nature research

Corresponding author(s): Hongtu Zhu

Last updated by author(s): 08/22/2019

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

There is no clear distinction between software/code used for data collection vs data analysis, thus we list all software here:

ANTs v2.1.0 and v2.2.0, http://stnava.github.io/ANTs/;
PLINK v1.90 beta, https://www.cog-genomics.org/plink2/;
GCTA v1.91.7beta, http://cnsgenomics.com/software/gcta/;
METAL v2011-03-25, https://genome.sph.umich.edu/wiki/METAL;
FUMA v1.3.4, http://fuma.ctglab.nl/;
MGAMA v1.07, https://ctg.cncr.nl/software/magma;
LD Score Regression v1.0.0, https://github.com/bulik/ldsc/;
LD Hub v1.9.1, http://ldsc.broadinstitute.org/ldhub/;
XWAS v3.0, http://keinanlab.cb.bscb.cornell.edu/content/xwas;
DEPICT v1 rel194, https://github.com/perslab/depict;
PRScs v2019-05-15, https://github.com/getian107/PRScs;
MSigDB v6.2, http://software.broadinstitute.org/gsea/msigdb;
MaCH-Admix v2.0.203, http://www.unc.edu/~yunmli/MaCH-Admix ;
NHGRI-EBI GWAS Catalog v2019-01-31, https://www.ebi.ac.uk/gwas/home/;
The atlas of GWAS Summary Statistics v20190131, http://atlas.ctglab.nl/ (for genetic variants);
The atlas of GWAS Summary Statistics v20190503, http://atlas.ctglab.nl/ (for genes);
UK Biobank, http://www.ukbiobank.ac.uk/resources/;
PING, http://pingstudy.ucsd.edu/resources/genomics-core.html/;
PNC, https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000607.v1.p1/;
ADNI, http://adni.loni.usc.edu/data-samples/;
HCP, https://www.humanconnectome.org/.

Data analysis | Please see above.

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 data used in this work were obtained from five publicly available datasets: the UK Biobank (UKB) study, the Human Connectome Project (HCP) study, the Pediatric Imaging, Neurocognition, and Genetics (PING) study, the Philadelphia Neurodevelopmental Cohort (PNC) study, and the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. This research has been conducted using the UK Biobank resource (application number 22783), subject to a data transfer agreement.
For UKB, the imputed genetic variants data was released in July 2017, and we used the imaging data of ~22,000 participants released until August 2018.
The raw MRI, covariates and genetic variants data were available from each data resource:
UK Biobank, http://www.ukbiobank.ac.uk/resources/;
PING, http://pingstudy.ucsd.edu/resources/genomics-core.html/;
PNC, https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000607.v1.p1/;
ADNI, http://adni.loni.usc.edu/data-samples/;
HCP, https://www.humanconnectome.org/.

We used 50 sets of publicly available GWAS summary statistics from several GWAS databases. The data resources are summarized in Supplementary Table 24.

The full set of UKB and meta-analysis GWAS summary statistics of ROI volumes are available at: https://med.sites.unc.edu/bigs2/data/gwas-summary-statistics/.

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 power calculation was needed in advance. We used all samples passing standard quality controls (please see below). The sample size in current analysis was greater than that of most of the previous GWAS on brain volumetric phenotypes.
Data exclusions	The main GWAS made use of data of individuals of British ancestry (self-reported ethnic background, Data-Field 21000) from the UKB study, and the other four GWAS (PING, PNC, HCP, ADNI) were performed on individuals of European ancestry. For the imaging data, we removed three ROIs (X5th ventricle and left/right lesion) with missing rates > 99%. For each phenotype and continuous covariate variable, we further removed values greater than five times the median absolute deviation from the median value. For the genetic variants data, we performed standard quality controls on each dataset, including 1) exclude subjects with more than 10% missing genotypes; 2) exclude variants with minor allele frequency less than 0.01; 3) exclude variants with larger than 10% missing genotyping rate; 4) exclude variants that failed the Hardy-Weinberg test at 1*10^{-7} level; and 5) remove variants with imputation INFO score less than 0.8. For X chromosome analysis, the following X-specific quality control steps were performed: 1) variants on chromosomes other than X were removed, as well as variants in the pseudoau-tosomal regions (PARs) on X; 2) variants were removed if they had significantly different MAF between male and female (p-value <1.76*10^{-7}, Bonferroni-corrected). All the data exclusion criteria were pre-established.
Replication	The significant genetic variants discovered in the UKB sample were supported by a joint analysis with other four independent studies. We checked whether the variant effect signs were concordant in the five studies and whether the p-value of top UKB variants decreased after meta-analysis. Below are the main results of joint analysis: The joint analysis was carried out on 3,841,911 genetic variants which were present in all five sets of GWAS results. For the 7,310 significant associations, 63.8% (4,666) associations had the same effect signs across the five studies, and 97.0% (7,090) associations had the same effect signs in at least four studies (including UKB). 93.9% (1,877) of the top 2,000 significant associations had smaller p-value after meta-analysis, and 91.4% (6,678) of all the 7,310 associations were enhanced.
Randomization	All the five datasets are from observational studies, and we used all samples available after data exclusions listed above. Therefore, there is no equivalent process of randomization in the present analysis.

Blinding	The data are not from controlled randomized studies, thus there is no step equivalent to blinding involved.

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	The main GWAS made use of data of individuals of British ancestry from the UKB study, and the other four GWAS were performed on individuals of European ancestry. The UKB genetic data has 8,944,375 genetic variants after genotyping quality controls, all individuals were ages between 40 and 80 with mean 62.51, the proportion of male is 0.47. More information about other study cohorts can be found in Supplementary Tables 29-30 and Supplementary Note.
Recruitment	Patients were recruited differently in each of the cohorts used, with some cohorts collected from general population and others from hospitals. Recruitment details and dataset overviews can be found in Sudlow et al. for UKB (https://doi.org/10.1371/journal.pmed.1001779), Satterthwaite et al. for PNC (https://doi.org/10.1016/j.neuroimage.2013.07.064), Weiner et al. for ADNI (https://doi.org/10.1016/j.jalz.2013.05.1769), and Jernigan et al. for PING (https://doi.org/10.1016/j.neuroimage.2015.04.057). The UKB significant variants were supported in the joint analysis and the UKB GWAS results had satisfactory prediction ability on four other independent cohorts suggesting that there was limited bias due to sample recruitment.
Ethics oversight	For UKB, the wide consultation, rigorous Ethics and Governance Framework, and Ethics and Governance Council oversight role have been essential in paving the way for UK Biobank to accomplish obtaining the multiple ethical and regulatory approvals required for participant recruitment, sample and data storage, linkages to routine health care data, enhancement studies, and the provision of access to data and samples for approved researchers. Substantial amounts of time, resources, patience, tenacity, and evidence of feasibility and/or acceptability from smaller scale pilot studies have also been required to provide regulatory bodies with the reassurance that they need of UK Biobank's rigorous approach and commitment to protecting the interests of its participants within an acceptable legal and ethical framework (details can be found in Sudlow et al. https://doi.org/10.1371/journal.pmed.1001779). More information about these study cohorts can be found in the above references in section "Recruitment", the acknowledgment of the main text, and the Supplementary Note.

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	Please see the Online Methods section for full details. This study made use of imaging data from Structural MRI and genetic variants data.
Design specifications	Details can be found in Miller et al. (https://doi:10.1038/nn.4393) and Alfaro-Almagro et al. (https://doi.org/10.1016/j.neuroimage.2017.10.034) for UKB, and Satterthwaite et al. for PNC (https://doi.org/10.1016/j.neuroimage.2013.07.064), Weiner et al. for ADNI (https://doi.org/10.1016/j.jalz.2013.05.1769), and Jernigan et al. for PING (https://doi.org/10.1016/j.neuroimage.2015.04.057). We then processed all the MRI data to generate imaging phenotypes using consistent procedures via advanced normalization tools (ANTs). The processing steps by ANTs were detailed in Tustison et al. (https://doi.org/10.1016/j.neuroimage.2014.05.044).
Behavioral performance measures	Behavioral performance measures were not used in this study.

Acquisition

Imaging type(s): Structural MRI.

Field strength: 3T in UKB, PNC, PING, and HCP; 1.5 T or 3T for ADNI.

Sequence & imaging parameters: Details can be found in Miller et al. (https://doi:10.1038/nn.4393) and Alfaro-Almagro et al. (https://doi.org/10.1016/j.neuroimage.2017.10.034) for UKB, and Satterthwaite et al. for PNC (https://doi.org/10.1016/j.neuroimage.2013.07.064), Weiner et al. for ADNI (https://doi.org/10.1016/j.jalz.2013.05.1769), and Jernigan et al. for PING (https://doi.org/10.1016/j.neuroimage.2015.04.057).

Area of acquisition: The whole brain scan was used.

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software: Details can be found in Miller et al. (https://doi:10.1038/nn.4393) and Alfaro-Almagro et al. (https://doi.org/10.1016/j.neuroimage.2017.10.034) for UKB, Satterthwaite et al. for PNC (https://doi.org/10.1016/j.neuroimage.2013.07.064), Weiner et al. for ADNI (https://doi.org/10.1016/j.jalz.2013.05.1769), and Jernigan et al. for PING (https://doi.org/10.1016/j.neuroimage.2015.04.057). The processing steps by ANTs were detailed in Tustison et al. (https://doi.org/10.1016/j.neuroimage.2014.05.044).

Normalization: Normalization/standardization using the ANTs software were detailed in Tustison et al. (https://doi.org/10.1016/j.neuroimage.2014.05.044) and Avants et al. (https://doi.org/10.1016/j.neuroimage.2010.09.025).

Normalization template: We use the OASIS-30 Atropos template for registration and Mindboggle-101 atlases for labeling. Details can be found in https://mindboggle.info/data.html, Klein and Tourville (https://doi.org/10.3389/fnins.2012.00171) and Tustison et al. (https://doi.org/10.1016/j.neuroimage.2014.05.044). The processing steps by ANTs were detailed in Tustison et al. (https://doi.org/10.1016/j.neuroimage.2014.05.044).

Noise and artifact removal: Noise and artifact removal from raw data can be found in Miller et al. (https://doi:10.1038/nn.4393) and Alfaro-Almagro et al. (https://doi.org/10.1016/j.neuroimage.2017.10.034) for UKB, Satterthwaite et al. for PNC (https://doi.org/10.1016/j.neuroimage.2013.07.064), Weiner et al. for ADNI (https://doi.org/10.1016/j.jalz.2013.05.1769), and Jernigan et al. for PING (https://doi.org/10.1016/j.neuroimage.2015.04.057). Further processing steps such as N4 bias correction by ANTs were detailed in Tustison et al. (https://doi.org/10.1016/j.neuroimage.2014.05.044).

Volume censoring: No volume censoring was used in processing structural images.

Statistical modeling & inference

Model type and settings: Statistical modeling was not used when generating imaging phenotypes. But within this study inference was applied at the level of the combined imaging-genetics modelling.

Effect(s) tested: Statistical modeling was not used when generating imaging phenotypes. But within this study inference was applied at the level of the combined imaging-genetics modelling.

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

Anatomical location(s): We use the OASIS-30 Atropos template for registration and Mindboggle-101 atlases for labeling. Details can be found in https://mindboggle.info/data.html, Klein and Tourville (https://doi.org/10.3389/fnins.2012.00171) and Tustison et al. (https://doi.org/10.1016/j.neuroimage.2014.05.044). The 101 brain region parcellation was performed by the Multi-Atlas joint label fusion using ANTs. The processing steps by ANTs were detailed in Tustison et al. (https://doi.org/10.1016/j.neuroimage.2014.05.044).

Statistic type for inference (See Eklund et al. 2016): Inference was not carried out when generating imaging phenotypes. But within this study inference was applied at the level of the combined imaging-genetics modelling.

Correction: Inference was not carried out when generating imaging phenotypes. But within this study inference was applied at the level of the combined imaging-genetics modelling.

Models & analysis

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