nature portfolio

Corresponding author(s): Michael C. Honigberg and Pradeep Natarajan

Last updated by author(s): Oct 13, 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 for data collection.
Data analysis	Data analysis was performed primarily using in-house code in R (version 4.1.0). Code used for the main analyses of this study can be accessed at https://github.com/aschuerm/ukbppp_cardiac_diseases. Protein imputation was performed using the impute.knn() function from the impute package in R (version 1.68.0). Time-to-event analyses were performed using the survival package (version 3.2-11) in R. Mendelian randomization analyses were performed using the TwoSampleMR (version 0.5.7) and MendelianRandomization (version 0.7.0) packages in R. Colocalization analyses were performed using the coloc.abf() function from the coloc package (version 5.1.0) in R. Protein-based, clinical, and combined risk scores were constructed and tested using the glmnet (version 4.1-2) and pROC (version 1.17.0.1) packages in R. Genetic instruments (used for Mendelian randomization analyses) were clumped into independent loci using PLINK (version 1.90). Figures were constructed using the ggplot2 package (version 3.4.2) in R.

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

Data supporting the results of the present study are available from the UK Biobank (https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access) to bona fide researchers with institutional review board and UK Biobank approval. These analyses were performed using the UK Biobank resource under application number 7089. The secondary use of these data was approved by the Mass General Brigham institutional review board. Pathway enrichment analyses were performed using the Gene Ontology resource via Enrichr (https://maayanlab.cloud/Enrichr/). The UK Biobank Pharma Proteomics Project was used for genetic association data for circulating proteins (i.e., protein quantitative trait locus data) through Synapse (https://doi.org/10.7303/syn51364943). FinnGen (freeze 9) was used for genetic association data for coronary artery disease (https://r9.finngen.fi/pheno/I9_CHD), heart failure (https://r9.finngen.fi/pheno/I9_HEARTFAIL), atrial fibrillation/flutter (https://r9.finngen.fi/pheno/I9_AF), and operated calcific aortic stenosis (https://r9.finngen.fi/pheno/I9_CAVS_OPERATED). The Human Protein Atlas was used for functional characterization of proteins (https://www.proteinatlas.org/). The Women's Health Initiative (WHI) was used for external validation analyses for the clinical, protein-based, and combined prediction models. Data from the Women's Health Initiative (https://www.whi.org/) can be accessed by researchers who meet the criteria for access to confidential data.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	The term "sex" was used throughout this study to indicate biological attribute. Sex was ascertained by participant self-report. Sex-stratified analyses were performed to evaluate differences in protein-disease associations between male (n=19,612) and female (n=24,701) participants in the UK Biobank. Sex was also incorporated as a covariate in statistical models where applicable.
Population characteristics	The UK Biobank is a population-based cohort of approximately 500,000 volunteers aged 40-69 years at the time of study enrolment, recruited from 22 assessment centers across the United Kingdom during 2006-2010. The UK Biobank Pharma Proteomics Project (UKB-PPP) is a project involving 13 biopharmaceutical companies that funded the profiling of the circulating proteome in a subset of approximately 55,000 UK Biobank participants. The final study sample included a total of 44,313 unrelated UKB-PPP participants without a history of coronary artery disease, heart failure, atrial fibrillation, or aortic stenosis at enrolment. The majority of participants were female (n=24,701 [55.7%]) and self-reported as white (n=41,481 [93.6%]). The mean (standard deviation) age was 56.4 (8.2) years. External validation analyses were performed in the Women's Health Initiative (WHI; a prospective study of women recruited at 40 centers across the United States from 1993 to 1998). A subset of WHI participants were invited for the Long Life Study (LLS) which consisted of a one-time in-person study visit (between March 2012 and May 2013) including a blood draw, clinical evaluation, and assessment of functional status. A total of 1,333 WHI-LLS participants underwent proteomic profiling. After excluding participants with missing values for >10% of measured proteins, missing data on time between enrolment and time of blood donation, or a history of heart disease, we included data from 1,083 WHI-LLS participants. All participants were female (n=1,083 [100%]). The majority of participants self-reported as white (n=872 [66.2%]). The mean (standard deviation) age was 79.9 (6.4) years.
Recruitment	UK Biobank participants were recruited from 22 assessment centers across the United Kingdom during 2006-2010. WHI participants were recruited from 40 centers across the United States from 1993 to 1998.
Ethics oversight	The UK Biobank was approved by the North West Multi-center Research Ethics Committee. All analyses were conducted under UK Biobank application number 7089. The Mass General Brigham Institutional Review Board approved the secondary use of these data.

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 size was determined by the data availability from the UK Biobank.
Data exclusions	UKB-PPP participants were excluded if they (1) had missing values for >10% of proteins; (2) had missing data on self-reported race/ethnicity; (3) had missing data on genetic ancestry; (4) were inferred to be related to at least one other included participant; and (5) had a history of coronary artery disease, heart failure, atrial fibrillation, or aortic stenosis at baseline. Proteins were excluded from analysis if they were missing for >10% of UKB-PPP participants.
Replication	Cinical, protein-based, and combined (i.e., clinical-proteomic) risk scores were validated externally in 1,083 participants from the Women's Health Initiative who attended the Long Life Study visit.
Randomization	Randomization was not applicable as the study design was observational and did not involve experimental groups.
Blinding	Blinding was not applicable as the study design was observational and did not involve experimental groups.

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

Methods

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