nature research

Corresponding author(s): Drs. Satizabal & Ikram NG-A47229R

Last updated by author(s): Aug 28, 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	No software were used
Data analysis	Associations at the individual study level: mach2qtl, probABEL, plink, merlin, SNPtest, RareMetalWorker. Neuroimaging processing: FreeSurfer (versions 3.0.2, 4.0.1, 4.3.0, 4.5.0, 5.0.0, 5.1.0, 5.3.0), SPM99, FSL-FIRST (versions 4.1.5, 4.1.7, 4.1.9, 5.0.4), custom image analysis pipeline based on the MNI software, in-house imaging software, MIPAV. Imputation on 1000 genomes or HRC: beagle, MaCH, IMPUTE, minimac, SHAPEIT. QC, heritability and meta-analysis: EASYQC, SOLAR, R, METAL, GCTA, LDSC, HASE. Functional follow-up: R, rAggr, Locus Zoom, LDSC, DIOPT, FlyBase, STRING.

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

We have included the following statement: "The genome-wide summary statistics that support the findings of this study will be made available through the CHARGE dbGaP (accession number phs000930) and ENIGMA (http://enigma.ini.usc.edu/research/download-enigma-gwas-results) websites."

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	We did not base our study on a calculated sample size. The sample size of each study sample included in the GWAS was determined by the number of individuals with both brain MRI outcomes and GWAS data. Following current practice for meta-analysis of GWAS, our study includes the largest possible sample from CHARGE and ENIGMA cohorts that were able to contribute data. We used the largest UKBB sample with QC'd imaging and GWAS data that was available to us.
Data exclusions	Our analysis pre-specified the exclusion of persons with prevalent stroke and dementia at the time of MRI, as well as those with neuroimaging abnormalities (i.e. brain tumor, large brain infarcts) that could have influenced the measurement of MRI outcomes.
Replication	We sought replication in three samples of African-American ancestry (up to n=769), and two of Asian (Chinese/Malay) ancestry (n=341). Due to differences in allele frequency in non-European samples, some of our lead variants were not present in one or more of our replication samples. Details are presented in Supplementary Table S7.
Randomization	Randomization was not used in this observational study.
Blinding	During data acquisition, the personnel in charge of processing MRI data were blinded to the genetic participant's data and vice-versa.

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	Our sample consisted of up to n=38,851 individuals of European ancestry. We additionally included three generalization samples of African-Americans (up to n=769), and two generalization samples of Asians (n=341). Participants' age ranged from 9 to 98 years and the percentage of females ranged between 0 and 73%.
Recruitment	The present effort included 53 study samples from the Cohorts of Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium, the Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) consortium, and the United Kingdom Biobank (UKBB). The CHARGE consortium is a collaboration of predominantly population-based cohort studies investigating the genomics of age-related complex diseases, including those of the brain. The ENIGMA consortium brings together various studies, approximately 75% of which are population-based, with the remainder using case-control designs for various neuropsychiatric or neurodegenerative diseases. The UKBB is a large-scale prospective epidemiological study of individuals aged 40-69 years from the United Kingdom, established to investigate the genetic and non-genetic determinants of middle and old age diseases.
Ethics oversight	The institutional review boards of Boston University and the University of Southern California, as well as the local ethics board of Erasmus University Medical Center approved this study.

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	GWAS of brain MRI-based subcortical structures.
Design specifications	Each study contributed MRI data obtained using diverse scanners, field strengths, and acquisition protocols as described in Supplementary Table S5.
Behavioral performance measures	No behavioral performance was assessed, only structural MRI.

Acquisition

Imaging type(s)	Structural MRI
Field strength	1T to 4T
Sequence & imaging parameters	Varied per study, specified in Supplementary Table S5. All studies had T1-weighted sequences
Area of acquisition	Whole brain scan, including brainstem

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software	Freesurfer, SPM, FSL-FIRST, MNI, in-house imaging software, MIPAV (Supplementary Table S5)
Normalization	Normalized as in specific approaches detailed above
Normalization template	Varied across cohorts based on software described above
Noise and artifact removal	Scans where volumes could not be assessed due to artifacts were excluded
Volume censoring	Varied across cohorts based on software described above

Statistical modeling & inference

Model type and settings	Fixed-effects meta-analysis assuming additive genetic model
Effect(s) tested	Associations of SNPs across the whole genome with MRI-based subcortical brain volumes

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

Statistic type for inference (See Eklund et al. 2016)	Association of genotype allele dosages with subcortical brain volumes
Correction	We applied Bonferroni correction for testing of multiple SNPs and MRI traits

Models & analysis

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