nature portfolio

Corresponding author(s): Ron Do

Last updated by author(s): Sep 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 to collect the data.
Data analysis	We used different publicly available softwares and R packages: ADMIXTURE 1.3: https://dalexander.github.io/admixture/index.html bigsnpr 1.8.1 (R package): https://cran.r-project.org/web/packages/bigsnpr/index.html GCTA 1.93.2beta (GREML-LDMS): https://yanglab.westlake.edu.cn/software/gcta/#Overview GENESIS (R package performing PC-AiR and PC-Relate): https://bioconductor.org/packages/3.18/bioc/html/GENESIS.html Scripts used for processing coronary artery snATAC data are available at https://github.com/MillerLab-CPHG/Coronary_snATAC SnpEff 4.1: https://pcingola.github.io/SnpEff/ All figures were generated using R software version 4.1.0 and later: https://www.R-project.org/

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 for each participating study can be accessed through dbGaP with the corresponding TOPMed accession numbers: Amish (phs000956), ARIC (phs001211), BioMe (phs001644), CARDIA (phs001612), CHS (phs001368), COPDGene (phs000951), DHS (phs001412), FHS (phs000974), GeneSTAR (phs001218), GENOA (phs001345), JHS (phs000964), MESA (phs001416), WHI (phs001237). Core support including centralized genomic read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1; contract HHSN268201800002I). Core support including phenotype harmonization, data management, sample-identity QC, and general program coordination were provided by the TOPMed Data Coordinating Center (R01HL-120393; U01HL-120393; contract HHSN268201800001I).
The latest gnomAD data set (v4.1.0) can be downloaded at https://gnomad.broadinstitute.org. phyloP scores can be downloaded at https://cgl.gi.ucsc.edu/data/cactus/241-mammalian-2020v2-hub/Homo_sapiens/241-mammalian-2020v2.bigWig. Raw and processed coronary artery snATAC data are available in Gene Expression Omnibus (GEO) under accession ID: GSE175621. The full dataset of the FAVORannotator's database can be downloaded at https://favor.genohub.org/favor-annotator.

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	Our findings apply to both sexes. Sex was used as a covariate in our heritability analyses. Concordance between annotated sex and genetic sex was inferred from the whole genome sequencing data by the TOPMed Data Coordinating Center (more details can be found at https://topmed.nhlbi.nih.gov/topmed-whole-genome-sequencing-methods-freeze-9).
Reporting on race, ethnicity, or other socially relevant groupings	In our comparison of self-identified race/ethnicity (SIRE) and genetic ancestry as inferred using ADMIXTURE and principal component analyses, we followed recommendations made by the TOPMed program (Khan, A. T. et al. Recommendations on the use and reporting of race, ethnicity, and ancestry in genetic research: Experiences from the NHLBI TOPMed program. Cell Genomics 2, 100155 (2022)). SIRE categories included White/European, Black/African American, Hispanic/Latino, East Asian and South Asian. All heritability estimates were adjusted by coronary artery disease status (case or control), age, sex, study and the first 15 genetic principal components (PCs).
Population characteristics	TOPMed Freeze 9 dataset "minDP10" includes ~800 million single nucleotide variants (SNVs) and ~62 million indels from autosomal chromosomes which were aligned to the GRCh38 human genome build. Coronary artery disease (CAD) cases and controls were identified from available medical records from 13 TOPMed studies totalling 64,397 samples (more details in Methods). After quality control, relatedness and genetic ancestry inference, we kept 22,443 samples (4,949 CAD cases and 17,494 controls) of inferred European ancestry with genotypic data for ~28 million biallelic SNVs. We also estimated CAD heritability in a restricted sample of 1,733 cases and 7,783 controls of inferred African genetic ancestry.
Recruitment	Recruitment differs for each study. Details of study-specific recruitment process can be found through dbGaP with the corresponding TOPMed accession numbers: Amish (phs000956), ARIC (phs001211), BioMe (phs001644), CARDIA (phs001612), CHS (phs001368), COPDGene (phs000951), DHS (phs001412), FHS (phs000974), GeneSTAR (phs001218), GENOA (phs001345), JHS (phs000964), MESA (phs001416), WHI (phs001237). Most participants included in these studies have been recruited on the basis of various cardiovascular disease diagnostics and their associated risk factors, and do not represent a representative sample of the U.S. population (more details can be found in the Supplementary Information).
Ethics oversight	This proposal has been discussed and approved by the TOPMed Atherosclerosis Working Group.

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	TOPMed studies included in our paper were based on the availability of CAD status for participants. Sample size in the European sample was considered large enough to provide interpretable variance contributions to overall heritability estimates within 1 standard error (SE). However, owing to the large SEs observed in the African sample, the main analyses presented in this paper were restricted to the European sample. Covariates like age and sex were accounted for in our heritability analyses since they represent potential confounders.

Data exclusions	Exclusion criteria were pre-established for variants and samples. Many rounds of quality control were applied. Variants were removed if they were: 1) indels or multiallelic; 2) share the same base pair position; 3) FILTER != PASS in the vcf files; 4) minor allele count (MAC) < 5; 5) have missing call rate > 1%. Then, 3,630 samples were removed due to ambiguous coronary artery disease (CAD) status, and 23 samples with missing call rate > 5%. We removed one member of each pair of participants related at degree four and above (kinship coefficient ≥ 2^-11/2) using PC-AiR and PC-Relate. We then selected participants with an ADMIXTURE European ancestry fraction > 98% and a Mahalanobis distance < 100 using PC scores of five European populations from the 1000 Genomes project. In addition, 603 samples were removed due to high inbreeding coefficient. Finally, SNVs were excluded if Hardy-Weinberg test p-value < 10^-6, and/or case-control missingness test p-value < 0.05. We repeated the same quality control steps as above in the TOPMed subset of self-identified Black/African American (BLK) participants. To reach a sufficient sample size, we applied a more relaxed threshold with respect to the inferred African ancestry (>75%).
Replication	No replication was done of our heritability analyses in another cohort. It is difficult to identify cases and controls using our precise definition of coronary artery disease and, in addition, having whole genome sequence data. However, we compared the CAD heritability estimate in our inferred European sample with the most recent estimate reported in 19,392 non-Hispanic White participants of the Million Veteran Program (MVP). Tcheandjieu et al. applied the same GREML-LDMS-I approach but binned the variants differently: quartiles of LD scores and six MAF bins were used. Using the GCTA REML EM algorithm and the same binning, we estimated a total observed heritability = 14.1% (SE = 5.5%), which is considerably less than their estimated heritability = 24.4% (SE = 4.7%). The major discrepancy came from SNVs in the lowest LD score quartile (Q1) of the lowest MAF bin (0.1% < MAF < 1%). However, when we added the ultra-rare variants (MAF < 0.1%), which were not included in the MVP heritability analyses, the missing gap was more than closed with heritability = 26.9% (SE = 10.7%). We note that there are differences that may contribute to discrepancies in CAD heritability estimates between our study and Tcheandjieu et al., including whole-genome sequencing vs. imputed data, different binning of variants and phenotypic definitions.
Randomization	Randomization was not relevant in our study, because only observational phenotype data and whole genome sequencing were available for analysis.
Blinding	Blinding was not relevant in our study, because only observational phenotype data and whole genome sequencing were available for analysis.

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

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.*