nature portfolio

Corresponding author(s): Jiang Qiu

Last updated by author(s): 2024/08/17

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	Behavioral and neuroimaging data were obtained at Southwest University. The behavior data was obtained byE-prime 2.0 and Psychophysics Toolbox (http://psychtoolbox.org/) for MATLAB.
Data analysis	SPM12, Interpreting machine learning models in neuroimaging toolbox (https://github.com/cocoanlab/interpret_ml_neuroimaging), Neurosynth Python package (https://github.com/neurosynth/neurosynth/blob/master/neurosynth/analysis/decode.py), neuromaps (https://netneurolab.github.io/neuromaps/), abagen toolbox (https://github.com/netneurolab/abagen), Gene Ontology enrichment analysis and visualization tool (GOrilla, http://cbl-gorilla.cs.technion.ac.il/), Pattern_Regression_Matlab (https://github.com/ZaixuCui/Pattern_Regression_Matlab), fMRIPrep (J.Q.).

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

The data and material used in this study are available from the corresponding author upon reasonable request.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	All participants' gender was based on self-reported sex in this study. We did not specially considersex in the study design and did not analyze it.
Population characteristics	57 in Sample1, 31 in Sample 2, 410 in SLIM, 304 in GBB, 600 in BBP
Recruitment	In all datasets, participants were recruited via online and poster advertisements.
Ethics oversight	All participants provided written informed consent and received payment for their time and taskparticipation, and the research protocol was approved by the ethics committee of the reviewcommittee of the Brain Imaging Center of Southwest University.

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	task-based fmri data Sample 1 (43 females, mean age = 19.23 years, SD = 0.96, range = 17-22 years). Sample 2 (20 females, mean age = 19.30 years, SD = 2.49, range = 17-28 years). rest-state fmri data SLIM (233 females, mean age = 20.01 years, SD = 1.24, range = 17-27 years), GBB (218 females, mean age = 19.70 years, SD = 1.73, range = 17-26 years), BBP (412 females, mean age = 20.7 years, SD = 0.99, range = 17-27 years) .
Data exclusions	55 subjects within sample 1 were include in the following analysis. Sample 2 has a total 31 participants. One subject was excluded from subsequent analyses due to falling asleep during scanning. Thus, 30 subjects were included in the following analyses.
Replication	Two samples as validation set for each other.
Randomization	re was no group allocation in this study.
Blinding	There was no group allocation in this study.

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

Methods

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

Magnetic resonance imaging

Experimental design

Design type	task and resting state
Design specifications	The procedure consists of 4 sessions, each session includes 20 trials, with 12 NU conditions and 8 GU conditions in sampe1. The procedure consists of 6 sessions, total 40 NU conditions and 20 GU conditions in sample2.
Behavioral performance measures	the subject is asked to quickly press "1" with the left hand and then write down the response on the paper as quickly as possible

Acquisition

Imaging type(s)	functional
Field strength	3 Tesla (all datasets)
Sequence & imaging parameters	Whole-brain imaging was acquired on a Siemens 3T Trio scanner (Siemens Medical Systems, Erlangen, Germany) at the Brain Imaging Center, Southwest University. Task fMRI images were acquired using Gradient Echo Type Echo Planar Imaging (GRE-EPI) sequence (TR/TE = 2000 ms/30 ms, FA = 90 degrees, resolution matrix = 64 × 64, FOV= 220 × 220 mm2, thickness = 3 mm, slices = 32, interslice gap = 1 mm, acquisition voxel size = 3.4 × 3.4 × 4 mm3). High-resolution three-dimensional T1-weighted structure images were obtained using a Magnetization Prepared Rapid Acquisition Gradient-echo (MPRAGE) sequence (TR/TE =1900 ms /2.52 ms, FA = 9 degrees, FOV= 256 × 256 mm2, slices = 176, thickness = 1.0 mm, voxel size = 1 × 1 × 1 mm3).
Area of acquisition	whole-brain

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software	fMRIprep 1.15.1
Normalization	Linear realignment to MNI152Nlin6Asym spaces
Normalization template	MNI152NLin2009cAsym space
Noise and artifact removal	Principal components are estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128s cut-off) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, the aCompCor masks are subtracted a mask of pixels that likely contain a volume fraction of GM. This mask is obtained by thresholding the corresponding partial volume map at 0.05, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration.
Volume censoring	No volume was discarded.

Statistical modeling & inference

Model type and settings	We conducted two separate GLM (general linear model) analysis in SPM12: subject-level analysis and single-trial analysis in two samples respectively. We applied whole-brain (restricted to a gray matter mask) multivariate machine-learning pattern analysis to obtain a brain activation pattern that best classified subjects in NU and GU.
Effect(s) tested	we used cross-validation (CV) strategies to overcome the loss of generalization due to the small training and testing sample

Effect(s) tested

size in the neuroimaging applications. To evaluate the performance, 10×10-fold CV procedure on the sample1 and sample2 was conducted respectively. We evaluated the performance of the SVM models from the following indices: accuracy, sensitivity, specificity, ROC, AUC.

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

Statistic type for inference
(See Eklund et al. 2016)

Bootstrap tests were conducted to provide P-values for voxel weights to threshold the classifier weights for display and interpretation. We constructed 10,000 bootstrap sample sets (with replacement) and ran SVMs on each bootstrap sample. Each bootstrap sample serves as training data for a new model. The hyper-parameters of the model were the same as the original model. Then, two-tailed, 5000 permutation tests were conducted to calculation P-values for each voxel based on the proportion of weights above or below zero.

Correction

we performed false discovery rate (FDR) multiple comparison correction for the results of permutation test.

Models & analysis

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

Functional and/or effective connectivity

pearson correlation

Multivariate modeling and predictive analysis

We used the support vector machine (SVM) algorithm using a linear kernel (C = 1) implemented in the Interpreting machine learning models in neuroimaging toolbox (https://github.com/cocoanlab/interpret_ml_neuroimaging) with individual beta maps (one condition for each subject) as feature to classify subjects in NU and GU while undergoing fMRI. we used cross-validation (CV) strategies to overcome the loss of generalization due to the small training and testing sample size in the neuroimaging applications. To evaluate the performance, 10×10-fold CV procedure on the sample1 and sample2 was conducted respectively. we tested the SVM classifier with the random labeling, and calculated the accuracy of SVM classifier in sample 1 and sample 2 respectively. The permutation test was repeated 5000 times. To test the generalization performance of the two classification models, we performed validation using sample 2 (sample 1) as the validation set. We calculated the dot-product of vectorized activation images in sample 2 (sample 1) with the threshold weights of the model in sample 1 (sample 2).