nature research

Corresponding author(s): Robert E. Gerszten

Last updated by author(s): 3/24/2021

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees 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 in this study.
Data analysis	Data analyses and machine learning were performed using open source packages from R 3.6.2 (R Core Team, R Foundation for Statistical Computing, Vienna, Austria). Gene set enrichment analysis (GSEA v4.0.3) was performed using the Molecular Signatures Database canonical pathways collection (MSigDB v7.0, http://software.broadinstitute.org/gsea/msigdb/collections.jsp). GSEA results were exported to Cytoscape (v3.7.1) for visualization with the Enrichment Map tool (v3.3.1). Predictive modeling was performed using: MMPC algorithm: MXM_1.4.5 (R library); global optimization algorithm: pso_1.0.3 (R library); k-nearest neighbor algorithm to impute missing values: impute_1.60.0 (R package); data preprocessing: caret_6.0-85 (R package); and elastic net: glmnet_3.0-2 (R package).

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

De-identified, individual level proteomics and phenotypic data that support the HERITAGE findings within this paper are included in the Data Supplement Table 1. Overlapping aptamer-based and antibody-based proteomics data on the HERITAGE sample are included in Supplementary Data Table 1. GWAS summary statistics in FHS and JHS are available through restricted access via the database of Genotypes and Phenotypes (dbGaP), a publicly available resource developed to archive data from human studies of genotype-phenotype relationship and can be accessed here (https://www.ncbi.nlm.nih.gov/gap/; FHS accession number: phs000363.v19.p13; JHS accession number: phs000964). FHS proteomics data have also been deposited in dbGaP and are available through the same accession

number. JHS proteomics data have been deposited in the JHS Data Coordinating Center and are being deposited in dbGaP; pending its receipt in dbGaP, all JHS data are available from the JHS Data Coordinating Center on request (JHSccdc@umc.edu). In addition, proteogenetics findings (precise SNP IDs) included in the Supplemental Table 15 from FHS/MDCS meta-analysis and JHS have been provided in Supplementary Data Tables 2 and 3, respectively.

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 sizing was not performed for these analyses. All HERITAGE study participants with complete clinical data were used for baseline analyses (N = 745) and longitudinal analyses (N = 654)
Data exclusions	Individuals with baseline VO2max measures who were missing post-training VO2max measures (N = 91) were excluded from longitudinal analyses. The clinical characteristics of individuals with baseline VO2max and those with complete longitudinal data are described in Table 1.
Replication	We derived protein-baseline VO2max relationships in the HERITAGE offspring and replicated our findings in the parents subgroup. We subsequently provided an external chronic exercise study to replicate our baseline VO2max analyses. The reproducibility of VO2max measures in the HERITAGE population has been described and is referenced in the manuscript. We validated aptamer specificity of our top findings using antibody-based assays. Quality control measures for proteomics data are also included in the manuscript.
Randomization	Randomization is not applicable for these secondary analyses of a single-arm exercise study. We adjusted for known clinical factors related to VO2max including age, sex, race, and body mass/composition. In addition we tested for these clinical factors effects on protein-VO2max relationships.
Blinding	All subjects in the HERITAGE study underwent the same standardized, exercise intervention. Proteomic analyses were performed in a blinded manner on de-identified samples and then data were returned to the HERITAGE principal investigators prior to analyses.

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
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

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

Human research participants

Policy information about studies involving human research participants

Population characteristics	The HERITAGE Family Study enrolled white and black participants from two-generation biologic families (parents and offspring; ages 17-65) from 4 clinical centers in the US and Canada. Participants were sedentary (over the past 3 months) and free from overt cardiometabolic disease. The Framingham Heart Study Offspring cohort included 5124 persons ages 5-70 who have been followed longitudinally every 4-8 years from 1971 until now. Participants who attended the 5th examination (1991-1995) and who underwent plasma proteomic profiling were including in this study. This included 821 participants from a case-control cohort (311 incident CVD cases and 588 control) who were free of prevalent CVD; and a random sample of 1014 participants. We included all available subjects with proteomics data, however both known and unknown group differences exist between cases and controls. The validation study (Ross et al.) included 300 abdominally obese adult (mean age = 51 years), sedentary white men and women. Thus, our discovery cohort (HERITAGE) findings are most applicable to sedentary but otherwise healthy individuals whereas the validation cohort findings are most relevant to abdominally obese individuals. No other selection biases exist.
Recruitment	HERITAGE participants were recruited using community based outreach and advertisements. Participants in the FHS Offspring

Recruitment	Cohort were recruited from families of the original FHS cohort. Men and women between 35-65 years old were recruited from the Kingston, Ontario region for participation in the validation study (Ross et al.)
Ethics oversight	HERITAGE, FHS, and the Validation exercise study consents were reviewed and the research performed in these analyses was approved by Beth Israel Deaconess Medical Center's institutional review board. Given these analyses involved the integration of peripheral blood sampling and clinical traits using de-identified data, we are not aware of any risks to the participants.

Note that full information on the approval of the study protocol must also be provided in the manuscript.