nature portfolio

Corresponding author(s): Professor Ruth Loos

Last updated by author(s): May 6, 2025

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 data was newly collected for this project.
Data analysis	Statistical analyses were performed with Stata v15 (ALSPAC cohort only) and R v4.2.0 (visualization, and regression/prediction analyses), GCTA v1.93 for conditional analyses to identify the quasi-independent, genome-wide significant variants, PRS-CS v1.0.0 and PRS-CSx v1.0.0 to develop the genome-wide polygenic scores, PLINK v1.9 and v2 for standard quality control and creation of PGS within PGS cohorts, and we created a figure using Biorender.com (under a paid subscription). We did not develop custom software for this project.

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 PGS variant-level weights for the GIANT-only and GIANT+23andMe multi-ancestry PGSs are available for download through the PGS Catalog (https://

www.pgscatalog.org/, publication ID PGP000724, score IDs PGS005198-PGS005204). The source data of the figures are provided in the Supplementary Tables. This research has been conducted using the UK Biobank Research under Application Number 1251. The UKBB genotype and phenotype data are available through the UKBB at https://www.ukbiobank.ac.uk. Pan-UKBB relatedness and ancestry determinations can be obtained via UKBB return 2442. Due to US Department of Veterans Affairs (VA) regulations and our ethics agreements, the analytic datasets used for this study are not permitted to leave the Million Veteran Program (MVP) research environment and VA firewall. This limitation is consistent with other MVP studies based on VA data. However, the MVP data are made available to researchers with an approved VA and MVP study protocol. Information regarding access to the BioMe Biobank is available at https://icahn.mssm.edu/research/ipm/programs/biome-biobank. For access to the Uganda GPC-UGR data, contact should be made to the director MRC/UVRI & LSHTM Uganda Research Unit, Entebbe by email to mrc@mrcuganda.org. The data underlying this article are available upon request to the PLCO Cancer Data Access System: https://cdas.cancer.gov/plco/. Requests from researchers to access ALSPAC data and samples are welcomed; relevant procedures can be found at https://www.bristol.ac.uk/alspac/researchers/access/. Data from the Diabetes Prevention Program [(V9)/https://doi.org/10.58020/3hw5-cf91] reported here are available for request at the NIDDK Central Repository (NIDDK-CR) website, Resources for Research (R4R), https://repository.niddk.nih.gov/. The Diabetes Prevention Program genotype data were downloaded from the dbGaP web site, under phs000681.v2.p1. Researchers can access data from the primary Look AHEAD data trial on the NIDDK repository. Due to limitations specified in informed consent, the Look AHEAD genetic data are not publicly available.

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	Body mass index (BMI) underwent rank-based inverse-normal transformation by sex, both for the underlying study-specific BMI GWAS as well as for the majority of PGS-related analyses. We performed sex-stratified analyses in several of the PGS cohorts: (i) UK Biobank: to determine whether the PGS shows an overall difference in prediction accuracy for BMI between men and women; (ii) ALSPAC: to examine whether the PGS associates with timing of adiposity rebound in both boys and girls; (iii) PLCO cancer screening trial: association between the PGS and weight change between age 20 and 50y, and AUC of the PGS for obesity outcomes at age 50 when additionally stratified on overweight status at age 20. Determination of sex and use in sample QC and as a covariable for the 8 PGS studies is reported in STable 3 and the methods section, respectively.
Reporting on race, ethnicity, or other socially relevant groupings	For the underlying BMI GWAS, analyses were stratified per study by similarity to major ancestry group (AFR, AMR, EAS, EUR, SAS; group assignment defined by each contributing study themselves). The PGS analyses were similarly stratified by population group, with population descriptors/labels (as provided by the PGS-studies) and manner of determination described in STable 3.
Population characteristics	We describe the population characteristics (age/sex/obesity status) of the PGS studies in STable 4.
Recruitment	We provide a list of the >200 studies contributing to the underlying GWAS meta-analyses from the GIANT consortium and 23andMe in STable 1. We describe the 8 PGS studies in greater detail in the methods section. These PGS studies recruited participants under various designs including population-based and hospital-based prospective studies and biobanks, a birth cohort, and two clinical trials.
Ethics oversight	Our research complies with all relevant ethical regulations. The UK Biobank Study was approved by the North West Multi-Centre Research Ethics Committee (Ref 11/NW/0382) and all participants provided written informed consent to participate in the UK Biobank Study. The VA Central Institutional Review Board approved the MVP study protocol in accordance with the principles outlined in the Declaration of Helsinki. Informed consent was obtained from all participants. The BioMe Biobank Program (Institutional Review Board 07-0529) operates under a Mount Sinai Institutional Review Board-approved research protocol. All study participants provided written informed consent. The Uganda Genome Resource was approved by the Science and Ethics Committee of the UVRI Research and Ethics Committee (UVRI-REC #HS 1978), the Uganda National Council for Science and Technology (UNCST #SS 4283), and the East of England-Cambridge South (formerly Cambridgeshire 4) NHS Research Ethics Committee UK. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Consent for biological samples has been collected in accordance with the Human Tissue Act (2004). Informed consent for the use of data collected via questionnaires and clinics was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time. The PLCO study was approved by the human subjects review boards at the National Cancer Institute and at the 10 study centres (National Institutes of Health Institutional Review Board #OH97CN041). Written informed consent was given by all participants. The Diabetes Prevention Program and Look AHEAD trials were approved by the institutional review board at each center, and all participants gave written informed consent.

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	For our PGS development, we build on the GWAS meta-analyses for BMI conducted by the GIANT consortium and 23andMe, comprising data from over 200 cohorts, to achieve the largest possible sample size (here, N=5.1 million). To evaluate the performance of the PGS, we did not

	perform sample size calculations. However, we leveraged data from large-scale observational biobanks and cohorts representing diverse populations of adult participants (UK Biobank, Million Veteran Program, PLCO Cancer Screening Trial, Uganda General Population Cohort, BioMe Biobank (combined N>560k)), a birth cohort (ALSPAC, up to 7,600 individuals), and two clinical trials of intensive lifestyle interventions (DPP and Look-AHEAD).
Data exclusions	For PGS development, we included variants present in the 1000G reference panels made available by the PRS-CS(x) developers (HapMap3 variants (N=1-1.2M) which were common in 1000G populations, excluding ambiguous A/T or G/C variants), which additionally had an INFO>0.3 in the overall UK Biobank. Ancestry-specific GWAS summary statistics were further restricted to the variants with a variant-specific sample size of at least a third of the maximum sample size for that ancestry. Sample and genotype QC for the 8 PGS cohorts are described in STable 3 and the methods section.
Replication	See Sample size. Initial selection of the best-performing PGS took place in the UK Biobank, with replication and other downstream analyses taking place in an independent subset of the UK Biobank and 7 other studies.
Randomization	N/A, as our analyses largely took place in observational setting, and post-hoc analyses in two clinical trials of intensive lifestyle intervention.
Blinding	N/A, as our analyses largely took place in observational setting, and post-hoc analyses in two clinical trials of intensive lifestyle intervention.

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	N/A
Novel plant genotypes	N/A
Authentication	N/A