nature portfolio

Corresponding author(s): Gamze Gürsoy

Last updated by author(s): 07/19/2024

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 used for data collection
Data analysis	Code for the PrecisionChain framework is available at https://github.com/G2Lab/PrecisionChain and https://doi.org/10.5281/zenodo.10067135. MultiChain v2.3.3 Community (GPLv3) was used to store data in blockchain infrastructure. Python v3.6.13 was used for data pre-processing , insertion and querying BCFTOOLS v1.9 was used for VCF data pre-processing PLINK version v1.90 was used for plaintext GWAS analysis

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 individual level data from NYGC ALS Consortium are available to authorized investigators through dbGaP via accession code phs003067 (being processed by dbGaP as of 07/19/2024). Data dictionaries and variable summaries are available on the dbGaP public FTP site: https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs003067/phs003067.v1.p1 Public summary-level phenotype data may be browsed at the dbGaP study report page: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs003067.v1.p1
Subject Sample Telemetry Report (SSTR) for Subject and Sample IDs, consents, summary counts, processing status, and molecular and sequence sample uses is available on the SSTR site: https://www.ncbi.nlm.nih.gov/gap/sstr/report/phs003067.v1.p1
The data are available under General Research and Health, Medical, Biomedical data use limitations.

GTAC data can be accessed via dbGaP accession code phs002973.v1.p1.

Genetic data for the publicly-accessible blockchain network are shared via the International Genome Sample Resource (1000 Genomes Project) and can be accessed through https://www.internationalgenome.org/data-portal/data-collection/30x-grch38.

UK Biobank data is available to authorized investigators through UKBiobank portal.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	~1964 (~42%) of 4734 samples used from NYGC ALS consortium data is female.
Population characteristics	We limited our analysis to patients with European genetic ancestry (>80%), known reported sex and known site of onset. This leaves 2,903 patients, 802 with bulbar and 2,101 with limb ALS. Average age of symptom onset for these patients are ~58. More information can be found at https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs003067.v1.p1
Recruitment	Samples were recruited from a range of hospitals and academic medical centers worldwide in the USA and Europe.
Ethics oversight	The NYGC ALS Consortium samples presented in this work were acquired through various institutional review board (IRB) protocols from member sites and the Target ALS postmortem tissue core and transferred to the NYGC in accordance with all applicable foreign, domestic, federal, state, and local laws and regulations for processing, sequencing, and analysis. The Biomedical Research Alliance of New York (BRANY) IRB serves as the central ethics oversight body for NYGC ALS Consortium. Ethical approval was given and is effective through 07/27/2024. Note that BRANY IRB determined that the request for waiver of informed consent satisfies the waiver criteria set forth in 45 CFR 46.116(d).

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	No sample size calculation was used. We limited our analysis to patients with European genetic ancestry (>80%), known reported sex and known site of onset. This leaves 2,903 patients, 802 with bulbar and 2,101 with limb ALS. Since our phenotype is site of onset, we are limited to this sample size with the available data. However, we think this is acceptable as sample sizes of >1000 are shown to be typically sufficient to power a genotype-phenotype association study for 500k- 1 million SNPs (Hong et al. 2012). Thus, we believe the study is sufficiently powered.
Data exclusions	Exclusion criteria were any of the following: We limited our analysis to patients with European genetic ancestry (>80%), known gender and

Data exclusions	known site of onset. This leaves 2,903 patients, 802 with bulbar and 2,101 with limb ALS. The first is essential as we wanted to make sure population does not confound our analysis and majority of NYGC ALS patients are of European ancestry. The second is needed as we use site of onset as the phenotype.
Replication	ALS GWAS findings were replicated once in a publicly available GTAC dataset (dbGaP accession code phs002973.v1.p1)
Randomization	The program was a resource generation program so no patient randomization was appropriate.
Blinding	Blinding is not relevant in this study as we are not assessing outcome of an intervention on separate groups. Our study measures only existing associations in patients' genetic variants and site of onset. Both features are fixed at time of analysis.

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

Methods

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