nature portfolio

Corresponding author(s): David Fardo
Last updated by author(s): 5/1/2024

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	No software was used .
Data analysis	bcftools 1.10.2 KING 2.2.7 MAGMA 1.10 METAL 2011-03-25 Minimac 4 PLINK 1.9 and 2.0 Python 3.8.16 and 3.10.8 R 4.2.1 and 4.2.2 R package coloc 5.2.2 R package data.table 1.14.10 R package GENESIS 2.26.0 R package ggplot2 3.4.2 R package GRAB 0.1.1 R package GWASTools 1.42.1 R package LDlinkR 1.2.3 R package MASS 7.3-60 R package mediation 4.5.0 R package ordinal 2023.12-04 R package pheatmap 1.012

R package psych 2.3.3
R package POLMM 0.2.3
R package SAIGE 1.1.3
R package SNPRelate 1.30.1
R package stringi 1.8.3
samtools 1.10
TOPMed Imputation Server 1.7.3

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

Code used for data preparation and analysis is available at https://zenodo.org/doi/10.5281/zenodo.11089995.
Meta-analysis summary statistics for each neuropathology endophenotype studied will be made available through NIAGADS upon publication at https://www.niagads.org/.
ROSMAP data can be requested at https://www.radc.rush.edu and downloaded from https://www.synapse.org by approved users.
ADGC data is can be requested from NIAGADS at https://www.niagads.org/resources/related-projects/alzheimers-disease-genetics-consortium-adgc-collection.
NACC neuropathology data can be requested at https://naccdata.org/.
ACT data can be requested at https://actagingresearch.org/.
The results published here are in whole or in part based on data obtained from the AD Knowledge Portal.
GTEx QTL data used is publicly available at https://www.gtexportal.org/home/datasets.
ROSMAP QTL data used is publicly available for download at http://mostafavilab.stat.ubc.ca/xqtl.
1000 Genomes Phase 3 data is available at https://www.internationalgenome.org/home.
Raw long-read RNAseq data generated and utilized in this manuscript are publicly available in both Synapse (https://www.synapse.org/#!Synapse:syn52047893) and NIH SRA (accession number: SRP456327; https://trace.ncbi.nlm.nih.gov/Traces/?view=study&acc=SRP456327). Processed long-read RNAseq data can be easily downloaded or viewed at https://ebbertlab.com/brain_rna_isoform_seq.html.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Biological sex was used as a covariate in all analyses.
Reporting on race, ethnicity, or other socially relevant groupings	Due to limited sample sizes of participants with substantial non-European descent, we excluded these participants from our study. This was done by merging each of our data sources individually with 1000 Genomes Phase 3 data, performing principal component (PC) analysis, and excluding participants whose positions on a plot of PC1 and PC2 were beyond a distance, determined via visual inspection, from the 1000 Genomes EUR superpopulation centroid.
Population characteristics	We used multiple independent data sets of participants in this study. We adjusted each genetic association analysis for principal components, data source/study, genotyping cohort, sex, and age at death. Demographic details of participants can be found in Table 1. In summary, 53% of participants were female, the mean age of death was 83 years with a SD of 10 years, 48% of participants had at least one APOE e4 allele.
Recruitment	Participants in NACC were recruited at individual Alzheimer's Disease Research Centers, each with its own recruitment criteria. Across different data sources used in this study, participants were recruited from clinics, religious organizations, communities, and hospitals. Participants in NACC have higher prevalence of ddementia and APOE e4 than the population at large. ROS recruited priests, brothers, and sisters in the Catholic church who have higher average educational attainment than the general population aged 80 years or older. MAP recruits from community members in the greater Chicago area. ACT recruited residents in the greater Seattle area aged 65 years and older without dementia at time of enrollment.
Ethics oversight	Each study was approved by the respective institutional review boards as appropriate. The IRB of the lead authors on this study deemed that this study did not include human subjects because all research participants were deceased.

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	Raw data used in this study were collected by the invidual studies: neuropathological data: NACC, ACT, ROSMAP; genotype data: ADGC, ROSMAP; RNA-Seq data: ROSMAP; DNA methylation data: ROSMAP. Sample sizes were not pre-determined; all available samples with relevant data that passed quality control and inclusion criteria (see Methods) were included in analyses. The total sample size available for GWAS after QC measures was N=7,463, though sample sizes in individual analyses were smaller based on available phenotype data.
Data exclusions	We excluded samples and variants based on standard quality control procedures for GWAS (Samples: heterozygosity, missingness, and population outliers. Variants: minor allele frequency, missingness, Hardy-Weinberg equilibirium, duplicated variants.) We also excluded participants with rare neuropathological phenotypes (e.g. brain malignancy and traumatic brain injury) that may affect interpretation of neuropathological outcomes used. For RNA-Seq and DNA methylation analyses, we used pre-QCed data from which samples with low RNA integrity scores or bisulfate conversion efficiency had been removed. Complete details of our quality control procedures are available in the Methods and Supplementary Note sections of the manuscript.
Replication	Cerebral amyloid angiopathy (CAA) had previously been studied using an independent cohort from the Mayo Clinical Brain Bank cohort with phenotype, covariate, and genotype data made available on synpapse.org. We used 815 available participants to successfully replicate the associated between rs7247551 and CAA while adjusting for APOE epsilon diplotype. We were unable to find suitable replication data sets for other associations because either phenotypes had not been previously studied with GWAS (LATE-NC) or had been studied using a subset of the participants in our study.
Randomization	Each study had different procedures for choosing participants for genotyping, which may have been based on relevant clinical diagnoses if not neuropathological endophenotypes. We adjusted for study and genotyping cohort in all relevant analyses.
Blinding	Genotyping was performed without knowing the neuropathological status of individuals. The analyses were not blinded to the status of individuals because harmonization and quality control procedures required knowing neuropathological status.

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
☒	☐ Plants

Methods

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