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