nature portfolio

Corresponding author(s): Alexander Bick

Last updated by author(s): Apr 6, 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 was used.
Data analysis	Polygenic Risk Scoring: Scores were calculated using the PRScs, PRScsx, and/or PLINK tools in R 4.1.2 Linear regression analyses: R 4.1.2 was used for analysis. R packages used for analysis: ggplot2, data.table, tidyverse, dplyr GSEA analyses: Analyses were performed using the fgsea and msigdbr packages.

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

Whole-genome sequencing data for individuals in the TOPMed cohort and the CHIP variant call sets are available through restricted access via the dpGaP TOPMed

Exchange Area available to TOPMed investigators in order to protect patient privacy. Individuals who want to apply to access TOPMed data should follow the steps listed here (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/about.html#request-controlled). Summary statistics for the measures of DNA methylation aging are available at https://datashare.ed.ac.uk/handle/10283/3645 , and summary statistics for the four significant proteins are available at www.omicspred.org

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	Sex was considered as a covariate in all analyses in this study.
Reporting on race, ethnicity, or other socially relevant groupings	Self-reported race was not used in this study, but the first 10 principal components of ancestry were used as covariates in the analyses.
Population characteristics	As part of 51 studies that contributed to the Freeze 8 TOPMed program, 127,946 individual samples underwent whole genome sequencing (WGS). All included individuals provided informed consent to participate. For the 82,807 samples with age data available, mean age was 52.5, and the samples come from diverse ancestral backgrounds, 40% European, 32% African, 16% Hispanic, 10% Asian. For the subset of individuals from the TOPMed cohort with CHIP, 61% were female, and the average age of the cohort was 68 years. 42% of the cohort reported ever smoking.
Recruitment	In TOPMed, 49 cohorts were included, each with differing sample recruitment protocols. More information can be found at the dbGaP page for each of the constituent cohorts.
Ethics oversight	Each of the 49 cohorts obtained IRB permission from their respective institutions and are supported by the NHLBI. Each of the studies obtained informed consent from each participant. The study design was approved by the Vanderbilt Institutional Review Board (IRB#210270)

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	4,370 participants with CHIP were included based on convenience sampling.
Data exclusions	No data was excluded.
Replication	The experimental data had sufficiently high n to perform statistical comparisons, but was not replicated in additional cohorts.
Randomization	Sample recruitment included no randomization mechanism, but relevant covariates are controlled for in all statistical analyses.
Blinding	Investigators were blinded with respect to generation of whole genome sequencing data from the TOPMed participants.

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