nature portfolio

Corresponding author(s): Eva L. Feldman, Maureen A. Sartor

Last updated by author(s): 2025/09/12

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

All sequencing files (BAM format) have been deposited into the Database of Genotypes and Phenotypes (dbGAP, accession number pending) and the European Genome-Phenome Archive (EGA) with accession number EGAS50000001019.

All software and code are freely available to the public and accessible via github. The following links provide unrestricted access:

The main github repository containing the code and a minimum subset of data to reproduce the figures, the results, and the supplementary figures and tables: https://github.com/yzhao80/ALS_RNA

The external Grima et al. test dataset was downloaded from NCBI Gene Expression Omnibus (GSE234297).
The external van Rheenen et al/Swindell et al. datasets were downloaded from NCBI Gene Expression Omnibus (GSE112676, GSE112680).

The list of differential expressed genes (DEG) from 2023 Grima et al was downloaded from https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Fnan.12943&file=nan12943-sup-0002-supplementary_tables.xlsx

The list of DEG from 2019 Swindell et al was downloaded from https://static-content.springer.com/esm/art%3A10.1186%2Fs12967-019-1909-0/MediaObjects/12967_2019_1909_MOESM10_ESM.xlsx

The list of DEG from iPSC derived neuron was downloaded from https://github.com/frattalab/unc13a_cryptic_splicing/tree/main/data

The list of DEG from post postmortem spinal cord tissue was downloaded from https://static-content.springer.com/esm/art%3A10.1038%

2Fs41467-023-37630-6/MediaObjects/41467_2023_37630_MOESM6_ESM.xlsx (Fig 4A)

Data analysis

The main github repository containing the code and a minimum subset of data to reproduce the figures, the results, and the supplementary figures and tables: https://github.com/yzhao80/ALS_RNA

RNA-seq data preprocessing
Raw RNA-seq reads (FASTQ files) were trimmed in Cutadapt v2.3, aligned to the human reference genome hg38 in STAR 2.5.3a, and quantified in featureCounts v2.0.3 (using "-s 2" for reverse strand reads). Rigorous RNA-seq quality control was implemented by FastQC v0.11.9, RSeQC v5.01., and FastQ Screen v0.15.2, and results were summarized with MultiQC v1.7.

Gene expression
Cell type deconvolution using DNAm was conducted using the FlowSorted.Blood.EPIC v 1.4.1 R package. Cell type deconvolution using RNA-seq was conducted in BayesPrism v2.1.2. Differential gene expression analysis was conducted using the limma R package (v3.48.3; functions: voom, lmFit, eBayes).

ALS case prediction
Analyses were run in Python 3.10.13 with the following package versions: kneed 0.8.5 for knee-point detection, xgboost 2.1.1 for model training, scikit-learn 1.3.0 for feature selection and performance metrics, matplotlib 3.8.0 for plotting, numpy 1.26.3 and pandas 2.1.4 for data manipulation, and pyreadr 0.5.0 for importing R data objects.

ALS survival prediction
WGCNA v1.73, MASS v7.3.60.2 and xgboost v0.1.0 R packages were used for feature selection in predicting survival. The R packages survival v 3.7.0, timeROC v0.4, and survminer v0.4.9 were used for visualization in survival predictions. The R package compareC v1.3.2 was used for comparing C-index of different CoxPH models.

Pathway and GO terms analysis
R packages clusterProfiler v4.12.2, org.Hs.eg.db v3.19.1, and org.Hs.eg.db v3.19.1 were used for pathway analysis. EnrichmentMap v3.5.0 in Cytoscape v3.10.0 was used for GO terms visualization.

Drug perturbation analysis
R packages cmapR v1.21.0, igraph v2.1.4 , and GGally v2.2.1 were used for drug perturbation analysis and network visualization.

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 RNA-seq data generated in this study have been deposited in the European Genome-Phenome Archive (EGA) database under accession code EGAS50000001019 (https://ega-archive.org/studies/EGAS50000001019) and Database of Genotypes and Phenotypes (dbGaP) database under accession code phs004288.v1.p1 (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs004288.v1.p1). The RNA-seq data and associated clinical/demographical data are available under restricted access for ALS disease-specific research, and access can be obtained by submitting requests through EGA or dbGaP. Requestors must provide documentation of local IRB approval, agree to make study results that used the data available to the larger scientific community, and provide a letter of collaboration with the primary study investigators. Requests are limited to not-for-profit organizations. The processed RNA-seq count data are available at dbGaP. Source data are provided with this paper.

The main github repository containing the code and a minimum subset of data to reproduce the figures, the results, and the supplementary figures and tables: https://github.com/yzhao80/ALS_RNA

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	Only the biological variable sex was employed. Sex was considered in the study design due to sex-differences in immune cell levels and activation state in ALS. Sex was ascertained in sequencing data. Findings were reported overall (i.e., in both sexes) or stratified by sex (i.e., in males or females), as indicated in the text and figure legends. Number of participants by sex outlined in the main text Tables 1 and 2.
Reporting on race, ethnicity, or other socially relevant groupings	Race was reported in the main text Tables 1 and 2 as either White, Black, or Other.
Population characteristics	Population characteristics that were collected are outlined in the "Participants" subsection of the "Methods". Results are outlined in the main text Tables 1 and 2 and Supplementary Table S3.
Recruitment	ALS participants were recruited during their clinical visit at the University of Michigan Pranger ALS Clinic between 2011 and

2021. ALS participants were over 18 years of age and met the Gold Coast definition of ALS. Control participants without ALS and without first- or second-degree relatives with ALS were identified and recruited via a University of Michigan research interest database, random address direct mailings, and Meta/Facebook advertisements, and were compensated USD $50 via a check in the mail. ALS participants did not receive monetary compensation.

Ethics oversight

This study was conducted according to all relevant ethical regulations; all participants were over 18 years of age and provided informed consent in English. The University of Michigan Institutional Review Board (IRBMED HUM28826) granted ethical approval for this research.

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	The sample size was determined based on budget constraints and available samples. The sample size is the largest compared to all previously published studies of ALS participant blood by RNA-seq.
Data exclusions	Expression of the X chromosome gene XIST and 49 male-specific Y chromosome genes was analyzed to compare observed sex to recorded sex, filtering out five samples with discrepancies, which were removed. Otherwise, all samples were included.
Replication	No replication was undertaken; however, select differentially expressed genes between ALS and control were validated by qPCR in an independent cohort (n=29 ALS cases, n=27 controls). The qPCR was not adjusted for covariates. Case-control and survival prediction models were validated in an independent external ALS case-control blood RNA-seq dataset.
Randomization	Randomization was not a feature of the study design.
Blinding	Sequencing of the samples was performed in a blinded manner.

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

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*