nature portfolio

Corresponding author(s): Amos Tanay

Last updated by author(s): Oct 15, 2023

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 for data collection.
Data analysis	Analysis was done using the R version 4.2.1 platform with the following R packages used: xgboost 1.2.01, cmprsk 2.2-11, umap 0.2.7.0, tglkmeans 0.3.4, BigSnpR 1.9.11, bigstatr 1.5.6, SusieR 0.11.92. Custom code is available in github repository https://github.com/tanaylab/Mendelson_et_al_2023

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

UKBB data is available to approved researchers via The UK Biobank Research Analysis Platform (https://www.ukbiobank.ac.uk/enable-your-research/research-analysis-platform). Longevity GWAS results will be made availableare available at https://gwasresults.s3.ap-south-1.amazonaws.com/

gwas_longevity_age_sex_covar_extended.tsv.gz. National Health and Nutrition Examination Survey (NHANES) data can be accessed from https://www.cdc.gov/nchs/nhanes/index.htm.

Access to the CHSDB data used for this study may be made available upon request, at CHS's discretion, subject to an internal review by Amos Tanay to ensure that participant privacy is protected, and subject further to completion of a data sharing agreement, approval from the institutional review board of CHS and institutional guidelines and in accordance with the current data sharing guidelines of CHS and Israeli law. Subject to receipt of the aforementioned CHS consent and subsequent approvals, data sharing will be made in a secure setting, on a per-case-specific manner, solely for the purpose of reproducing the analysis carried in the research paper, as will be defined by the chief information security officer of CHS. Please submit such requests to Amos Tanay.

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 stratified in all analysis and reported accordingly.
Reporting on race, ethnicity, or other socially relevant groupings	Race and socio-economic status were considered and reported where applicable.
Population characteristics	Medical histories of 4.57M individuals from (CHSDB), providing a total of 29.5M patient-years between ages 30-85, with a median tracking duration of 16.6 years. Demographic characteristics are described in extended data figure 1. Briefly, a total of 3.73M patients age 30-80 were included in CHSDB modeling in 5-year intervals. Mean (SD) size of each patient age group is 339518 (70120), with 44.2% (55.8%) male (female).
Recruitment	Clalit EHR, UK biobank and NHANES recruitment policies are well documented
Ethics oversight	CHS institutional review board approved this study and it was deemed exempt from the requirement for informed consent (Reference 0158-16-COM2)

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	Due to the nature of the retrospective study, no statistical methods were used to pre-determine sample sizes. Data were collected from electronic health records. No data were excluded from the analyses.
Data exclusions	None
Replication	Models were cross validated on Clalit EHR (5-fold cross validation) and then in UK Biobank and NHANES
Randomization	Cross validation group association was done while controlling for sex and outcome
Blinding	Due to the nature of this retrospective study, blinding was not applicable.

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