nature portfolio

Corresponding author(s): Mahdi Moqri
Vadim N. Gladyshev

Last updated by author(s): Oct 20, 2025

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	Data were collected from publicly available sources including Gene Expression Omnibus (GEO), NHLBI TOPMed MESA (phs001416.v3.p1), Parkinson's Progression Markers Initiative (PPMI via LONI IDA), Generation Scotland (GS via application: https://genscot.ed.ac.uk/for-researchers/access). Data from the Mass General Brigham Biobank (MGB) were generated for this study (DNAm) or obtained internally (MGB4K DNAm and mortality). Some original datasets (RA, Grady, GENOA, JenAge, GC6, 500FG) were accessed via GEO or specific repositories as listed in Table 1. New MGB500 DNAm data generated in this study is available on GEO (GSE246337).
Data analysis	All analyses were performed in Python 3.10 using the Biolearn Python library (version 0.3.0, https://github.com/Bio-Learn/biolearn), which incorporates statsmodels (version 0.13.5) and seaborn (version 0.12.2) for statistical analysis and visualization. R (version 4.3.3) was used with packages including SeSAMe (version 1.22.1) for methylation preprocessing, clusterProfiler and enrichplot for functional enrichment analysis, and DOSE for disease gene network analysis. Cell composition estimation used software provided by the Clock Foundation.

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 new MGB500 DNA methylation dataset generated for this study is available on NCBI GEO under accession number GSE246337. Previously published DNA methylation and RNA-seq data can be accessed through the Gene Expression Omnibus (GEO) database (accession numbers listed in Table 1: e.g., GSE42861, GSE132203, GSE157131, GSE103232, GSE75337, GSE94438, GSE134080). MESA data are available through dbGaP (phs001416.v3.p1). PPMI data are available through LONI IDA (https://ida.loni.usc.edu/). Generation Scotland data are available via application (https://genscot.ed.ac.uk/for-researchers/access). The Biolearn library, which includes data loading functions, is freely available at https://Bio-Learn.github.io. MGB4K data access requires specific agreements.

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	Term sex (male/female) is used.
Reporting on race, ethnicity, or other socially relevant groupings	This study utilized datasets from diverse populations, including participants identified as White, African American, Hispanic, and Asian (e.g., MESA). The newly generated MGB500 cohort is generally representative of the racial/ethnic distribution of the Boston area (Fig. 3a). Socioeconomic status and educational background information varied by the original dataset.
Population characteristics	The study included diverse populations from various publicly available and newly generated datasets. Detailed characteristics of each dataset are provided in Table 1 of the manuscript, including sample sizes, age ranges, and sex distributions where available.
Recruitment	Participants were not directly recruited for this study, except for the establishment of the Mass General Brigham (MGB) Biobank from which samples were drawn. Most data were obtained from existing public or controlled-access datasets. Details of recruitment for each original study can be found in their respective publications.
Ethics oversight	This study involved secondary analysis of existing datasets and generation of new data from banked samples. Ethical approval for the original data collection was obtained by the respective study teams. Our use of these data was in accordance with the terms of use for each dataset. The generation and analysis of new methylation data from the MGB Biobank was approved by the Mass General Brigham IRB (protocol 2021P003059).

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	Sample sizes for each dataset are provided in Table 1 of the manuscript. These were predetermined by the original studies from which the data were obtained or by the selection criteria for the new MGB500 cohort.
Data exclusions	Diseased participants from the GC6 cohort were excluded. Standard quality control filtering was applied to methylation data (samples and CpGs with >10% missing values were excluded).
Replication	The analyses were performed on multiple independent datasets to ensure reproducibility of findings across different populations and study designs. Findings are primarily based on blood samples. Methylation data was generated using array technology (EPIC v1/v2), not whole-genome sequencing.
Randomization	Randomization was not relevant to this study as it involved analysis of existing observational datasets or banked samples.
Blinding	Blinding was not relevant to this study as it involved computational analysis of existing datasets.

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	Not applicable. This study did not involve a clinical trial.
Study protocol	Not applicable. This study involved analysis of existing datasets and banked samples.
Data collection	Data were collected from publicly available sources (e.g., GEO, dbGaP, LONI IDA) or obtained through application (e.g., Generation Scotland) or internal biobanks (MGB). Original data collection periods vary by dataset and are detailed in the respective original publications. The new MGB500 methylation data was generated from samples collected prospectively by the MGB Biobank.
Outcomes	The study analyzed molecular outcomes (DNA methylation levels, gene expression levels) and their association with chronological age. It also assessed the association of identified "multi-omic aging genes" (specifically CpG methylation levels) with all-cause mortality and performed enrichment analysis for associations with age-related diseases.

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.*