subtype 4 serving as the reference group. The Cox model relies on the fundamental assumption of proportional hazards, which posits that the relative hazard remains constant over time across various levels of covariates[119]. To ensure this assumption, we employed the Schoenfeld residuals method[119], which tested the nonzero slope of each time-dependent covariate in the Cox model. The analyses were adjusted for sex, age, BMI, education qualifications and Townsend index[118]. The analyses included 11 mental disorders, including Alzheimer's disease (International Classification of Diseases (ICD)-10 F00 and G30), anxiety (ICD-10 F40 and F41), bipolar disorder (ICD-10 F31), depression (ICD-10 F32 and F33), dementia (ICD-10 F00, F01, F02, F03 and G30), eating disorder (ICD-10 F50), Parkinson's disease (ICD-10 G20), stroke (ICD-10 G45, G46, I60, I61, I63 and I64), sleep disorder (ICD-10 G47), migraine (ICD-10 G43) and schizophrenia (ICD-10 F20). The duration of follow-up, defined as the time elapsed from the participants' first occurrence of a mental disorder until their death, loss of follow-up or 19 July 2022 (whichever came first), was used as the timescale. The data were provided by 180,173 participants from the UK Biobank. The results of the models were presented as HRs and 95% CI, representing the averaged ratio of hazard of mental disorders between the other three subtypes compared with subtype 4 within 15 years of follow-up. The FDR corrections were used for multiple comparisons. Multivariate Cox regression analyses were performed using the 'survival' package in R[120].

SEM. A SEM[121] was employed to investigate the associations between food-liking and three latent variables: mental health, cognitive function, and brain MRI trait. We constructed three separate SEMs for each of the food-liking comparisons between the subtypes (subtypes 1, 2 or 3) and subtype 4. Food-liking was treated as a group variable indicating the other subtypes or subtype 4. The three latent variables were created by combining measurements that exhibited significant differences between other subtypes and subtype 4 in post hoc tests. Confirmatory factor analysis was used to estimate the latent variables in the model. The cognitive function latent variable was assessed on the basis of symbol–digit substitution, fluid intelligence, pair matching and reaction time. The mental health latent variable was evaluated using anxiety symptoms, depressive symptoms, mental distress, psychotic experiences, self-harm, trauma and well-being scores, which were obtained from the MHQ. The brain MRI trait latent variables in the three SEMs were constructed from the GMV of specific brain regions, the mean FA of the white matter tracts and the MD of the white matter tracts that showed significant differences between other subtypes and subtype 4. The RMSEA was used to assess model fitness. The analyses were conducted using the 'lavaan 0.6–14' package in R.

Case–control GWAS. Genotype data were obtained for all 500,000 participants from the UK Biobank v3 imputation. The comprehensive genotyping and quality-control procedures from the UK Biobank are

described in a previous publication[122]. We performed quality control for the genotype data from ~500,000 participants extracted from UKB v3 imputation, excluding SNPs with call rate <95%, minor allele frequency <0.1% and deviation from Hardy–Weinberg equilibrium ($P < 1\times10^{-10}$). We included only participants who were estimated to have recent British ancestry and no more than ten putative third-degree relatives in the analyses. After quality control, a total of 337,199 participants with 8,894,431 SNPs were included in our analysis. In this study, we utilized the genetic data from 181,551 participants.

To explore the genetic underpinnings of distinct subtypes of food-liking, we performed GWAS using logistic regression in PLINK 2.0 (refs. 123) (https://www.cog-genomics.org/plink/2.0/) on a binary phenotype distinguishing between other subtypes (subtype 1, $n = 32,843$; subtype 2, $n = 10,056$; and subtype 3, $n = 35,178$) and subtype 4 (controls, $n = 103,474$), while adjusting for sex, age, BMI, the top ten ancestry principal components and genotype measurement batch.

Gene expression and enrichment analysis. To provide further biological insights into the GWAS results, we utilized gene set enrichment analysis via FUMA[124]. First, we employed the SNP2GENE function in FUMA to map the SNPs with significant differences ($P < 5\times10^{-8}$) between subtype 3 and subtype 4 in the GWAS results to a set of prioritized genes based on positional, expression quantitative trait loci (eQTL) and chromatin interaction information of the SNPs. FUMA identifies independent, significant SNPs and their surrounding genomic loci based on LD structure and defines lead SNPs and genomic risk loci from the provided summary statistics. Next, we used the GENE2FUNC function to obtain information on gene expression and test for enrichment of the mapped genes from SNP2GENE in predefined pathways. The gene expression analysis was based on the GTEx (v8 54 tissue types) dataset[125] and provided averaged expression values ($\log_2$ transformed) per gene per label (for example, tissue types or developmental stage). For enrichment analysis on the mapped genes, hypergeometric tests were performed to determine if the mapped genes were overrepresented in any of the predefined gene sets. The FUMA platform provides access to three prominent gene sets for conducting enrichment analyses, namely the Molecular Signatures Database[126], WikiPathways[127] and the GWAS catalog[128]. A multiple test correction was conducted using the Benjamini–Hochberg FDR with an adjusted P value cutoff of 0.05 and a minimum of two overlapped genes.

Inclusion and ethics statement

This work involved a collaboration between scientists in China and the United Kingdom. All contributors have been listed as coauthors in acknowledgment to their work. This publication has considered the Global Code of Conduct. All research complies with the Declaration of Helsinki. The UK Biobank study was approved by the National Information Governance Board for Health and Social Care and the North West Multi-Centre Research Ethics Committee (ref. no. 11/NW/0382). All participants provided written informed consent.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Access to individual-level data from the UK Biobank, including phenotypic, neuroimaging and genotype information, is accessible to bona fide researchers via application through the UK Biobank website found at https://www.ukbiobank.ac.uk. Further details regarding registration for data access can be found at http://www.ukbiobank.ac.uk/register-apply/. The main dataset utilized in this study was obtained from the publicly accessible UK Biobank Resource under application number 19542. The AAL2 atlas utilized for GMV parcellation is publicly available via https://www.gin.cnrs.fr/en/tools/aal/. The summary statistics from previous GWAS studies on depression, which were utilized in this study for computing PRSs, are accessible via the Psychiatric Genomics Consortium, which can be downloaded from their website at https://pgc.unc.edu/for-researchers/download-results/. Additionally, the summary statistics from previous GWAS studies on suicide, employed in this study for PRS calculations, is accessible via application from the International Suicide Genetics Consortium at https://tinyurl.com/ISGC2021.

Code availability

MATLAB 2018b was mainly used to perform the analyses in this study. MATLAB functions 'pca', 'linkage' and 'anova1' were used to perform PCA, Hierarchical Clustering and one-way ANCOVA, respectively. For the analyses conducted in R (version 4.2.1), R package 'survival' (version 3.4.0) was used to perform a multivariate Cox Regression, and R package 'lavaan 0.6–14' was used to conduct Structural Equation Model analyses. PLINK 2.0, found at https://www.cog-genomics.org/plink/2.0/, and PRSice-2 version 2.3.5, found at http://www.prsice.info, were used to perform GWAS and calculate PRS, respectively. The primary code used in this study has been made publicly accessible through the GitHub repository found at https://github.com/RuohanZhang97/UKB_Diet (ref. 129).

References

1. Berridge, K. C., Robinson, T. E. & Aldridge, J. W. Dissecting components of reward: 'liking','wanting', and learning. *Curr. Opin. Pharmacol.* **9**, 65–73 (2009).
2. Recio-Román, A., Recio-Menéndez, M. & Román-González, M. V. Food reward and food choice. An inquiry through the liking and wanting model. *Nutrients* **12**, 639 (2020).
3. Bechthold, A. et al. Food groups and risk of coronary heart disease, stroke and heart failure: a systematic review and dose-response meta-analysis of prospective studies. *Crit. Rev. Food Sci. Nutr.* **59**, 1071–1090 (2019).
4. Schwingshackl, L. et al. Food groups and risk of type 2 diabetes mellitus: a systematic review and meta-analysis of prospective studies. *Eur. J. Epidemiol.* **32**, 363–375 (2017).
5. Castro-Barquero, S., Ruiz-León, A. M., Sierra-Pérez, M., Estruch, R. & Casas, R. Dietary strategies for metabolic syndrome: a comprehensive review. *Nutrients* **12**, 2983 (2020).
6. Steck, S. E. & Murphy, E. A. Dietary patterns and cancer risk. *Nat. Rev. Cancer* **20**, 125–138 (2020).
7. Muth, A.-K. & Park, S. Q. The impact of dietary macronutrient intake on cognitive function and the brain. *Clinical Nutrition* **40**, 3999–4010 (2021).
8. Jais, A. et al. Myeloid-cell-derived VEGF maintains brain glucose uptake and limits cognitive impairment in obesity. *Cell* **165**, 882–895 (2016).
9. Malaeb, S., Bakker, C., Chow, L. S. & Bantle, A. E. High-protein diets for treatment of type 2 diabetes mellitus: a systematic review. *Adv. Nutr.* **10**, 621–633 (2019).
10. Choi, S., DiSilvio, B., Fernstrom, M. H. & Fernstrom, J. D. Meal ingestion, amino acids and brain neurotransmitters: effects of dietary protein source on serotonin and catecholamine synthesis rates. *Physiol. Behav.* **98**, 156–162 (2009).
11. Ford, P. A., Jaceldo-Siegl, K., Lee, J. W., Youngberg, W. & Tonstad, S. Intake of Mediterranean foods associated with positive affect and low negative affect. *J. Psychosom. Res.* **74**, 142–148 (2013).
12. Ruusunen, A. et al. Dietary patterns are associated with the prevalence of elevated depressive symptoms and the risk of getting a hospital discharge diagnosis of depression in middle-aged or older Finnish men. *J. Affect. Disord.* **159**, 1–6 (2014).
13. Bear, T. L. K. et al. The role of the gut microbiota in dietary interventions for depression and anxiety. *Adv. Nutr.* **11**, 890–907 (2020).

14. van Zoonen, K. et al. Preventing the onset of major depressive disorder: a meta-analytic review of psychological interventions. *Int. J. Epidemiol.* **43**, 318–329 (2014).
15. Lassale, C. et al. Healthy dietary indices and risk of depressive outcomes: a systematic review and meta-analysis of observational studies. *Mol. Psychiatry* **24**, 965–986 (2019).
16. Firth, J., Gangwisch, J. E., Borisini, A., Wootton, R. E. & Mayer, E. A. Food and mood: how do diet and nutrition affect mental wellbeing? *Brit. Med. J.* **369**, m2382 (2020).
17. Román, G. C., Jackson, R. E., Gadhia, R., Román, A. N. & Reis, J. Mediterranean diet: the role of long-chain ω-3 fatty acids in fish; polyphenols in fruits, vegetables, cereals, coffee, tea, cacao and wine; probiotics and vitamins in prevention of stroke, age-related cognitive decline, and Alzheimer disease. *Revue Neurologique* **175**, 724–741 (2019).
18. Gangwisch, J. E. et al. High glycemic index and glycemic load diets as risk factors for insomnia: analyses from the Women's Health Initiative. *Am. J. Clin. Nutr.* **111**, 429–439 (2019).
19. Castro-Diehl, C. et al. Mediterranean diet pattern and sleep duration and insomnia symptoms in the Multi-Ethnic Study of Atherosclerosis. *Sleep* **41**, zsy158 (2018).
20. Więckowska-Gacek, A., Mietelska-Porowska, A., Wydrych, M. & Wojda, U. Western diet as a trigger of Alzheimer's disease: from metabolic syndrome and systemic inflammation to neuroinflammation and neurodegeneration. *Ageing Res. Rev.* **70**, 101397 (2021).
21. Marx, W., Moseley, G., Berk, M. & Jacka, F. Nutritional psychiatry: the present state of the evidence. *Proc. Nutr. Soc.* **76**, 427–436 (2017).
22. Bloch, M. H. & Hannestad, J. Omega-3 fatty acids for the treatment of depression: systematic review and meta-analysis. *Mol. Psychiatry* **17**, 1272–1282 (2012).
23. Schuch, F. B. et al. Physical activity and incident depression: a meta-analysis of prospective cohort studies. *Am. J. Psychiatry* **175**, 631–648 (2018).
24. Berding, K. et al. Diet and the microbiota–gut–brain axis: sowing the seeds of good mental health. *Adv. Nutr.* **12**, 1239–1285 (2021).
25. McGuinness, A. J. et al. A systematic review of gut microbiota composition in observational studies of major depressive disorder, bipolar disorder and schizophrenia. *Mol. Psychiatry* **27**, 1920–1935 (2022).
26. Berding, K. et al. Feed your microbes to deal with stress: a psychobiotic diet impacts microbial stability and perceived stress in a healthy adult population. *Mol. Psychiatry* **28**, 601–610 (2023).
27. Dinan, T. G. & Cryan, J. F. Gut microbiota: a missing link in psychiatry. *World Psychiatry* **19**, 111 (2020).
28. Boraxbekk, C. J. et al. Diet-induced weight loss alters functional brain responses during an episodic memory task. *Obes. Facts* **8**, 261–272 (2015).
29. Luciano, M. et al. Mediterranean-type diet and brain structural change from 73 to 76 years in a Scottish cohort. *Neurology* **88**, 449–455 (2017).
30. Prehn, K. et al. Caloric restriction in older adults—differential effects of weight loss and reduced weight on brain structure and function. *Cereb. Cortex* **27**, 1765–1778 (2017).
31. Mosconi, L. et al. Lifestyle and vascular risk effects on MRI-based biomarkers of Alzheimer's disease: a cross-sectional study of middle-aged adults from the broader New York City area. *BMJ Open* **8**, e019362 (2018).
32. *Series of Systematic Reviews on the Relationship between Dietary Patterns and Health Outcomes* (US Department of Agriculture, 2018).
33. Perng, W. et al. A prudent dietary pattern is inversely associated with liver fat content among multi-ethnic youth. *Pediatr. Obes.* **16**, e12758 (2021).
34. Hargreaves, S. M., Rosenfeld, D. L., Moreira, A. V. B. & Zandonadi, R. P. Plant-based and vegetarian diets: an overview and definition of these dietary patterns. *Eur. J. Nutr.* **62**, 1109–1121 (2023).
35. Kapoor, A., Baig, M., Tunio, S. A., Memon, A. S. & Karmani, H. Neuropsychiatric and neurological problems among vitamin B12 deficient young vegetarians. *Neurosci. J.* **22**, 228–232 (2017).
36. Paslakis, G. et al. Prevalence and psychopathology of vegetarians and vegans—results from a representative survey in Germany. *Sci. Rep.* **10**, 6840 (2020).
37. Beezhold, B. L. & Johnston, C. S. Restriction of meat, fish, and poultry in omnivores improves mood: a pilot randomized controlled trial. *Nutr. J* **11**, 1–5 (2012).
38. Rodríguez, J. & Gonzaléz, M. Indicators of anxiety and depression in subjects with different kinds of diet: vegetarians and omnivores. *Boletin de la Asociacion Medica de Puerto Rico* **90**, 58–68 (1998).
39. Pfeiler, T. M. & Egloff, B. Do vegetarians feel bad? Examining the association between eating vegetarian and subjective well-being in two representative samples. *Food Qual. Preference* **86**, 104018 (2020).
40. Bègue, L. & Shankland, R. Is vegetarianism related to anxiety and depression? A cross-sectional survey in a French sample. *J. Health Popul. Nutr.* **41**, 1–6 (2022).
41. Reedy, J., Subar, A. F., George, S. M. & Krebs-Smith, S. M. Extending methods in dietary patterns research. *Nutrients* **10**, 571 (2018).
42. Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987).
43. Joyce, E. M. Organic psychosis: the pathobiology and treatment of delusions. *CNS Neurosci. Ther.* **24**, 598–603 (2018).
44. Alexandros Lalousis, P. et al. Inflammatory subgroups of schizophrenia and their association with brain structure: a semi-supervised machine learning examination of heterogeneity. *Brain Behav. Immun.* **113**, 166–175 (2023).
45. Stein, D. J. et al. Global mental health and neuroscience: potential synergies. *Lancet Psychiatry* **2**, 178–185 (2015).
46. Agarwal, U. et al. A multicenter randomized controlled trial of a nutrition intervention program in a multiethnic adult population in the corporate setting reduces depression and anxiety and improves quality of life: the GEICO study. *Am. J. Health Promot.* **29**, 245–254 (2015).
47. Baines, S., Powers, J. & Brown, W. J. How does the health and well-being of young Australian vegetarian and semi-vegetarian women compare with non-vegetarians? *Public Health Nutr.* **10**, 436–442 (2007).
48. Lindeman, M. in *Health Psychology* (ed Whitaker, E. D.) 495–501 (Routledge, 2015).
49. Hibbeln, J. R., Northstone, K., Evans, J. & Golding, J. Vegetarian diets and depressive symptoms among men. *J. Affect. Disord.* **225**, 13–17 (2018).
50. Michalak, J., Zhang, X. C. & Jacobi, F. Vegetarian diet and mental disorders: results from a representative community survey. *Int J. Behav. Nutr. Phys. Act.* **9**, 1–10 (2012).
51. Matta, J. et al. Depressive symptoms and vegetarian diets: results from the constances cohort. *Nutrients* **10**, 1695 (2018).
52. Iguacel, I., Huybrechts, I., Moreno, L. A. & Michels, N. Vegetarianism and veganism compared with mental health and cognitive outcomes: a systematic review and meta-analysis. *Nutr. Rev.* **79**, 361–381 (2021).
53. Beezhold, B., Radnitz, C., Rinne, A. & DiMatteo, J. Vegans report less stress and anxiety than omnivores. *Nutr. Neurosci.* **18**, 289–296 (2015).

54. Phillips, C. M., Shivappa, N., Hébert, J. R. & Perry, I. J. Dietary inflammatory index and mental health: a cross-sectional analysis of the relationship with depressive symptoms, anxiety and well-being in adults. *Clin. Nutr.* **37**, 1485–1491 (2018).
55. Lai, J. S. et al. A systematic review and meta-analysis of dietary patterns and depression in community-dwelling adults. *Am. J. Clin. Nutr.* **99**, 181–197 (2014).
56. House, J., DeVoe, S. E. & Zhong, C.-B. Too impatient to smell the roses: exposure to fast food impedes happiness. *Soc. Psychol. Pers. Sci.* **5**, 534–541 (2014).
57. Nakamura, M. et al. Low zinc, copper, and manganese intake is associated with depression and anxiety symptoms in the japanese working population: findings from the eating habit and well-being study. *Nutrients* **11**, 847 (2019).
58. Holder, M. D. The contribution of food consumption to well-being. *Ann. Nutr. Metab.* **74**, 44–52 (2019).
59. Bremner, J. D. et al. Diet, stress and mental health. *Nutrients* **12**, 2428 (2020).
60. Hamilton, M. K., Boudry, G., Lemay, D. G. & Raybould, H. E. Changes in intestinal barrier function and gut microbiota in high-fat diet-fed rats are dynamic and region dependent. *Am. J. Physiol. Gastrointest. Liver Physiol.* **308**, G840–G851 (2015).
61. Dinan, T. G. & Cryan, J. F. Melancholic microbes: a link between gut microbiota and depression? *Neurogastroenterol. Motility* **25**, 713–719 (2013).
62. Kris-Etherton, P. M. et al. Nutrition and behavioral health disorders: depression and anxiety. *Nutr. Rev.* **79**, 247–260 (2021).
63. Petermann-Rocha, F. et al. Vegetarians, fish, poultry, and meat-eaters: who has higher risk of cardiovascular disease incidence and mortality? A prospective study from UK Biobank. *Eur. Heart J.* **42**, 1136–1143 (2021).
64. Jensen, D. E. A., Leoni, V., Klein-Flügge, M. C., Ebmeier, K. P. & Suri, S. Associations of dietary markers with brain volume and connectivity: A systematic review of MRI studies. *Ageing Res. Rev.* **70**, 101360 (2021).
65. Gu, Y. et al. Mediterranean diet and brain structure in a multiethnic elderly cohort. *Neurology* **85**, 1744–1751 (2015).
66. Jacka, F. N., Cherbuin, N., Anstey, K. J., Sachdev, P. & Butterworth, P. Western diet is associated with a smaller hippocampus: a longitudinal investigation. *BMC Med.* **13**, 215 (2015).
67. Soares, J. M., Marques, P., Alves, V. & Sousa, N. A hitchhiker's guide to diffusion tensor imaging. *Front. Neurosci.* **7**, 31 (2013).
68. Booth, T. et al. Personality, health, and brain integrity: the Lothian birth cohort study 1936. *Health Psychol.* **33**, 1477 (2014).
69. Safadi, Z. et al. Functional segmentation of the anterior limb of the internal capsule: linking white matter abnormalities to specific connections. *J. Neurosci.* **38**, 2106–2117 (2018).
70. Bubb, E. J., Metzler-Baddeley, C. & Aggleton, J. P. The cingulum bundle: anatomy, function, and dysfunction. *Neurosci. Biobehav. Rev.* **92**, 104–127 (2018).
71. Kamali, A., Kramer, L. A., Butler, I. J. & Hasan, K. M. Diffusion tensor tractography of the somatosensory system in the human brainstem: initial findings using high isotropic spatial resolution at 3.0 T. *Eur. Radiol.* **19**, 1480–1488 (2009).
72. Haber, S. N. & Behrens, T. E. The neural network underlying incentive-based learning: implications for interpreting circuit disruptions in psychiatric disorders. *Neuron* **83**, 1019–1039 (2014).
73. Dalboni da Rocha, J. L., Bramati, I., Coutinho, G., Tovar Moll, F. & Sitaram, R. Fractional anisotropy changes in parahippocampal cingulum due to Alzheimer's disease. *Sci. Rep.* **10**, 2660 (2020).
74. Armentano, M. et al. COUP-TFI regulates the balance of cortical patterning between frontal/motor and sensory areas. *Nat. Neurosci.* **10**, 1277–1286 (2007).
75. Shinoura, N. et al. Fibers connecting the primary motor and sensory areas play a role in grasp stability of the hand. *Neuroimage* **25**, 936–941 (2005).
76. Yehuda, S. Omega-6/omega-3 ratio and brain-related functions. *World Rev. Nutr. Diet.* **92**, 37–56 (2003).
77. Thesing, C. S., Bot, M., Milaneschi, Y., Giltay, E. J. & Penninx, B. W. Omega-3 and omega-6 fatty acid levels in depressive and anxiety disorders. *Psychoneuroendocrinology* **87**, 53–62 (2018).
78. Song, P. et al. Association between dietary patterns and low HDL-C among community-dwelling elders in North China. *Nutrients* **13**, 3308 (2021).
79. Nagpal, R., Neth, B. J., Wang, S., Craft, S. & Yadav, H. Modified Mediterranean-ketogenic diet modulates gut microbiome and short-chain fatty acids in association with Alzheimer's disease markers in subjects with mild cognitive impairment. *EBioMedicine* **47**, 529–542 (2019).
80. Hill, E., Goodwill, A. M., Gorelik, A. & Szoeke, C. Diet and biomarkers of Alzheimer's disease: a systematic review and meta-analysis. *Neurobiol. Aging* **76**, 45–52 (2019).
81. Lassale, C. et al. Healthy dietary indices and risk of depressive outcomes: asystematic review and meta-analysis of observational studies. *Mol. Psychiatry* **24**, 965–986 (2019).
82. Marx, W. et al. Diet and depression: exploring the biological mechanisms of action. *Mol. Psychiatry* **26**, 134–150 (2021).
83. McEvoy, C. T. et al. Dietary patterns during adulthood and cognitive performance in midlife: The CARDIA study. *Neurology* **92**, e1589–e1599 (2019).
84. Fry, A. et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am. J. Epidemiol.* **186**, 1026–1034 (2017).
85. Kałużna-Czaplińska, J., Gątarek, P., Chirumbolo, S., Chartrand, M. S. & Bjørklund, G. How important is tryptophan in human health? *Crit. Rev. Food Sci. Nutr.* **59**, 72–88 (2019).
86. Cheung, F. & Lucas, R. E. Assessing the validity of single-item life satisfaction measures: results from three large samples. *Qual. Life Res.* **23**, 2809–2818 (2014).
87. Diener, E., Emmons, R. A., Larsen, R. J. & Griffin, S. The satisfaction with life scale. *J. Pers. Assess.* **49**, 71–75 (1985).
88. Zhu, X. et al. Multidimensional assessment of subjective well-being and risk of dementia: findings from the UK Biobank Study. *J. Happiness Stud.* **24**, 629–650 (2023).
89. Miller, K. L. et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* **19**, 1523–1536 (2016).
90. Peryam, D. R. & Pilgrim, F. J. Hedonic scale method of measuring food preferences. *Food Technol.* **11**, 9–14 (1957).
91. Fildes, A. et al. Nature and nurture in children's food preferences. *Am. J. Clin. Nutr.* **99**, 911–917 (2014).
92. Troyanskaya, O. et al. Missing value estimation methods for DNA microarrays. *Bioinformatics* **17**, 520–525 (2001).
93. Mwangi, B., Tian, T. S. & Soares, J. C. A review of feature reduction techniques in neuroimaging. *Neuroinformatics* **12**, 229–244 (2014).
94. Jolliffe, I. T. *Principal Component Analysis* (Springer, 2002).
95. Nielsen, F. *Introduction to HPC with MPI for Data Science* (Springer, 2016).
96. Kessler, R. C., Andrews, G., Mroczek, D., Ustun, B. & Wittchen, H. U. The World Health Organization composite international diagnostic interview short-form (CIDI-SF). *Int. J. Meth. Psych. Res.* **7**, 171–185 (1998).
97. Lee, W. et al. Mental health in UK Biobank-development, implementation and results from an online questionnaire completed by 157,366 participants. *Bjpsych. Open* **4**, 83–90 (2018).

98. Davis, K. & Hotopf, M. Mental health phenotyping in UK Biobank. *Prog. Neurol. Psychiatry* **23**, 4–7 (2019).
99. Fawns-Ritchie, C. & Deary, I. J. Reliability and validity of the UK Biobank cognitive tests. *PLoS ONE* **15**, e0231627 (2020).
100. Julkunen, H. et al. Atlas of plasma nuclear magnetic resonance biomarkers for health and disease in 118,461 individuals from the UK Biobank. *Nat. Commun.* https://doi.org/10.1038/s41467-023-36231-7 (2023).
101. Alfaro-Almagro, F. et al. Image processing and quality control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage* **166**, 400–424 (2018).
102. Rolls, E. T., Joliot, M. & Tzourio-Mazoyer, N. Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *NeuroImage* **122**, 1–5 (2015).
103. Andersson, J. L. & Sotiropoulos, S. N. Non-parametric representation and prediction of single- and multi-shell diffusion-weighted MRI data using Gaussian processes. *NeuroImage* **122**, 166–176 (2015).
104. Andersson, J. L. R. & Sotiropoulos, S. N. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage* **125**, 1063–1078 (2016).
105. Douaud, G. et al. DTI measures in crossing-fibre areas: increased diffusion anisotropy reveals early white matter alteration in MCI and mild Alzheimer's disease. *NeuroImage* **55**, 880–890 (2011).
106. Ennis, D. B. & Kindlmann, G. Orthogonal tensor invariants and the analysis of diffusion tensor magnetic resonance images. *Magn. Reson. Med.* **55**, 136–146 (2006).
107. Raichle, M. E. et al. A default mode of brain function. *Proc. Natl Acad. Sci. USA* **98**, 676–682 (2001).
108. de Groot, M. et al. Improving alignment in Tract-based spatial statistics: evaluation and optimization of image registration. *Neuroimage* **76**, 400–411 (2013).
109. Smith, S. M. et al. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *NeuroImage* **31**, 1487–1505 (2006).
110. Wakana, S., Jiang, H., Nagae-Poetscher, L. M., van Zijl, P. C. & Mori, S. Fiber tract-based atlas of human white matter anatomy. *Radiology* **230**, 77–87 (2004).
111. Wartolowska, K. A. & Webb, A. J. S. Blood pressure determinants of cerebral white matter hyperintensities and microstructural injury: UK Biobank cohort study. *Hypertension* **78**, 532–539 (2021).
112. Thompson, D. J. et al. UK Biobank release and systematic evaluation of optimised polygenic risk scores for 53 diseases and quantitative traits. Preprint at *medRxiv* https://doi.org/10.1101/2022.06.16.22276246 (2022).
113. Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
114. Mullins, N. et al. Dissecting the shared genetic architecture of suicide attempt, psychiatric disorders, and known risk factors. *Biol. Psychiatry* **91**, 313–327 (2022).
115. Wickens, T. D. & Keppel, G. *Design and Analysis: A Researcher's Handbook* (Pearson Prentice-Hall Upper Saddle River, 2004).
116. Olkin, I. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling* (Stanford Univ. Press, 1960).
117. Hsu, J. *Multiple Comparisons: Theory and Methods* (CRC Press, 1996).
118. Townsend, P., Phillimore, P. & Beattie, A. *Health and Deprivation: Inequality and the North* Vol. 8, 1st edn (Taylor & Francis, 2023).
119. Schoenfeld, D. Partial residuals for the proportional hazards regression model. *Biometrika* **69**, 239–241 (1982).
120. Therneau, T. A package for survival analysis in S. R Package version 2 (2015).
121. Bollen, K. A. *Structural Equations with Latent Variables* Vol. 210 (Wiley, 1989).
122. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
123. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, s13742–015 (2015).
124. Watanabe, K., Taskesen, E., Van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
125. Aguet, F. et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
126. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
127. Kutmon, M. et al. WikiPathways: capturing the full diversity of pathway knowledge. *Nucleic Acids Res.* **44**, D488–D494 (2016).
128. MacArthur, J. et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).
129. RuohanZhang97. UKB_Diet. *GitHub* https://github.com/RuohanZhang97/UKB_Diet (2024).

Acknowledgements

W.C. was supported by grants from the Shanghai Rising-Star Program (21QA1408700), the National Natural Sciences Foundation of China (82071997) and the National Key Research and Development Program of China (grant no. 2023YFC3605400). J.F. was supported by the National Key R&D Program of China (2018YFC1312900 and 2019YFA0709502), the Shanghai Municipal Science and Technology Major Project (2018SHZDZX01), the ZJ Lab, Shanghai Center for Brain Science and Brain-Inspired Technology and the 111 Project (B18015). The funders played no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Authors Contributions

All authors had full access to the data in the study and accepted the responsibility to submit it for publication. W.C. and J.F. conceived and designed the study. R.Z. performed the data analysis. Z.L., W.Z., Y.Z., Y.L. and C.S. preprocessed the data. W.C., J.F. and B.J.S. contributed to interpretation of the results. R.Z. wrote the manuscript. B.J.S., R.Z., B.Z., C.S. and W.C. revised the manuscript. R.Z. contributed to the visualization. All authors considered how to analyze data and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44220-024-00226-0.

Correspondence and requests for materials should be addressed to Jianfeng Feng or Wei Cheng.

Peer review information *Nature Mental Health* thanks Kostas Anastasiou, Oleg Medvedev and Marie Pierre Moisan for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Jianfeng Feng
Wei Cheng

Last updated by author(s): Feb 25, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data collection did not involve the use of any software. All the data used was obtained directly from the UK Biobank, as elaborated in detail in the paper.
Data analysis	MATLAB 2018b was mainly used to perform the analyses in this study: MATLAB functions "pca", "linkage", "anova1" were used to perform Principal Component Analysis (PCA), Hierarchical Clustering, and One-Way Analysis of Covariance (ANCOVA), respectively. R version 4.2.1 package: R package "survival" (version 3.4.0) was used to perform Multivariate Cox Regression. R package "lavaan 0.6-14" was used to conduct Structural Equation Model analyses. PLINK 2.0 alpha (https://www.cog-genomics.org/plink/2.0/) and PRSice-2 version 2.3.5 (http://www.prsice.info) were used to perform genome-wide association analyses and calculate PRS, respectively. The primary code used in this study has been made publicly accessible through the GitHub repository (https://github.com/RuohanZhang97/UKB_Diet).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Access to individual-level data from the UK Biobank, including phenotypic, neuroimaging, and genotype information, is accessible to bona fide researchers via application through the UK Biobank website (https://www.ukbiobank.ac.uk). Further details regarding registration for data access can be found at http://www.ukbiobank.ac.uk/register-apply/. The main dataset utilized in this study was obtained from the publicly accessible UK Biobank Resource under application number 19542. The AAL2 atlas utilized for gray matter volume parcellation is publicly available via https://www.gin.cnrs.fr/en/tools/aal/. Summary statistics from previous GWAS studies on depression, which were utilized in this study for computing polygenic risk scores (PRSs), are accessible via the Psychiatric Genomics Consortium (PGC), which can be downloaded from their website at https://pgc.unc.edu/for-researchers/download-results/. Additionally, the summary statistics from previous GWAS studies on suicide, employed in this study for PRS calculations, is accessible via application from the International Suicide Genetics Consortium at https://tinyurl.com/ISGC2021.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	The study included both male and female participants from the UK biobank study. Summary statistics on sex distributions were reported in Tables 1 and S3. All statistic models were adjusted for sex.
Reporting on race, ethnicity, or other socially relevant groupings	Within the UK Biobank cohort, 94.6% of participants identified as belonging to the white ethnicity group. Therefore, we did not incorporate race as a covariate in our statistical analyses. Among the 181,990 UK Biobank participants included in this study, the majority held either college or university degrees (n = 81,391, 44.72%), followed by O levels/GCSEs or equivalent qualifications (n = 36,163, 19.87%).
Population characteristics	We analyzed a dataset consisting of 181,990 participants (with a mean age of 70.7 ± 7.7 years, of which 57.08% were females) to identify distinct food-liking subtypes. The demographic characteristics of these participants were summarized in Table 1. Our results revealed the presence of four distinct food-liking subtypes, with proportions of 18.09%, 5.54%, 19.39%, and 56.98% for Subtypes 1 to 4, respectively. Further demographic details for these subtypes can be found in Table S3. The demographic characteristics focused in this study encompassed sex, age, BMI, education qualifications, and Townsend index, all of which were used as covariates in the statistical analyses. For comprehensive statistical insights, Tables S4 to S7 in the Supplementary Materials present detailed information on the sample sizes employed in ANCOVA analyses and post hoc tests on mental health symptoms, cognitive function assessments, as well as blood and metabolic biomarkers.
Recruitment	The UK Biobank is a prospective, population-based cohort that recruited more than 500,000 participants aged 37 - 73 years who attended 1 of 22 assessment centers across the United Kingdom between 2006 and 2010. Previous investigation showed UK biobank subject to a healthy sample bias.
Ethics oversight	The UKB has received approval from the National Information Governance Board for Health and Social Care and the National Health Service North West Centre for Research Ethics Committee (Ref: 11/NW/0382). All participants provided informed consent through electronic signature at the baseline assessment. Additionally, the risks of participants experiencing harm due to their involvement are minimal, and UK Biobank is equipped with insurance to offer compensation for any instances of negligence resulting in harm during participation.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were not predetermined using statistical methods. We utilized the entire available dataset from the UK Biobank, comprising 181,990 participants with Food-liking data, 42,665 participants with blood and metabolic biomarker data, 32,715 participants with T1-weighted structural MRI, 31,195 participants with diffusion MRI, and 181,551 participants with genotype data. Detailed sample size information can be found in the Methods section and Supplementary Materials.
Data exclusions	In our analyses of food liking data, we included only 140 of the 150 items related to food and beverages. Additionally, out of the 182,176

Data exclusions	participants, only 186 participants (0.1%) who responded “never tried” or “do not wish to answer” on more than 30% of the 140 food and beverage items were excluded, resulting in a final sample of 181,990 participants (as detailed in the Methods section). In the Cox proportional hazard regression model, we excluded participants based on several criteria, including patients with missing data regarding food-liking preferences, missing baseline information on the analysed mental disorders (as detailed in the Methods section), unavailability of baseline data, missing follow-up data, and individuals who self-reported these mental disorders. In the structural equation model (SEM), we excluded participants with any missing data related to mental health, cognitive function, or brain MRI traits. For the genome-wide association analyses on food-liking, we excluded single-nucleotide polymorphisms (SNPs) with call rate < 95%, minor allele frequency < 0.1%, and deviation from Hardy-Weinberg equilibrium (P < 1×10−10), and included only participants who were estimated to have recent British ancestry and no more than ten putative third-degree relatives in the analyses.
Replication	All available data were used to maximize statistical power of the analysis therefore we did not repeat the analysis.
Randomization	To ensure the accuracy of our results, we included standard covariates of no interest, such as sex, age, BMI, education qualifications, Townsend deprivation index, and scanning sites (the last only applied to brain MRI traits). Additional covariates were also included to account for potential confounding factors, including the intracranial volume (regressed out in the analyses of T1-weighted structural imaging), and the PRS genetic principal components (regressed out in the analyses of polygenic risk scores of mental disorders). The genome-wide association analysis (GWAS) was adjusted for gender, age, BMI, the top 10 ancestry principal components and genotype measurement batch.
Blinding	Blinding was not applicable to this study as this study is observational.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Magnetic resonance imaging

Experimental design

Design type	Structural MRI and diffusion MRI
Design specifications	The UK Biobank designed the imaging acquisition protocols including 6 modalities, covering structural, diffusion and functional imaging. The collection order is T1-weighted structural image, resting-state functional MRI, task functional MRI, T2-weighted FLAIR structural image, Diffusion MRI and susceptibility-weighted imaging. The T1-weighted structural image was acquired using straight sagittal orientation for 5 minutes. The diffusion MRI data was acquired for 7 minutes (including 36 seconds phase-encoding reversed data).
Behavioral performance measures	Food-liking data was gathered via an online questionnaire (https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/foodpref.pdf) from 182,176 participants. The questionnaire comprises 150 items that assessed both sensory attributes (e.g., bitter, sweet) and food preferences (e.g., fruit, vegetables, meat). In addition, non-food items are included to measure liking for health-related activities (e.g., physical activity, watching television). Liking is measured using a 9-point hedonic scale, where 1 represents "Extremely dislike" and 9 represents "Extremely like." This widely-used scale has good statistical properties, discrimination between different points, and linearity. The UK Biobank provided two response options, "never tried" or "do not wish to answer," in addition to the 9-point scale. For our analyses, we included only the 140 items related to food and beverages, and classified these items into 10 internally reliable categories based on the Food Preference Questionnaire utilized in previous research: alcohol, beverages, dairy, flavourings, fruits, fish, meat, snacks, starches, and vegetables. The use of these classification criteria serves to underscore our primary research objective, which is to explore the intricate association between different food categories and brain health. The individual items of each category were listed in Table S1. Out of the 182,176 participants, only 186 participants (0.1%) who responded “never tried” or “do not wish to answer” on more than 30% of the 140 food and beverage items were excluded, resulting in a final sample of 181,990 participants. The UK Biobank issued an online mental health self-assessment questionnaire (MHQ) in 2016. The questionnaire aimed to comprehensively evaluate self-reported symptoms of mental health and associated major environmental factors (https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=136). Note that, the MHQ in the UK Biobank is a composite questionnaire, incorporating previously existing and validated measures, which is based, in part, on the World Health Organization’s Composite International Diagnostic Interview – Short Form (CIDI-SF), alongside complementary tools that

have been widely used in mental health research and have established validity and reliability. The CIDI-SF forms the basis of many other major research studies, including those contributing to the work of the international Psychiatric Genomics Consortium. The self-reported diagnosed mental disorder rates in the MHQ align with population estimates from the Health Survey England. For more details regarding the rationale and procedure for administration of the MHQ, refer to the UK Biobank website (http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=22). In this study, we analyzed eight mental health symptoms, including anxiety symptoms, depressive symptoms, mania symptoms, mental distress, psychotic experience, self-harm, trauma and well-being. Quantitative measures of these mental health symptoms were obtained by calculating an average score of the items used to assess each mental health symptom. The sample size of the data utilized in this study was from 118,616 participants.

Several of the cognitive function tests administered via touchscreen during the initial assessment visit were re-implemented as web-based questionnaires and participants were invited to complete them remotely (https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=116). Six cognitive function tests were analysed in the current study, including fluid intelligence, trail making, symbol-digit substitution, pairs matching, reaction time and numeric memory. These tests showed moderate correlations (ranging from 0.33 to 0.64, all $p < 0.001$) with their respective reference test(s) that were judged to be assessing the same cognitive capability or domain, suggesting substantial concurrent validity and test-retest reliability. For instance, slower response on the reaction time test in the UK Biobank was associated with slower responses on Deary-Liewald Reaction Time Test (DLRT) Simple Reaction Time ($r = 0.52$, $p < 0.001$). Additional information on the validity of cognitive tests can be found in Fawns-Ritchie et al. (2020). Additionally, the cognitive assessment data utilized in this analysis was from a substantial sample size of 179,740 participants.

Acquisition

Imaging type(s): T1-weighted structural imaging

Field strength: 3T

Sequence & imaging parameters: The EPI-based acquisitions utilize simultaneous multi-slice (multiband) acceleration. Biobank uses pulse sequences and reconstruction code from the Center for Magnetic Resonance Research (CMRR), University of Minnesota https://www.cmrr.umn.edu/multiband. The resolution is 1x1x1 mm and the field of view is 208x256x256 matrix. Straight sagittal orientation is used. TR and TE are 2000ms and 2.01ms respectively. The flip angle is 8 deg. Detailed sequence and imaging parameters are openly available here: https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/brain_mri.pdf

Area of acquisition: Whole brain

Diffusion MRI ☒ Used ☐ Not used

Parameters: Diffusion MRI data in the UK Biobank is obtained with two b-values (b = 1,000 and 2,000 s/mm2) at a spatial resolution of 2 mm using a multiband acceleration factor of 3, which allows for the acquisition of three slices simultaneously. For the two diffusion-weighted shells, 50 distinct diffusion-encoding directions were acquired (and all 100 directions are distinct). The diffusion preparation is a standard ("monopolar") Stejskal-Tanner pulse sequence. This enables higher SNR due to a shorter echo time (TE=92ms) than a twice-refocused ("bipolar") sequence. This improvement comes at the expense of stronger eddy current distortions, which are removed in the image processing pipeline.

Preprocessing

Preprocessing software: The T1-weighted structural imaging data were preprocessed with the Statistical Parametric Mapping software version 12 (https://www.fil.ion.ucl.ac.uk/spm/) using the CAT12 toolbox (https://neuro-jena.github.io/cat/) with default settings. The preprocessing involved high-dimensional spatial normalization with an integrated Dartel template in Montreal Neurological Institute (MNI) space, followed by nonlinear modulations and correction for each individual's head size. Following these procedures, gray matter images (voxel size: 1.5×1.5×1.5 mm3) were obtained for all participants.

The diffusion MRI data are corrected for distortions, eddy currents, and head motion and then modelled using FMRIB's Diffusion Toolbox for diffusion modelling and tractography analysis. Neurite orientation dispersion and density imaging modelling are conducted using Accelerated Microstructure Imaging via the Convex Optimization tool. White matter pathways are aligned cross-subject for extracting image-derived phenotypes (IDPs) using tract-based spatial statistics (TBSS), in which a high-dimensional warp maps a standard-space white matter skeleton to each participant.

Normalization: see above

Normalization template: see above

Noise and artifact removal: see above

Volume censoring: see above

Statistical modeling & inference

Model type and settings: First, we employed one-way analysis of covariance (ANCOVA) and post hoc tests to assess differences in measures of interest among the subtypes. Second, we applied Cox proportional hazards models to investigate differences in survival rates of common mental disorders among the subtypes. Third, we employed structural equation models (SEMs) to explore the associations between food-liking and three latent variables: mental health, cognitive function and brain MRI biomarkers.

Fourth, to explore the genetic underpinnings of distinct subtypes of food-liking, we performed GWAS analysis using logistic regression to compare subtypes. Finally, we performed gene expression and enrichment analysis on the identified genes in GWAS to provide further biological insights into the genetic associations with food-liking.

Effect(s) tested

In the one-way ANCOVAs and subsequent post hoc tests, we assessed statistical significance using F-values, t-values, and Cohen's d values. Levene's test was conducted to assess the equality of variances prior to one-way ANCOVAs. For the Cox proportional hazard regression model, we evaluated the proportional hazards assumption using the Schoenfeld residuals method and confirmed its satisfaction. The outcomes of the Cox models were presented as hazard ratios (HRs) along with their corresponding 95% confidence intervals (CIs). These HRs represent the averaged ratio of the hazard of developing mental disorders between the other three subtypes when compared to Subtype 4 within a 15-year follow-up period. In the SEMs, we employed confirmatory factor analysis (CFA) to estimate the latent variables within the model. Subsequently, we conducted Wald tests to determine two-sided P-values for the standardized coefficients, which were adjusted for multiple comparisons using the false discovery rate (FDR) correction. In addition, the root mean square error of approximation (RMSEA) of the SEM was used to assess model fitness.

Specify type of analysis: ☒ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference

(See Eklund et al. 2016)

Voxel-wise association, voxel-wise FDR correction

Correction

FDR correction

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☒	☐ Multivariate modeling or predictive analysis