nature research

Corresponding author(s): Jeroen Mollink

Last updated by author(s): 22-02-2019

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees 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 was not explicitly acquired in this study, but originated from the UK Biobank project. Details about genetics data acquisition can be found at Bycroft et al 2017. Genotyping of all subjects was performed using the Applied BiosystemsTM UK BiLEVE AxiomTM Array by Affymetrix1 (807,411 markers) or using the closely-related Applied BiosystemsTM UK Biobank AxiomTM Array (825,927 markers). Both arrays were purpose-designed specifically for the UK Biobank genotyping project and share 95% of marker content. MRI data acquisition and protocols can be found in Alfaro-Almagro et al., 2018 and Miller et al., 2016 or at http://biobank.ctsu.ox.ac.uk/crystal/docs/brain_mri.pdf. All MRI data were acquired with 3T Siemens Skyra (software platform VD13).
Data analysis	FMRIB's Software Library (FSL) v5.0 for MRI (pre)processing. Part of FSL: - BIANCA for estimating white matter hyperintensities from T2-FLAIR data - MELODIC tool for estimating resting-state networks from fMRI data using ICA - FSLNets toolbox for processing the resting-state time-series and network analyses - BedpostX for estimating sub-voxel fibre configurations from diffusion MRI data. - Probtrackx2 to delineate white matter bundles using tractography. - TBSS for estimating the tract skeleton of white matter bundles. AMICO for estimating NODDI parameters from diffusion MRI. Matlab r2016a for data analysis, multivariate modeling and plotting. BGENIE v1.2 for genome-wide associations.

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

The UK Biobank Brain imaging protocol consists of 6 distinct modalities covering structural, diffusion and functional imaging, summarised in Supplementary Table 1. For this study, we primarily used data from the February 2017 release of ~10,000 participants' imaging data (and an additional ~5,000 subjects' data released in January 2018 provided the larger replication sample). The raw data from these 6 modalities has been processed for UK Biobank to create a set of imaging derived phenotypes (IDPs). These are available from UK Biobank, and it is these IDPs from the 2017/18 data releases that we used in this study (see http://www.ukbiobank.ac.uk/register-apply).

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

No sample-size calculation was performed, as large epidemiological studies like the UK Biobank project (and others) aim to maximize sample sizes to detect small variations across subjects. The sample sizes were determined based on the availability of usable MRI data in addition to genetics data without artefacts from subjects with a recent British ancestry. These sample sizes deemed sufficient to detect effect sizes reported in our work (>0.1% variance explained).

Data exclusions

Exclusion criteria were pre-established. With regards to the MRI data, a subject can be excluded based on the T1-weighted scan if registration to standard space fails, likely due to excessive head motion, atypical structure and/or anatomical abnormalities (e.g., large ventricles). Subjects can additionally be excluded from further analysis on the basis of their dMRI and fMRI data due to bad EPI distortions, failed registration to T1, extreme bias fields, unusable fieldmaps and/or severe motion artefacts. 87% of the dMRI datasets and 94% of the rfMRI datasets were considered suitable for further analysis based on these QC measures.

Similarly, subjects were selected on usable genetics data. As in Elliott et al, to avoid confounding effects that may arise from population structure or environmental effects, we selected unrelated subjects with recent British ancestry. Ancestry was determined using sample quality control information provided by UK Biobank. We then filtered the genetic data to remove SNPs with minor allele frequency < 0.01% or a Hardy-Weinberg equilibrium p-value of less than 10-7, yielding a total of 11,734,353 SNPs distributed across the 22 autosomes.
All subjects used in our study passed these quality control criteria, yielding a total of 11354 subjects

Replication

All models estimating functional connectivity (derived from resting-state fMRI) from white matter microstructural metrics (derived from diffusion MRI) were trained on the main cohort of 7481 subjects. We used an additional subset of 3873 subjects to replicate these findings. Likewise, we performed a genome-wide association in the main cohort and replicated the reported genetic variants in the replication cohort.

Randomization

Subjects were randomly assigned to the two cohorts (main and replication).

Blinding

There were no experimental groups, so no blinding steps were involved 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
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

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

Human research participants

Policy information about studies involving human research participants

Population characteristics

Sample size, 11354 subjects (7481 subjects in the main cohort + 3873 subjects in the replication cohort); 5393 females; age, 62.8 (SD 7.4) years. All subjects had recent British ancestry.

Recruitment

Participants were selected using the NHS register, and invited to volunteer for the study. Recruitment was carried out between 2007 and 2010. Full details of the recruitment process are available in UK Biobank: Protocol for a large-scale prospective epidemiological resource, 2007 (http://www.ukbiobank.ac.uk/wp-content/uploads/2011/11/UK-Biobank-Protocol.pdf)

Ethics oversight

Ethics is handled by the UK Biobank Ethics Advisory Committee. Further details can be found at https://www.ukbiobank.ac.uk/ethics/

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

Magnetic resonance imaging

Experimental design

Design type

Resting-state functional MRI data, no task.

Design specifications

No tasks were performed during fMRI scanning, so no design.

Behavioral performance measures

No tasks were performed during fMRI scanning, so no behavioral performance measures.

Acquisition

Imaging type(s)

T1-weighted MRI, T2-weighted MRI, diffusion MRI and resting-state functional MRI

Field strength

3T

Sequence & imaging parameters

T1-weighted MRI: 3D MPRAGE protocol
TI = 880 ms, TR = 2000 ms and an in-plane acceleration factor of 2.
Field of view 208x256x256 matrix
Resolution 1.0x1.0x1.0 mm

T2-weighted MRI: Fluid-attenuated inversion recovery (FLAIR) protocol (3D SPACE).
TI = 1800, TR = 5000 ms and an in-plane acceleration factor of 2.
Field of view: 192x256x56 matrix.
Resolution: 1.05x1.0x1.0 mm.

Diffusion MRI: Diffusion weighted spin-echo EPI sequence using multi-band (MB) acceleration.
MB = 3, R = 1, TE/TR = 92/3600 ms, no iPAT, PF 6/8, fat saturation.
Field of view: 104x104x72 matrix
Resolution: 2x2x2mm.
b = 0 (5x + 3× phase-encoding reversed), b = 1000, (50×), b = 2000 (50×).

Resting-state functional MRI: Gradient echo EPI sequence using multi-band acceleration.
TE/TR = 39/735 ms, MB = 8, R = 1, no iPAT, flip angle 52°, fat saturation.
Field of view: 88x88x64 matrix.
Resolution: 2.4x2.4x2.4mm.
490 time-points.

Area of acquisition

Whole brain

Diffusion MRI ☒ Used ☐ Not used

Parameters: 100 gradient directions over two shells, 50 directions/shell. b-values = 1000, 2000 s/mm2

Preprocessing

Preprocessing software	FMRIB's Software Library (FSL), v5.0
Normalization	Diffusion MRI: Subject's fractional anisotropy (FA) maps (obtained after fitting the diffusion tensor model to the raw diffusion MRI data) were transformed using FNIRT (part of FSL) to a 1-mm FA template in MNI-space.
Normalization template	Diffusion MRI: FA template FMRIB58_FA_1mm included in the FSL v5.0 software package. Resting-functional MRI: T1 template, MNI152_T1_2mm included in the FSL v5.0 software package.
Noise and artifact removal	The resting-state functional MRI data was motion corrected (Jenkinson et al., 2002) and FIX-cleaned (Salimi-Khorshidi et al., 2014) to remove physiological noise and image artefacts.
Volume censoring	Motion was corrected using MCFLIRT as implemented in FSL for the resting-state fMRI data. Diffusion MRI data was motion corrected with EDDY, also implemented in FSL. Subjects with severe head motion were not included in the preprocessed data released by the UK Biobank.

Statistical modeling & inference

Model type and settings	Multiple linear regression models, see below: Multivariate modeling and predictive analysis.
Effect(s) tested	No effects tested, because no task fMRI was performed.

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

Anatomical location(s)	The resting-state fMRI data were fed into an Independent Component Analysis (ICA) using the MELODIC tool (Beckmann and Smith, 2004) to identify resting-state networks present on average in the whole population. First, data was reduced to 100 dimensions using PCA and then fed into spatial ICA, from which 55 components corresponded to functional regions, and the other 45 judged to reflect physiological noise or image artifacts ("noise"). A functional component was split if it consisted of noncontiguous brain regions, yielding 81 bilateral (homotopic) regions that were further split between the hemispheres to estimate interhemispheric connectivity (see Supplementary Table 1). The 81 homotopic region pairs were all located in cortical grey matter. Cerebellar and sub-cortical components were not included in this analysis.
Statistic type for inference (See Eklund et al. 2016)	See above, clusters were obtained using spatial independent components analysis.
Correction	Statistical significance of the regression models was assessed by means of permutation testing. A null distribution was constructed for each regressor by randomly permuting the functional connectivity values (the number of permutations was set to 100,000). A p-value (two-sided) was then determined in the non-permuted model from the null distribution. Because multiple models were evaluated, we corrected for the family wise error as in (Winkler et al., 2014). Here, we generated a maximum t-statistic distribution across all homotopic region pairs and regressors (i.e., the microstructural principal components) of the permuted t-statistics. From this maximum t-statistics null-distribution a corrected p-value was estimated based for each of the non-permuted t-statistics. See Multivariate modelling and predictive analysis for more information.

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	Functional connectivity was estimated between all pairs of homotopic regions (2x81) by means of partial correlation of the cleaned time-series using Ridge regression with a regularization factor ρ=1. Partial correlation aims to measure direct connectivity between two areas by first regressing out all other regions' time-series before calculating the correlation (i.e., established through inversion of the covariance matrix).
Multivariate modeling and predictive analysis	We used a multiple linear regression model to predict homotopic connectivity from a set of regressors describing the spatial pattern of microstructure along a white matter tract. The regression model was constructed for each pair of homotopic regions separately: $Y_i = X_i \beta + \varepsilon_i$, with $i=1,\dots,n$ Here Y_i (Nsubjects x 1) is a vector that contains the functional connectivity values of all subjects derived from homotopic region i (over n = 81 regions). To build a model using p microstructural regressors, we need to estimate a set of regression coefficients β (p x 1) that describe the relative contribution from the microstructural metrics X_i (Nsubjects x p) along the white matter tract. The regressors are derived in two stages. First, the microstructural metrics were extracted from the TBSS-

voxels (white matter skeleton) corresponding to the tract of interest for every subject, yielding a matrix X†i (Nsubjects x Nvoxels). As the matrix X†i is very large, a direct regression with functional connectivity is ill conditioned. We therefore perform a dimensionality reduction on X†I to derive a set of regressors reflecting the primary modes of variation of a given microstructural metric across space for the cohort of subjects. A singular value decomposition (SVD) was computed from matrix X†i, from which the top p components were retained, yielding matrix Xi (Nsubjects x p). In practice, p was set to 30 principal components, which approximately corresponded to a transition in the spectrum of singular values in terms of variance explained, above which variance explained roughly tracked noise singular vectors (Supplementary Fig. 1).

Matrices Xi were constructed for each of the microstructure metrics separately, yielding six single-metric linear regression models per homotopic region. In addition, a multimodal regression model was created that combined across all microstructure metrics. For the multimodal regression, all raw microstructure matrices (X†i) were normalized through division by their first singular value to ensure comparable range of values. The six normalized matrices were then concatenated and an SVD was performed on the concatenated matrix to reduce back to the top 30 components.

Finally, we defined a set of confound variables of no interest (age, age^2, sex, age*sex, age^2*sex, resting-state fMRI head motion, and head size) that could correlate with estimated microstructural measures (e.g. through artefacts such as partial volume) and thereby bias the estimated regressors. The confound variables were regressed out of the functional and microstructural data before fitting each regression model.