nature portfolio

Corresponding author(s): John F. Crary, MD-PHD
Adam Naj, PHD

Last updated by author(s): Aug 6th, 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	Genotyping: Illumina Human660W, Illumina OmniExpress 2.5, and Illumina Global Screening Array
Data analysis	TTOPMED Imputation Server, Eagle v2.4.1 were used for Imputation. PLINK v1.90, Trimmomatic v0.36, STAR v2.7a, FASTQC v0.11.8, featureCounts v1.3.1, Picard v2.20.0, RSEM v1.3.1, AlphaEaseFC software v4.0.1, SNPTEST v2.5.6, EIGENSTRAT: v6.1.4, COLOC v5.2.3, SparkINFERNO v1.0.0 were used for association and statistical analyses. LiftOver was used for lifting over chromosomal positions. FineMapping was performed using FINEMAP and PolyFun + FINEMAP. LD correlation matrices were acquired from UK Biobank reference panel. RAPiD-nf pipeline was used to process the RNA-Seq data and differential gene expression analysis was conducted using Limma package. R v4.0 was used for data transformations, plotting of the results and statistical analyses including packages ggplot2 v3.4.2, ggh4x v0.2.8, ggpubr v0.6.0, corrplot v0.92, ggcorrplot v0.1.4.1 and data.table v1.15.4. Image analysis was performed with NeuronJ v1.4.3 package. All other code parameters are available at https://github.com/jackhump/PSP_GWAS and are permanently referenced with DOI: https://doi.org/10.5281/zenodo.12668541

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

The genotype summary statistics in this study have been deposited in the NIAGADS database under accession code NG00169 [https://dss.niagads.org/open-access-data-portal/#NG00169]. GWAS summary statistics P values only are open access and available to download here: https://dss.niagads.org/open-access-data-portal/#NG00169. Access to full summary statistics is controlled due to the risk of identifiable information, please fill out this application for access: https://dss.niagads.org/datasets/ng00067/ng00169/. Please allow two weeks for a response to the request. The genotype raw data are available under restricted access as the data contains identifiable information, but can be obtained by emailing adamnaj@pennmedicine.upenn.edu and kurt.farrell@mssm.edu. Please allow four weeks for a response to the request. Data is available for general research use according to the following requirements for data access and data attribution: https://www.niagads.org/data/request/data-request-instructions. We anticipate the individual level genotypes will be available on NIAGADS under restricted access in 6-12 months. The additional data generated in this study are provided in the Supplementary Information/Source Data file.

The publicly available data used here can be found in the following repositories:
GTEx web portal, https://gtexportal.org/home/datasets
eQTL single cell data, https://zenodo.org/record/5543735
AD GWAS summary statistics, https://www.niagads.org/datasets/ng00075
PD GWAS summary statistics, https://drive.google.com/drive/folders/10bGj6HfAXgl-JslpI9ZJIL_JIgZyktxn
Mayo Clinic RNAseq Study, https://adknowledgeportal.synapse.org/Explore/Studies/DetailsPage/StudyDetails?Study=syn5550404
Whole blood microarray data, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE140830
Brain PLAC-seq, https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001373.v2.p2
Picard https://github.com/broadinstitute/picard/releases
C4 imputation panel, https://github.com/freeseek/imputec4
1000 Genomes reference panel, https://www.internationalgenome.org

All other data supporting the findings described in this manuscript are available in the article and its Supplementary Information files. Please see legends for these files for details.

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	The cohort includes 8,703 total (4850 female sex, 3853 male sex). Sex was determined based on XY chromosome. Sex was included as a covariate in GWAS.
Reporting on race, ethnicity, or other socially relevant groupings	All tissue, included in the study, is from individuals of European descent.
Population characteristics	The mean age of participants at death was 66.5 in cases and 72.8 in controls.
Recruitment	We did not recruit any new subjects for this study; instead, we utilized existing specimens from established brain banks at various institutions.
Ethics oversight	Tissue was obtained from donors who had provided consent for research use either directly or via their next of kin. Research with de-identified autopsy material does not meet the federal regulatory definition of human subject research as defined in 45 CFR part 46 and is otherwise exempt. However, HIPAA requirements still apply. Thus, all material was de-identified. For the living subjects, the study was reviewed and approved by the institutional board (IRB#11-001142) at University of California, Los Angeles.

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	For genotyping and quality control, we included 2,595 cases diagnosed with PSP and 5,584 control samples. Westernblot samples included 7 controls and 6 PSP cases. Immunohistochemistry image analysis included 10 controls and 10 PSP cases. These sample sizes were chosen based on the availability of specimens in brain banks. The number of obtained is deemed sufficient as we detected significant differences in genotyping, gene expression, and immunohistochemistry analyses related to PSP pathology.
Data exclusions	Samples were excluded if they significantly deviated from the European subcluster, had discordant sex (i.e., genotyped sex did not match reported sex), exhibited genotyping failure of more than 10%, or showed high relatedness (Pi-Hat > 0.4). Additionally, samples were filtered by TOPMed imputation based on a quality threshold (R2 ≥ 0.8) and excluded if they had a minor allele frequency (MAF) greater than 0.01.
Replication	Our focus was on creating the largest genome-wide association study (GWAS) of progressive supranuclear palsy (PSP) to date. Despite the lack of a replication GWAS, we demonstrated biological validation through protein-level analysis in human post-mortem brain tissue by staining for C4A and showing its elevated presence in the blood of those with PSP. Imputed C4A/C4B copy numbers were successfully replicated using dPCR. Differential gene expression analysis, immunohistochemistry, and biochemical analysis were conducted to validate genetic and expression findings at the molecular level.
Randomization	Not relevant to the study, as no samples were allocated into experimental groups.
Blinding	Experimenters were not blind to the genotype, as ll functional experiments were performed by one experimentalist.

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

Antibodies

Antibodies used	Olig2 Rabbit Monoclonal (EP112) PAb Recombinant Anti-C4a Rabbit monoclonal [EPR10143] AT8 phospho-Tau Mouse monoclonal(Ser 202, Thr205)
Validation	Olig2 Rabbit Monoclonal (EP112) PAb: https://www.cellmarque.com/antibodies/EP/2682/OLIG2_EP112 Recombinant Anti-C4a Rabbit monoclonal [EPR10143]: https://www.abcam.com/products/primary-antibodies/c4a-antibody-epr10143-ab170942.html AT8 phospho-Tau Mouse monoclonal(Ser 202, Thr205): https://www.thermofisher.com/antibody/product/Phospho-Tau-Ser202-Thr205-Antibody-clone-AT8-Monoclonal/MN1020

Plants

Seed stocks	NA
Novel plant genotypes	NA
Authentication	NA

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	Refer to https://molecularneurodegeneration.biomedcentral.com/articles/10.1186/s13024-018-0270-8#Sec15
Files in database submission	Refer to https://molecularneurodegeneration.biomedcentral.com/articles/10.1186/s13024-018-0270-8#Sec15
Genome browser session (e.g. UCSC)	Refer to https://molecularneurodegeneration.biomedcentral.com/articles/10.1186/s13024-018-0270-8#Sec15

Methodology

Replicates	Refer to https://molecularneurodegeneration.biomedcentral.com/articles/10.1186/s13024-018-0270-8#Sec15
Sequencing depth	Refer to https://molecularneurodegeneration.biomedcentral.com/articles/10.1186/s13024-018-0270-8#Sec15
Antibodies	Refer to https://molecularneurodegeneration.biomedcentral.com/articles/10.1186/s13024-018-0270-8#Sec15
Peak calling parameters	Refer to https://molecularneurodegeneration.biomedcentral.com/articles/10.1186/s13024-018-0270-8#Sec15
Data quality	Refer to https://molecularneurodegeneration.biomedcentral.com/articles/10.1186/s13024-018-0270-8#Sec15
Software	Refer to https://molecularneurodegeneration.biomedcentral.com/articles/10.1186/s13024-018-0270-8#Sec15