Trans-ethnic Polygenic Risk Scores for Body Mass Index: An International Hundred K+ Cohorts Consortium Study

Huiqi Qu¹, John J Connolly¹, Peter Kraft², Jirong Long³, Alexandre Pereira²,⁴, Christopher Flatley⁵, Constance Turman², Bram Prins⁶, Frank Mentch¹, Paulo A Lotufo⁷,⁸, Per Magnus⁹,¹⁰, Meir J Stampfer²,¹¹,¹², Rulla Tamimi², A Heather Eliassen², Wei Zheng³, Gun Peggy Stromstad Knudsen⁵, Oyvind Helgeland⁵, Adam S. Butterworth⁶,¹³,¹⁴,¹⁵, Hakon Hakonarson¹,¹⁶,¹⁷,¹⁸,¹⁹*, Patrick M. Sleiman¹,¹⁶,¹⁷, on behalf of the IHCC consortium

Affiliations:

¹The Center for Applied Genomics, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania, 19104, USA.

²Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

³Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

⁴Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA;

⁵Division of Health Data and Digitalization, Department of Genetics and Bioinformatics, Norwegian Institute of Public Health, Oslo, Norway.

⁶British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.

⁷Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil;

⁸Centro de Pesquisas Clínicas e Epidemiológicas, Hospital Universitário, Universidade de São Paulo, São Paulo, Brazil.

⁹University of Oslo, Oslo, Norway.

¹⁰Center for Fertility and Health, Norwegian Institute of Public Health, Oslo, Norway.

¹¹Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

¹²Channing Division of Network Medicine, Department of Medicine, Harvard Medical School, Boston, MA, USA.

¹³The National Institute for Health Research Blood and Transplant Research Unit (NIHR BTRU) in Donor Health and Genomics, Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB1 8RN, UK.

¹⁴British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
15 Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, UK.

16 Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, 19104, USA.

17 Division of Human Genetics, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania, 19104, USA.

18 Division of Pulmonary Medicine, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania, 19104, USA.

19 Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland.

*Correspondence:

Dr. Hakon Hakonarson
Center for Applied Genomics
3615 Civic Center Blvd
Abramson Building
Philadelphia, PA 19104, United States of America
Telephone: 267-426-0088
Fax: 267-426-0363
Email: hakonarson@chop.edu

Word count: 2,192

Number of tables and figures: 5 Tables and 2 figures.
Abstract

Background While polygenic risk scores hold significant promise in estimating an individual’s risk of developing a complex trait such as obesity, their application in the clinic has, to date, been limited by a lack of data from non-European populations. As a collaboration model of the International Hundred K+ Cohorts Consortium (IHCC), we endeavored to develop a globally applicable trans-ethnic PRS for body mass index (BMI) through this relatively new international effort.

Methods The PRS model was developed and tested at the Center for Applied Genomics (CAG) of The Children’s Hospital of Philadelphia (CHOP) based on a BMI meta-analysis from the GIANT consortium. The validated PRS models were subsequently disseminated to the participating sites. Scores were generated by each site locally on their cohorts and summary statistics returned to CAG for final analysis.

Results We show that in the absence of a well powered trans-ethnic GWAS from which to derive SNPs and effect estimates, trans-ethnic scores can be generated from European ancestry GWAS using Bayesian approaches such as LDpred to adjust the summary statistics using trans-ethnic linkage disequilibrium reference panels. The ported trans-ethnic scores outperform population specific-PRS across all non-European ancestry populations investigated including East Asians and three-way admixed Brazilian cohort.

Conclusions Widespread use of PRS in the clinic is hampered by a lack of genotyping data in individuals of non-European ancestry for the vast majority of traits. Here we show that for a truly polygenic trait such as BMI adjusting the summary statistics of a well powered European ancestry study using trans-ethnic LD reference results in a score that is predictive across a range of ancestries including East Asians and three-way admixed Brazilians.

Key words: body mass index; obesity; polygenic risk score; population admixture; trans-ethnic

Introduction
Obesity is a global health issue with an adult prevalence of about 13% across the world (https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight). As a difficult condition to treat, obesity prevention is important which has led us to develop a trans-ethnic polygenic risk score (PRS) for body mass index (BMI) through the International HundredK+ Cohorts Consortium (IHCC). PRS aggregates the effects of many genetic variants across the human genome into a single score, which may effectively improve the prediction of a complex disease/trait and assist the differential diagnosis. Our obesity PRS was based on the published GWAS meta-analysis of BMI that included 339,224 individuals of European ancestry. The study identified 97 genome-wide significant BMI-associated loci that account for approximately 2.7% of BMI variation alone. Genome-wide estimates suggest that common variation accounts for over 20% of variation in BMI. Various approaches for PRS calculation have been developed to date. The standard approaches for calculating risk scores involve linkage disequilibrium (LD)-based marker pruning followed by p-value thresholding of GWAS-based association statistics. While effective, these approaches lose information and can reduce predictive accuracy particularly where the test population differs in genetic ancestry from the GWAS sample. Bayesian approaches such LDpred, a method that infers the posterior mean effect size of each marker by using a prior on effect sizes and LD information from an external reference panel may therefore improve prediction accuracy in multi-ethnic studies of diverse populations. As most large scale GWAS have been conducted using only individuals of European ancestry there is a need to develop approaches that can port PRS using European ancestry derived effect estimates. More importantly, BMI and obesity are closely related to environmental factors. It is essential to validate a multi-ethnic PRS in different regional populations, especially admixed populations. Leveraging an international effort supported by the IHCC that has brought together large scale cohorts with genotyping data from around the world, we explored the development of a LDpred-based, trans-ethnic (TE) obesity PRS through a collaboration of 6 international research centers (Table 1).

Methods

**Model training at the Children’s Hospital of Philadelphia (CHOP)**

The PRS model development and training was carried out at the Center for Applied Genomics (CAG) according to the pipeline shown in Fig.1. SNP weights (i.e. posterior mean effect sizes) were calculated using Markov chain Monte Carlo (MCMC) Gibbs sampling as implemented in LDpred. The summary statistics of genetic association with BMI were based on the meta-analysis of genome-wide association studies by the GIANT consortium. We restricted the variants to SNPs included in the HapMap data. Five sets of SNP weights were generated based on the respective LD patterns from the following populations, African American (AA), Hispanic American (HAMR), East Asian (ASN), Northern European (EUR) and all of the above populations (i.e. trans-ethnic). For each group we selected 2500 CAG participants who clustered with the 1000 genome project reference data to generate the LD files. The trans-ethnic group included all populations tested. Each group of SNP weights includes 8 different sets corresponding respectively to the mixture probability values (i.e. fractions of causal markers) of [infinitesimal prior (LDpred-inf), 1, 0.3, 0.1, 0.03, 0.01, 0.003, and 0.001].
Initial validation of PRS models in-house

The initial assessment of the PRS models was based on the genomic data in-house at CAG, CHOP. The pediatric biobank built at CAG has archived samples from 500,000 children from USA, Europe, South America, Canada, Saudi Arabia and Australia. For the initial validation of the trans-ethnic PRS model, 57,613 randomly selected individuals including individuals of European ancestry, African Americans, Hispanics/Latinos, and East / South Asians in order of frequency, with genotypes and BMI data from the CAG biobank were used for the validation. The principal component analysis (PCA) plot of the population structure is shown in Supplementary Figure 1. All the individuals have been genotyped with an Illumina Genotyping BeadChip with at least 550,000 SNPs genotyped. Genome-wide imputation was done using the TOPMed Imputation Server (https://imputation.biodatacatalyst.nhlbi.nih.gov) with the TOPMed (Version R2 on GRCh38) Reference Panel. Population ancestries of the research subjects were confirmed by principal component analysis (PCA) with genomic DNA markers, compared with the reference populations in the 1000 Genomes project. Harmonization of SNP alleles in the PRS model was confirmed by comparing with the reference alleles of the TOPMed imputation.

Validation of PRS models in regional populations

Having validated the trans-ethnic PRS models, we shared the protocol of the PRS models, as well as the SNPs and weights with the IHCC collaborators for assessment in 7 different cohorts of regional populations. The population sites included a three-way admixed Brazilian cohort from ELSA-Brasili, the Norwegian Mother, Father and Child Cohort Study (MoBa, conducted by the Norwegian Institute of Public Health), two US based studies The Nurses’ Health Study (NHS) and Nurses’ Health Study II (NHSII), the UK based INTERVAL BioResource and two Chinese population samples from the Shanghai Men’s Health Study (SMHS) and Shanghai Women’s Health Study (SWHS).

Each site generated the PRS on their cohort following the same protocol and using the same tools and the software packages. To compare the performance of trans-ethnic PRS vs. that of population-specific PRS in different regional populations, we requested each collaboration group run two calculations if possible, i.e., one PRS calculation for the specific population that is closest to their dataset, and one PRS for the trans-ethnic score. To do the dichotomous receiver operating characteristic (ROC) curve analysis, the top 5% and 1% of the BMI distribution within each study were defined as cases.

Participating IHCC cohorts:

The Nurses’ Health Study I recruited 121,700 married registered nurses in 1976. Blood samples were collected from 33,000 participants in 1989-90, and cheek cells from another 33,000 in 2001-4. Genome-wide association data are available on over 17,000 participants as part of studies of multiple complex diseases and traits, including breast cancer, type 2 diabetes, venous thromboembolism, and depression. All women included in these analyses are of European ancestry (cluster with European reference samples and do not self-report as other than European ancestry).

The Nurses’ Health Study II recruited 116,430 married registered nurses between the ages of 25
and 42 in 1989. Blood samples were collected from 29,000 participants in 1996-99, and cheek cells from another 30,000 in 2006. Genome-wide association data are available on over 12,000 participants as part of studies of multiple complex diseases and traits, including breast cancer, type 2 diabetes, venous thromboembolism, and depression. All women included in these analyses are of European ancestry (cluster with European reference samples and do not self-report as other than European ancestry).

Body mass index was self-reported at time of blood draw or cheek cell collection. Diabetes cases were defined as self-reported diabetes confirmed by a validated supplementary questionnaire.

The Shanghai Women's Health Study (SWHS) is a large population-based prospective cohort study initiated in 1996. Approximately 75,000 Chinese women who lived in Shanghai were recruited into the study. In addition to survey data, blood and urine samples were collected from most study participants at the baseline recruitment.

The Shanghai Men's Health Study (SMHS) is a population-based cohort study of 61,480 Chinese men between ages 40 and 74 who lived in eight urban communities in Shanghai at enrollment (2002-2006). Detailed information on dietary and other lifestyle factors was collected at baseline and updated in follow-up surveys. Biological samples (blood, and or urine) were collected from 89% of cohort members.

The Norwegian Mother, Father and Child Cohort Study (MoBA) was established and is conducted by the Norwegian Institute of Public Health (NIPH). MoBa is an ambitious family-oriented cohort study that aims to find causes of diseases and explain trajectories and variability of health-related traits over a life-course span. Between 1999 - 2008 pregnant women were invited to take part in the study around the time of the ultrasound examination in week 17-20 of gestation. The fathers of the children were also invited to participate. Biological material has been collected from mothers, fathers and children and has been stored in a biobank. Self-reported data are collected from regular questionnaires about general health, diet and environmental exposure. The cohort includes approximately 109,000 children, 91,000 women and 71,700 men. 50,290 Northern European adult males and females were analysed in this study, with mean BMI=24.96(SD=3.9).

The Brazilian Longitudinal Study of Adult Health (ELSA-Brasil) enrolled 15,105 civil servants aged 35 to 74 years-old living in six cities, addressing the incidence of non-communicable diseases. From the 15105 participants, 9333 DNA samples were analyzed for genetic ancestry using a software tool for maximum likelihood estimation of individual ancestries from multi locus SNP genotype datasets.

The INTERVAL BioResource recruited 45,263 whole blood donors (22,466 men and 22,797 women) between June 11, 2012, and June 15, 2014. Donors were aged 18 years or older from 25 NHS Blood and Transplant (NHSBT) blood donation centres distributed across England, UK. Donors provided blood samples at baseline to enable DNA extraction and self-reported their height and weight for estimation of BMI. Genotyping was conducted using the Affymetrix Axiom UK Biobank array with imputation using a combined 1000 Genomes Phase 3 / UK10K
reference panel. A total of 38,319 European ancestry adult participants were included in the final analyses

**Results**

**Initial validation of PRS models in-house**

The performance of the trans-ethnic PRS model tested in-house is shown in Fig.2. As shown by our analysis, AUC>0.720 to predict top 1% BMI was achieved in all the PRS models with mixture probability≥0.03. The highest AUC is seen at the mixture probabilities of 0.03 to 0.1.

**Validation of the trans-ethnic PRS models in regional populations**

By the Chinese population in the SMHS and SWHS cohorts, the PRS models were tested in 866 men in the SMHS cohort and 3,120 women in the SWHS cohort and the combination of both. Across all three analyses the trans-ethnic (TE) score outperformed the ancestry specific East-Asian reference with a TE AUC of 0.76 compared with 0.479 for the ASN reference in men, TE AUC of 0.737 vs ASN AUC of 0.719 in women, and a TE AUC of 0.717 vs ASN AUC of 0.706 in the combined dataset at the most informative fraction (Table 2).

The Brazilian population in the ELSA-Brasil cohort is three way admixed. We therefore compared the trans-ethnic score in this population to each of the three founders: American, African and European. The trans-ethnic score outperformed all three founder population ancestry-specific scores at the most informative fraction of 0.03 at both 95th and 99th percentiles (Table 3).

**Evaluation of the PRS in the four European ancestry cohorts**

Across all four European ancestry cohorts both European-specific PRS models and the trans-ethnic PRS models had the largest AUCs with a mixture probability of 0.03. Across the three cohorts where both the European-specific PRS model and the trans-ethnic PRS model were run, (INTERVAL, NHS and NHSII), the performance of the scores was comparable (Table 4).

Only the European-specific PRS models were tested in the Norwegian MoBa study. The PRS model showed AUC≥0.686 to predict top 5% BMI, and AUC≥0.735 to predict top 1% BMI, in the models with mixture probability≥0.03 (Table 5).

**Discussion**

In order for polygenic risk scores to achieve their clinical potential, and avoid exacerbating health disparities due to the lack of genomic information in minorities, they have to be universally applicable regardless of a patient’s genetic ancestry. Ideally, trans-ethnic PRS would be generated from true trans-ethnic GWAS however, well powered trans-ethnic studies remain the exception for the majority of traits and phenotypes. In this study we evaluated the performance of a BMI PRS that is based on European ancestry GWAS effect sizes combined with trans-ethnic LD patterns using a Bayesian approach as implemented in LDpred. In a federated model, we developed the PRS score at CAG and disseminated the standard operating procedure along with the SNPs and weights files and the population specific LD matrices to
participant sites within the IHCC. This model allowed us to quickly test the hypothesis in various world populations without the need for data transfer and hence time-consuming data sharing agreements. By providing detailed protocols and all required files to run the scores, we made efforts to minimize the work load on each collaboration group while enabling direct comparisons between groups.

It is worth to mention that, in contrast to the LDpred4 used in this study, a new version of the method LDpred2 has been released\textsuperscript{20}. LDpred2 made a significant effort to address the potential bias of Gibbs sampling in the human leukocyte antigen (\textit{HLA}) region at chromosome 6 with extended LD\textsuperscript{20}. In contrast, the \textit{HLA} region has been removed in the LDpred modeling. The \textit{HLA} region is highly polymorphic, with highly diverse frequencies across different populations, as well as extended and strong LD due to significant evolutionary selection pressure in human populations\textsuperscript{21,22}. Including the \textit{HLA} region will cause significant difference of PRS across different ethnicities, as we have observed in the trans-ethnic scoring of an autoimmune disease, type 1 diabetes (T1D), with \textit{HLA} as a major risk factor\textsuperscript{23}. In addition, there has been no GWAS study to date suggesting the role of the \textit{HLA} region in BMI or obesity. Nevertheless, it is interesting to leverage the IHCC resources to examine the performance of LDpred2 in the trans-ethnic scoring of BMI.

As the results show, our trans-ethnic PRS models outperformed the ancestry specific models in the non-European populations tested including Chinese and Brazilian. Importantly the UK and US data from the INTERVAL and NHS cohorts demonstrate that there is no appreciable loss in predictive power in European ancestry individuals when using a trans-ethnic score. As such, and in the absence of trans-ethnic effect sizes from diverse GWAS, we propose that using trans-ethnic reference data to adjust the summary statistics for the effects of LD patterns improves the performance of PRS in populations that have not been included in the generation of the summary stats. On the other hand, in the absence of reference panels for different populations, the genome-wide LD scores and matrices from the Pan-UK Biobank resource\textsuperscript{24}, or calculated by genome sequencing data of the Genome Aggregation Database (gnomAD)\textsuperscript{25}, may provide sufficient resolution.

As the first effort by the IHCC to leverage the existing datasets that reside within this large scale consortium for a transethnic PRS on BMI, we envision an opportunity to scale this to other cohorts within the consortium and expand the number of traits that can be analyzed. As such, the IHCC presents a rich resource of data for collaborative research with transethnic focus, where there is much unmet need at the present time and an area of research that has been largely ignored.

\textbf{Acknowledgements} We thank all the participants who contributed to and enabled this study.

\textbf{Ethical Approval}

This study was exempted by the Institutional Review Board (IRB) of the Children’s Hospital of Philadelphia. Human participants and personal information are inaccessible to the research group due to sample and data deidentification. All human subjects or their proxies provided written
informed consent for the respective studies involved.

**Funding**

The study was supported by:

All authors are supported by the International HundredK+ Cohorts Consortium (IHCC), which has been created in collaboration with the Global Alliance for Genomics and Health (GA4GH) and the Global Genomics Medicine Collaborative (G2MC) with support from the National Institute of Health and the Wellcome Trust.

CHOP funding: Institutional Development Funds from the Children’s Hospital of Philadelphia to the Center for Applied Genomics, and The Children’s Hospital of Philadelphia Endowed Chair in Genomic Research.

The Shanghai Women's Health Study (US NIH, R37CA070867, UM1CA182910), and the Shanghai Men's Health Study (US NIH, R01CA082729, UM1CA173640), are funded by the National Institutes of Health.

ELSA-Brasil is funded by the Brazilian Ministry of Health (Department of Science and Technology) and the Ministry of Science, Technology and Innovation (FINEP, Financiadora de Estudos e Projetos), and CNPq (the National Council for Scientific and Technological Development).

The Norwegian Mother, Father and Child Cohort Study (MoBa) is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research. We are grateful to all the participating families in Norway who take part in this on-going cohort study. We thank the Norwegian Institute of Public Health (NIPH) for generating high-quality genomic data. This research is part of the HARVEST collaboration, supported by the Research Council of Norway (#229624). We also thank the NORMENT Centre for providing genotype data, funded by the Research Council of Norway (#223273), South East Norway Health Authorities and Stiftelsen Kristian Gerhard Jebsen. We further thank the Center for Diabetes Research, the University of Bergen for providing genotype data and performing quality control and imputation of the data funded by the ERC AdG project SELECTionPREDISPOSED, Stiftelsen Kristian Gerhard Jebsen, Trond Mohn Foundation, the Research Council of Norway, the Novo Nordisk Foundation, the University of Bergen, and the Western Norway Health Authorities.

Participants in the INTERVAL randomized controlled trial were recruited with the active collaboration of NHS Blood and Transplant (www.nhsbt.nhs.uk), which has supported fieldwork and other elements of the trial. DNA extraction and genotyping were co-funded by the National Institute for Health Research (NIHR), the NIHR BioResource (http://bioresource.nihr.ac.uk) and the NIHR Cambridge Biomedical Research Centre (BRC) (no. BRC-1215-20014). The academic coordinating centre for INTERVAL was supported by core funding from the NIHR Blood and Transplant Research Unit in Donor Health and Genomics (no. NIHR BTRU-2014-10024), UK Medical Research Council (MRC) (no. MR/L003120/1), British Heart Foundation (nos SP/09/002, RG/13/13/30194 and RG/18/13/33946) and the NIHR Cambridge BRC (no. BRC-1215-20014). A complete list of the investigators and contributors to the INTERVAL trial is provided in ref.17. The academic coordinating centre thanks blood donor centre staff and blood
donors for participating in the INTERVAL trial. This work was supported by Health Data Research UK, which is funded by the UK MRC, Engineering and Physical Sciences Research Council (EPSRC), Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome. The views expressed in this manuscript are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Declaration of Conflicting Interests

A.S.B declares grants outside of this work from AstraZeneca, Bayer, Biogen, BioMarin, Merck and Sanofi. All other authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Consent for publication

All authors have provided consent for publication of the manuscript.

Availability of supporting data

Supporting data from this study can be obtained by emailing the corresponding author Dr. Hakon Hakonarson.

Figure Legend

**Figure 1** Development of a Transethnic Polygenic Risk Score for Body Mass Index.

**Figure 2** AUC values following the initial validation of PRS models in-house. 57,613 randomly selected individuals with genotypes and BMI data from the the Center for Applied Genomics biobank at the Children’s Hospital of Philadelphia were used for the validation.

**Supplementary Figure 1.** The principal component analysis (PCA) plot of the population structure of CAG samples for initial validation.
Table 1 Six international research centers for the development of the trans-ethnic PRS of BMI

<table>
<thead>
<tr>
<th>Cohort Name</th>
<th>PI - Lead</th>
<th>Participating countries</th>
<th>Current Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELSA-Brasil</td>
<td>Paulo A. Lotufo</td>
<td>Brazil: six cities</td>
<td>15,105</td>
</tr>
<tr>
<td>Norwegian Mother and Child Cohort Study (MoBa)</td>
<td>Per Magnus</td>
<td>Norway</td>
<td>284,000</td>
</tr>
<tr>
<td>Children's Hospital of Philadelphia (CHOP)</td>
<td>Hakon Hakonarson</td>
<td>USA, Europe, South America, Canada, Saudi Arabia and Australia</td>
<td>500,000</td>
</tr>
<tr>
<td>Biorepository</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHS (Nurses' Health Study, NCI)</td>
<td>Meir Stampfer, Rulla Tamimi</td>
<td>USA</td>
<td>121,700</td>
</tr>
<tr>
<td>NHSII (Nurses' Health Study II, NCI)</td>
<td>Walter Willett, Heather Eliassen</td>
<td>USA</td>
<td>116,430</td>
</tr>
<tr>
<td>Shanghai Men and Women's Health Study (2 cohorts)</td>
<td>Wei Zheng</td>
<td>Shanghai, China</td>
<td>136,000</td>
</tr>
<tr>
<td>Fraction of causal variants</td>
<td>Racial groups of weights</td>
<td>SMHS (N=866)</td>
<td>SWHS (N=3,120)</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------------------------</td>
<td>--------------</td>
<td>----------------</td>
</tr>
<tr>
<td>INF</td>
<td>ASN</td>
<td>P* 0.07</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>TE</td>
<td>0.04486 1</td>
<td>0.68 4</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>ASN 0.05</td>
<td>0.68 2</td>
</tr>
<tr>
<td></td>
<td>TE</td>
<td>0.04016 9</td>
<td>0.68 4</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>ASN 0.03</td>
<td>0.70 7</td>
</tr>
<tr>
<td></td>
<td>TE</td>
<td>0.03012 9</td>
<td>0.69 2</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>ASN 0.02</td>
<td>0.72 1</td>
</tr>
<tr>
<td></td>
<td>TE</td>
<td>0.01729 9</td>
<td>0.71 6</td>
</tr>
<tr>
<td></td>
<td>0.03</td>
<td>ASN 0.03</td>
<td>0.71 9</td>
</tr>
<tr>
<td></td>
<td>TE</td>
<td>0.01589 2</td>
<td>0.73 3</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>ASN 0.22</td>
<td>0.65 8</td>
</tr>
<tr>
<td></td>
<td>TE</td>
<td>0.11957 7</td>
<td>0.96 3</td>
</tr>
<tr>
<td></td>
<td>0.003</td>
<td>ASN 0.21</td>
<td>0.62 2</td>
</tr>
<tr>
<td></td>
<td>TE</td>
<td>0.73 7</td>
<td>0.52 7</td>
</tr>
<tr>
<td></td>
<td>0.001</td>
<td>ASN 0.87</td>
<td>0.47 9</td>
</tr>
<tr>
<td></td>
<td>TE</td>
<td>0.01778 9</td>
<td>0.76 13</td>
</tr>
</tbody>
</table>

* Logistic regression analyses were performed after coding top 99% distribution of BMI as cases and the remainings as controls. # Linear regression analyses were performed using BMI as continuous variable.
Table 3 BMI PRS test by the Brazilian population in the ELSA-Brasil cohort

<table>
<thead>
<tr>
<th>ELSA-Brazil</th>
<th>BMI %</th>
<th>AFR</th>
<th>AMR</th>
<th>EUR</th>
<th>TE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inf*</td>
<td>99</td>
<td>0.569</td>
<td>0.575</td>
<td>0.548</td>
<td>0.589</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>0.575</td>
<td>0.553</td>
<td>0.538</td>
<td>0.564</td>
</tr>
<tr>
<td>1</td>
<td>99</td>
<td>0.604</td>
<td>0.582</td>
<td>0.558</td>
<td>0.595</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>0.575</td>
<td>0.554</td>
<td>0.542</td>
<td>0.569</td>
</tr>
<tr>
<td>0.1</td>
<td>99</td>
<td>0.635</td>
<td>0.63</td>
<td>0.579</td>
<td>0.616</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>0.581</td>
<td>0.559</td>
<td>0.544</td>
<td>0.57</td>
</tr>
<tr>
<td>0.3</td>
<td>99</td>
<td>0.612</td>
<td>0.587</td>
<td>0.563</td>
<td>0.602</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>0.575</td>
<td>0.554</td>
<td>0.543</td>
<td>0.566</td>
</tr>
<tr>
<td>0.03</td>
<td>99</td>
<td>0.558</td>
<td>0.636</td>
<td>0.606</td>
<td>0.645</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>0.517</td>
<td>0.571</td>
<td>0.553</td>
<td>0.579</td>
</tr>
<tr>
<td>0.01</td>
<td>99</td>
<td>0.472</td>
<td>0.56</td>
<td>0.634</td>
<td>0.538</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>0.546</td>
<td>0.526</td>
<td>0.573</td>
<td>0.511</td>
</tr>
<tr>
<td>0.003</td>
<td>99</td>
<td>0.524</td>
<td>0.615</td>
<td>0.525</td>
<td>0.535</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>0.522</td>
<td>0.547</td>
<td>0.497</td>
<td>0.521</td>
</tr>
<tr>
<td>0.001</td>
<td>99</td>
<td>0.475</td>
<td>0.585</td>
<td>0.538</td>
<td>0.448</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>0.519</td>
<td>0.521</td>
<td>0.508</td>
<td>0.492</td>
</tr>
</tbody>
</table>

* Infinitesimal prior
## Table 4 BMI PRS test by the NHS and NHSII cohorts*

<table>
<thead>
<tr>
<th>Top 0.03 AUC</th>
<th>1% / P-</th>
<th>EUR-NHS1</th>
<th>TE-NHS1</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHS</td>
<td>0.730</td>
<td>0.727</td>
<td></td>
</tr>
<tr>
<td>NHSII</td>
<td>0.736</td>
<td>0.735</td>
<td></td>
</tr>
<tr>
<td>Top 0.03 AUC</td>
<td>5% / P-</td>
<td>EUR-NHS1</td>
<td>TE-NHS1</td>
</tr>
<tr>
<td>NHS</td>
<td>0.703</td>
<td>0.701</td>
<td></td>
</tr>
<tr>
<td>NHSII</td>
<td>0.721</td>
<td>0.718</td>
<td></td>
</tr>
</tbody>
</table>

* NHS: the Nurses’ Health Study.
Table 5 BMI PRS test by the Norwegian Mother and Child Cohort Study (MoBA)

<table>
<thead>
<tr>
<th>MoBa</th>
<th>95% BMI</th>
<th>99% BMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inf*</td>
<td>0.686</td>
<td>0.735</td>
</tr>
<tr>
<td>1</td>
<td>0.689</td>
<td>0.738</td>
</tr>
<tr>
<td>0.3</td>
<td>0.694</td>
<td>0.745</td>
</tr>
<tr>
<td>0.1</td>
<td>0.699</td>
<td>0.751</td>
</tr>
<tr>
<td>0.03</td>
<td>0.704</td>
<td>0.756</td>
</tr>
<tr>
<td>0.01</td>
<td>0.658</td>
<td>0.7</td>
</tr>
<tr>
<td>0.003</td>
<td>0.594</td>
<td>0.626</td>
</tr>
<tr>
<td>0.001</td>
<td>0.552</td>
<td>0.565</td>
</tr>
</tbody>
</table>

* Infinitesimal prior
Reference:

Weight files generated with 8 different fractions of causal variants, starting at 1 and decreasing to 0.001

- African American Weight file
- Asian Weight file
- Hispanic Weight file
- European Weight file
- Trans-ethnic Weight file

40 files distributed to sites

Sites generate scores on their data using all 8 weight files
Report the AUC for the fraction with highest $R^2$

- AUC Closest matching population
- AUC Trans ethnic data

Binary trait for