nature portfolio

Corresponding author(s): Pyarajan, Saiju

Last updated by author(s): Jun 14, 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	Data collection: Clinical and genetic data in MVP were collected as previously described in the referenced works Hunter-Zinck et al. (2020), Gaziano et al. (2016), and Fang et al. (2019). No custom software was used for data collection.
Data analysis	HARE (https://github.com/tanglab/HARE); PLINK2 (v2.00alpha-3LM; for PCA and PRS); PLINK2 (v2.00alpha-2LM; for GWAS); REGENIE (v1.0.6.7); Eagle2 (v2.4); Minimac4; METAL (released 2010-05-05); METASOFT (v2.0.1); Matrix eQTL (R; v2.3); FastQTL (v7); susieR (v0.9.1.0); PRS-CS (released 2020-9-10); FINEMAP (v1.4); LDSTORE 2.0; GCTA-mtCOJO (v1.93.2 beta); PheWAS (R; v0.1); R (v4.0.2); LDSC (v1.0.1); coloc (R; v5.2.1)

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

The full summary level association data from the individual population analyses in MVP are available via the dbGaP study accession number phs001672 [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001672]. ILCCO [ref 7] summary statistics can be found in GWAS Catalog accession numbers GCST004748, GCST004744, and GCST004750 [https://www.ebi.ac.uk/gwas/publications/28604730]. OncoArray Consortium [refs 8,40] summary statistics used for replication can be found in dbGaP study accession number phs001273 [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001273]. GTEx v8 lung eQTL summary data were accessed on the GTEx portal [https://gtexportal.org]; full data are available via the dbGaP study accession number phs000424.v8.p2 [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000424.v8.p2]. Source data are provided with this paper.

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	All analyses were performed using sex assessed using genetic markers, and the term "sex" is used throughout the manuscript.
Reporting on race, ethnicity, or other socially relevant groupings	Analyses were performed using HARE-defined harmonized race/ethnicity, described in Fang et al. (2019), and this characteristic is referred to as "ancestry" throughout the manuscript. Supplementary Data 1 summarizes participant numbers relevant to each cohort as well as their age and sex distribution.
Population characteristics	Characteristics of the MVP cohort and the International Lung Cancer Consortium OncoArray study cohort are described in Gaziano et al. (2016), and McKay et al. (2017), respectively.
Recruitment	This study did not do any recruitment. Subjects were recruited from VA Medical Centers participating in the Million Veteran Program. Veterans were identified from VA databases and recruited via invitational and appointment mailings. In addition, Veterans were recruited at selected VA Medical Centers. More details on recruitment are previously provided in PMID 26441289.
Ethics oversight	The study was approved by the VA central IRB and the VA Boston IRB. ILCCO and OncoArray participants provided written informed consent in IRB approved protocols.

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	Study was performed on the MVP Biobank data. Sample size was determined from summing the case and control numbers in each component study in the meta-analysis.
Data exclusions	Participant exclusions were made using standard criteria for GWAS, which included withdrawn consent, data missingness, and poor quality genotyping.
Replication	Replication for lung cancer and its subtypes was performed in an external OncoArray cohort. Nine novel loci were replicated.
Randomization	Randomization was not applicable to this GWAS study which is a retrospective analysis of lung cancer cases and controls.
Blinding	Blinding was not applicable to this GWAS study which is a retrospective analysis of lung cancer cases and controls.

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