nature neuroscience

Corresponding Author: Karla Miller

Manuscript Number: NN-RS54965C

Manuscript Type: Resource

Main Figures: 8

Supplementary Figures: 8

Supplementary Tables: 2

Supplementary Videos: 0

Reporting Checklist for Nature Neuroscience

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. For more information, please read Reporting Life Sciences Research.

Please note that in the event of publication, it is mandatory that authors include all relevant methodological and statistical information in the manuscript.

Statistics reporting, by figure

- Please specify the following information for each panel reporting quantitative data, and where each item is reported (section, e.g. Results, & paragraph number).
- Each figure legend should ideally contain an exact sample size (n) for each experimental group/condition, where n is an exact number and not a range, a clear definition of how n is defined (for example x cells from x slices from x animals from x litters, collected over x days), a description of the statistical test used, the results of the tests, any descriptive statistics and clearly defined error bars if applicable.
- For any experiments using custom statistics, please indicate the test used and stats obtained for each experiment.
- Each figure legend should include a statement of how many times the experiment shown was replicated in the lab; the details of sample collection should be sufficiently clear so that the replicability of the experiment is obvious to the reader.
- For experiments reported in the text but not in the figures, please use the paragraph number instead of the figure number.

Note: Mean and standard deviation are not appropriate on small samples, and plotting independent data points is usually more informative. When technical replicates are reported, error and significance measures reflect the experimental variability and not the variability of the biological process; it is misleading not to state this clearly.

		TEST USED		n			DESCRIPTIVE STATS (AVERAGE, VARIANCE)		P VALUE		DEGREES OF FREEDOM & F/t/z/R/ETC VALUE	
	FIGURE NUMBER	WHICH TEST?	SECTION & PARAGRAPH #	EXACT VALUE	DEFINED?	SECTION & PARAGRAPH #	REPORTED?	SECTION & PARAGRAPH #	EXACT VALUE	SECTION & PARAGRAPH #	VALUE	SECTION & PARAGRAPH #
example	1a	one-way ANOVA	Fig. legend	9, 9, 10, 15	mice from at least 3 litters/group	Methods para 8	error bars are mean +/- SEM	Fig. legend	p = 0.044	Fig. legend	F(3, 36) = 2.97	Fig. legend
example	results, para 6	unpaired t-test	Results para 6	15	slices from 10 mice	Results para 6	error bars are mean +/- SEM	Results para 6	p = 0.0006	Results para 6	t(28) = 2.808	Results para 6

		TEST USED		n			DESCRIPTIVE STATS (AVERAGE, VARIANCE)		P VALUE		DEGREES OF FREEDOM & F/t/z/R/ETC VALUE	
	FIGURE NUMBER	WHICH TEST?	SECTION & PARAGRAPH #	EXACT VALUE	DEFINED?	SECTION & PARAGRAPH #	REPORTED?	SECTION & PARAGRAPH #	EXACT VALUE	SECTION & PARAGRAPH #	VALUE	SECTION & PARAGRAPH #
+ -	1g-h, 2e, 3e, 4c-d, 6,8, Supp 1,2	Pearson correlation (& multiple regression in Fig 8)	Fig legend	5430	All subjects with non-missing data	Results p3 para 4	-log10(Puncorrected)	Fig legend	individual p-values plotted (1100 p-values per IDP)	Figure	DoF 5430, multiple comparison correction (Bonferroni and FDR)	Fig legend & online methods ("simple associations")
+ -	3d	Fixed effects one-group t-test	Fig legend	4111	All subjects with non-missing data	Results p6 para 2	Fixed effects Z-statistic	Fig legend	Z>100	Online methods p20 para	DoF>10,000	Results p6 para 2
+ -	5a-d	Pearson correlation	Fig legend	3722	All subjects with non-missing data	Results p6 para 4	r	Fig legend	|r|>0.1; p<0.05 Bonferroni corrected	Results p6 para 4	DoF 3722	Results p6 para 4
+ -	5e	Pearson correlation	Fig legend	3781	All subjects with non-missing data	Results p6 para 4	r	Fig legend	|r|>0.1; p<0.05 Bonferroni corrected	Results p6 para 4	DoF 3781	Results p6 para 4
+ -	7 & Supplemental figures 3-8	CCA+ICA (canonical correlation analysis + independent component analysis)	Fig legend; Results p11-13; Online Methods	5034	All subjects with non-missing rfMRI data	Fig legend; Results p11-13; Online Methods	correlations and % variance explained, for subject measures & IDPs vs ICA canonical variates.	Fig legend; Results p11-13; Online Methods	All 9 modes: p<0.002, family-wise-error corrected using permutation testing (nperm=10000)	Fig legend; Results p11-13; Online Methods	>200 (See Online Methods; DoF and multiple comparison correction directly accounted for by permutation testing)	Fig legend; Results p11-13; Online Methods

Representative figures

1. Are any representative images shown (including Western blots and immunohistochemistry/staining) in the paper?

   If so, what figure(s)?

   Figs 1-5,7-8; Supp Figs 1,3-7

2. For each representative image, is there a clear statement of how many times this experiment was successfully repeated and a discussion of any limitations in repeatability?

   If so, where is this reported (section, paragraph #)?

   N/A. All statistical images are described above; all other representative images are (clearly marked as) either group-averages (N>3500; see individual figures) or single-subject example images. The study for each subject is carried out just once, with large subject numbers (see above).

Statistics and general methods

1. Is there a justification of the sample size? If so, how was it justified? Where (section, paragraph #)? Even if no sample size calculation was performed, authors should report why the sample size is adequate to measure their effect size.	Discussion of study size in Results section 1 and Discussion. With respect to the sample size used in these analyses; we used all datasets available and complete for any given test (see also QC discussions in Online Methods).
2. Are statistical tests justified as appropriate for every figure? Where (section, paragraph #)?	Primary CCA+ICA mode results are tested for significance using permutation testing that accounts for multiple comparisons (Online Methods). Other tests are simple univariate association tests using Pearson correlation (see Online Methods for details of preprocessing).
a. If there is a section summarizing the statistical methods in the methods, is the statistical test for each experiment clearly defined?	yes (permutation test details, 10,000 permutations) and see above for all details.
b. Do the data meet the assumptions of the specific statistical test you chose (e.g. normality for a parametric test)? Where is this described (section, paragraph #)?	Yes - See Online Methods: data were normalised to Gaussian distributions before association/ICA tests.
c. Is there any estimate of variance within each group of data? Is the variance similar between groups that are being statistically compared? Where is this described (section, paragraph #)?	N/A (no multiple group comparison tests).
d. Are tests specified as one- or two-sided?	Two-sided
e. Are there adjustments for multiple comparisons?	Yes - as described above, for different tests, full correction was made for numbers of voxels or numbers of tests (IDPs and other variables). In association tests, both Bonferroni and FDR threshold levels are shown.
3. To promote transparency, *Nature Neuroscience* has stopped allowing bar graphs to report statistics in the papers it publishes. If you have bar graphs in your paper, please make sure to switch them to dot-plots (with central and dispersion statistics displayed) or to box-and-whisker plots to show data distributions.	N/A
4. Are criteria for excluding data points reported? Was this criterion established prior to data collection? Where is this described (section, paragraph #)?	See above - all non-missing data included.

5. Define the method of randomization used to assign subjects (or samples) to the experimental groups and to collect and process data. If no randomization was used, state so. Where does this appear (section, paragraph #)?	N/A
6. Is a statement of the extent to which investigator knew the group allocation during the experiment and in assessing outcome included? If no blinding was done, state so. Where (section, paragraph #)?	N/A
7. For experiments in live vertebrates, is a statement of compliance with ethical guidelines/regulations included? Where (section, paragraph #)?	N/A
8. Is the species of the animals used reported? Where (section, paragraph #)?	N/A
9. Is the strain of the animals (including background strains of KO/ transgenic animals used) reported? Where (section, paragraph #)?	N/A
10. Is the sex of the animals/subjects used reported? Where (section, paragraph #)?	yes - detailed in Online Methods
11. Is the age of the animals/subjects reported? Where (section, paragraph #)?	yes - detailed in Online Methods
12. For animals housed in a vivarium, is the light/dark cycle reported? Where (section, paragraph #)?	N/A
13. For animals housed in a vivarium, is the housing group (i.e. number of animals per cage) reported? Where (section, paragraph #)?	N/A
14. For behavioral experiments, is the time of day reported (e.g. light or dark cycle)? Where (section, paragraph #)?	N/A
15. Is the previous history of the animals/subjects (e.g. prior drug administration, surgery, behavioral testing) reported? Where (section, paragraph #)?	Fully described in the core UK Biobank literature and URLs referenced.

a. If multiple behavioral tests were conducted in the same group of animals, is this reported?

Where (section, paragraph #)?

N/A

16. If any animals/subjects were excluded from analysis, is this reported?

Where (section, paragraph #)?

Yes, see above.

a. How were the criteria for exclusion defined?

Where is this described (section, paragraph #)?

See above; QC section in Online Methods; otherwise all non-missing data was used.

b. Specify reasons for any discrepancy between the number of animals at the beginning and end of the study.

Where is this described (section, paragraph #)?

N/A

Reagents

1. Have antibodies been validated for use in the system under study (assay and species)?

N/A

a. Is antibody catalog number given?

Where does this appear (section, paragraph #)?

N/A

b. Where were the validation data reported (citation, supplementary information, Antibodypedia)?

Where does this appear (section, paragraph #)?

N/A

2. Cell line identity

N/A

a. Are any cell lines used in this paper listed in the database of commonly misidentified cell lines maintained by ICLAC and NCBI Biosample?

Where (section, paragraph #)?

b. If yes, include in the Methods section a scientific justification of their use--indicate here in which section and paragraph the justification can be found.

N/A

c. For each cell line, include in the Methods section a statement that specifies:

- the source of the cell lines
- have the cell lines been authenticated? If so, by which method?
- have the cell lines been tested for mycoplasma contamination?

Where (section, paragraph #)?

N/A

Data availability

Provide a Data availability statement in the Methods section under "Data availability", which should include, where applicable:

- Accession codes for deposited data
- Other unique identifiers (such as DOIs and hyperlinks for any other datasets)
- At a minimum, a statement confirming that all relevant data are available from the authors
- Formal citations of datasets that are assigned DOIs
- A statement regarding data available in the manuscript as source data
- A statement regarding data available with restrictions

See our data availability and data citations policy page for more information.

Data deposition in a public repository is mandatory for:

a. Protein, DNA and RNA sequences
b. Macromolecular structures
c. Crystallographic data for small molecules
d. Microarray data

Deposition is strongly recommended for many other datasets for which structured public repositories exist; more details on our data policy are available here. We encourage the provision of other source data in supplementary information or in unstructured repositories such as Figshare and Dryad.

We encourage publication of Data Descriptors (see Scientific Data) to maximize data reuse.

Where is the Data Availability statement provided (section, paragraph #)?

See first 2 sections of Online Methods for a full description of the data access procedure for UK Biobank; all data is available openly for any researchers after following an application procedure.

For new results shown in this paper, see the Online Methods section "Data, code and results availability". All will be made freely available at the point of publication.

Computer code/software

Any custom algorithm/software that is central to the methods must be supplied by the authors in a usable and readable form for readers at the time of publication. However, referees may ask for this information at any time during the review process.

1. Identify all custom software or scripts that were required to conduct the study and where in the procedures each was used.

All software used (Matlab, FSL, nearestSPD, AMICO) is identified in detail in Online Methods.

2. If computer code was used to generate results that are central to the paper's conclusions, include a statement in the Methods section under "**Code availability**" to indicate whether and how the code can be accessed. Include version information as necessary and any restrictions on availability.

See the Online Methods section "Data, code and results availability". All matlab code will be made freely available at the point of publication.

Human subjects

1. Which IRB approved the protocol? Where is this stated (section, paragraph #)?	Fully described in the core UK Biobank literature / URLs referenced here; the paper is only using publicly available datasets. A detailed statement on this is included in the "Protocol considerations" section of Online Methods.
2. Is demographic information on all subjects provided? Where (section, paragraph #)?	Yes; see Online Methods (throughout).
3. Is the number of human subjects, their age and sex clearly defined? Where (section, paragraph #)?	Yes; see Online Methods (throughout).
4. Are the inclusion and exclusion criteria (if any) clearly specified? Where (section, paragraph #)?	Yes; see above.
5. How well were the groups matched? Where is this information described (section, paragraph #)?	N/A
6. Is a statement included confirming that informed consent was obtained from all subjects? Where (section, paragraph #)?	Fully described in the core UK Biobank literature / URLs referenced here; the paper is only using publicly available datasets.
7. For publication of patient photos, is a statement included confirming that consent to publish was obtained? Where (section, paragraph #)?	N/A

fMRI studies

For papers reporting functional imaging (fMRI) results please ensure that these minimal reporting guidelines are met and that all this information is clearly provided in the methods:

1. Were any subjects scanned but then rejected for the analysis after the data was collected?	Yes - see above for full details regarding exclusions.
a. If yes, is the number rejected and reasons for rejection described? Where (section, paragraph #)?	See above and Online Methods throughout; Of 5430 subjects considered for fMRI analysis, 5034 had complete data (see above for criteria).
2. Is the number of blocks, trials or experimental units per session and/or subjects specified? Where (section, paragraph #)?	Yes, depicted in Fig 3.
3. Is the length of each trial and interval between trials specified?	Yes - see above and Online Methods (tfMRI protocol).
4. Is a blocked, event-related, or mixed design being used? If applicable, please specify the block length or how the event-related or mixed design was optimized.	Blocked. Paradigm described fully in reference Barch 2013.

5. Is the task design clearly described? Where (section, paragraph #)?	See above.
6. How was behavioral performance measured?	Recorded (and available in the Biobank database) but not used for the analyses shown here.
7. Is an ANOVA or factorial design being used?	N/A
8. For data acquisition, is a whole brain scan used? If not, state area of acquisition.	Yes
a. How was this region determined?	N/A
9. Is the field strength (in Tesla) of the MRI system stated?	yes (3T; see Online Methods)
a. Is the pulse sequence type (gradient/spin echo, EPI/spiral) stated?	yes - full details in Online Methods.
b. Are the field-of-view, matrix size, slice thickness, and TE/TR/ flip angle clearly stated?	yes - full details in Online Methods.
10. Are the software and specific parameters (model/functions, smoothing kernel size if applicable, etc.) used for data processing and pre-processing clearly stated?	yes - full details in Online Methods and Biobank technical documentation.
11. Is the coordinate space for the anatomical/functional imaging data clearly defined as subject/native space or standardized stereotaxic space, e.g., original Talairach, MNI305, ICBM152, etc? Where (section, paragraph #)?	Stated as MNI152 in Online Methods.
12. If there was data normalization/standardization to a specific space template, are the type of transformation (linear vs. nonlinear) used and image types being transformed clearly described? Where (section, paragraph #)?	Linear+Nonlinear; see Online Methods
13. How were anatomical locations determined, e.g., via an automated labeling algorithm (AAL), standardized coordinate database (Talairach daemon), probabilistic atlases, etc.?	N/A
14. Were any additional regressors (behavioral covariates, motion etc) used?	Yes - see full description of confound regressors in Online Methods.
15. Is the contrast construction clearly defined?	Yes - see Online Methods
16. Is a mixed/random effects or fixed inference used?	For group-average task activation, fixed-effects is used, alongside number of subjects activated.

a. If fixed effects inference used, is this justified?	N/A (the numbers are so large - several thousand subjects - that fixed-effects - i.e. an "average subject" in this context - is more meaningful than mixed-effects.) Of much more relevance to quantitating subject reliability is Fig 3c, showing fraction of subjects activating (see also Barch ref for further examples).
17. Were repeated measures used (multiple measurements per subject)?	No - see above.
a. If so, are the method to account for within subject correlation and the assumptions made about variance clearly stated?	N/A
18. If the threshold used for inference and visualization in figures varies, is this clearly stated?	N/A
19. Are statistical inferences corrected for multiple comparisons?	Yes
a. If not, is this labeled as uncorrected?	N/A
20. Are the results based on an ROI (region of interest) analysis?	N/A
a. If so, is the rationale clearly described?	N/A
b. How were the ROI's defined (functional vs anatomical localization)?	N/A
21. Is there correction for multiple comparisons within each voxel?	N/A
22. For cluster-wise significance, is the cluster-defining threshold and the corrected significance level defined?	N/A

Additional comments

Additional Comments	