Combining Asian-European Genome-Wide Association Studies of Colorectal Cancer Improves Risk Prediction Across Race and Ethnicity


1Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA.

2Department of Medicine (Medical Genetics), University of Washington Medical Center, Seattle, WA 98195, USA

3Division of Research, Kaiser Permanente Northern California, Oakland, California, USA.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
4 Genomic Medicine Institute, Cleveland Clinic, Cleveland, Ohio, USA.

5 Population and Cancer Prevention Program, Case Comprehensive Cancer Center, Cleveland, OH, USA.

6 Danish Institute for Advanced Study (DIAS), Epidemiology, Biostatistics and Biodemography, Department of Public Health, University of Southern Denmark, Odense, Denmark.

7 Colon Cancer Genetics Group, Medical Research Council Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, EH4 2XU, UK

8 Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt University Medical Center, Nashville, Tennessee, USA

9 Instituto de Investigacion Sanitaria de Santiago (IDIS), Choupana sn, 15706, Santiago de Compostela, Spain

10 Edinburgh Cancer Research Centre, Institute of Genomics and Cancer, University of Edinburgh, Crewe Road, Edinburgh EH4 2XU, UK

11 Division of Genetics and Epidemiology, The Institute of Cancer Research, London, SW7 3RP, UK


13 Digestive Diseases and Microbiota Group, Girona Biomedical Research Institute (IDIBGI), Salt, 17190, Girona, Spain.

14 Unit of Biomarkers and Susceptibility, Oncology Data Analytics Program, Catalan Institute of Oncology, Barcelona 08908, Spain.

15 Colorectal Cancer Group, ONCOBELL Program, Bellvitge Biomedical Research Institute, Barcelona 08908, Spain.

16 Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, Barcelona 08908, Spain.
17 Division of Human Nutrition and Health, Wageningen University & Research, Wageningen, The Netherlands.

18 Department of Preventive Medicine, Seoul National University College of Medicine, Seoul National University Cancer Research Institute, Seoul, South Korea.

19 Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden.

20 Department of Epidemiology, University of Washington, Seattle, WA, USA.

21 Institute for Health Research, Kaiser Permanente Colorado, Denver, Colorado, USA.

22 Center for Cancer Research, Medical University Vienna, Vienna, Austria.

23 Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA.

24 Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA.

25 Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA.

26 Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA.

27 Department of Epidemiology, Harvard T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts, USA.

28 Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts, USA.

29 Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York, USA.

30 University of Southern California, Preventative Medicine, Los Angeles, California, USA.
31 Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden.

32 Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden.

33 Department of Population Science, American Cancer Society, Atlanta, Georgia, USA.

34 SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA.

35 Colorado Center for Personalized Medicine, University of Colorado - Anschutz Medical Campus

36 Department of Population Science, American Cancer Society, Atlanta, Georgia, USA.

37 Center for Genetic Epidemiology, Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, California, USA.

38 Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, UK.

39 Colorectal Oncogenomics Group, Department of Clinical Pathology, The University of Melbourne, Parkville, Victoria 3000 Australia

40 University of Melbourne Centre for Cancer Research, Victorian Comprehensive Cancer Centre, Parkville, Victoria 3000 Australia

41 Genomic Medicine and Family Cancer Clinic, The Royal Melbourne Hospital, Parkville, Victoria 3000 Australia.

42 Department of Medicine (Medical Genetics), University of Washington Medical Center, Seattle, WA 98195, USA

43 Department of Bioinformatics and Medical Education, University of Washington Medical Center, Seattle, WA 98195, USA

44 Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, California, USA.
Department of Social and Preventive Medicine, Hallym University College of Medicine, Okcheon-dong, South Korea

VA Cooperative Studies Program Epidemiology Center, Durham Veterans Affairs Health Care System, Durham, North Carolina, USA.

Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC, USA.

Department of Epidemiology, University of Washington School of Public Health, Seattle, Washington, USA.

Cancer Epidemiology Program, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA.

Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, Ohio, USA.

Department of Community Medicine and Epidemiology, Lady Davis Carmel Medical Center, Haifa, Israel.

Ruth and Bruce Rappaport Faculty of Medicine, Technion-Israel Institute of Technology, Haifa, Israel.

Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, Australia.

Division of Human Genetics, Department of Internal Medicine, The Ohio State University Comprehensive Cancer Center, Columbus, Ohio, USA.

Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany.

Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany.

Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany.
58 Division of Cancer Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan.

59 Research Institute and Hospital, National Cancer Center, Goyang, South Korea, South Korea.

60 Department of Gastroenterology, Kaiser Permanente San Francisco Medical Center, San Francisco, California, USA.

61 Division of Research, Kaiser Permanente Northern California, Oakland, California, USA.

62 Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany.

63 University Medical Centre Hamburg-Eppendorf, University Cancer Centre Hamburg (UCCH), Hamburg, Germany.

64 Department of Cancer Biomedical Science, Graduate School of Cancer Science and Policy, National Cancer Center, Gyeonggi-do, South Korea.

65 Department of Medicine I, University Hospital Dresden, Technische Universität Dresden (TU Dresden), Dresden, Germany.

66 Department of Pathology, University of Michigan, Ann Arbor, Michigan 48104, USA.

67 Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia.

68 Department of Epidemiology, School of Public Health and Institute of Health and Environment, Seoul National University, Seoul, South Korea.

69 Slone Epidemiology Center, School of Medicine, Boston University, Boston, Massachusetts.

70 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA.

71 Division of Molecular and Clinical Epidemiology, Aichi Cancer Center Research Institute, Nagoya, Japan.
Laboratory of Clinical Genome Sequencing, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, University of Tokyo, Tokyo, Japan.

Institute for Health Promotion, Graduate School of Public Health, Yonsei University, Seoul, Korea.

Department of Family Medicine, University of Virginia, Charlottesville, Virginia, USA.

University of Hawaii Cancer Center, Honolulu, Hawaii, USA.

Department of Molecular Biology of Cancer, Institute of Experimental Medicine of the Czech Academy of Sciences, Prague, Czech Republic.

Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University, Prague, Czech Republic.

Faculty of Medicine and Biomedical Center in Pilsen, Charles University, Pilsen, Czech Republic.

Department of Gastroenterology, Biodonostia Health Research Institute. Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd). Universidad del País Vasco (UPV/EHU). San Sebastián, Spain

Nutrition and Metabolism Section, International Agency for Research on Cancer, World Health Organization, Lyon, France.

Moffitt Cancer Center, Tampa, FL, USA

Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia.

University of Melbourne Centre for Cancer Research, Victorian Comprehensive Cancer Centre, Parkville, Victoria 3000 Australia

Department of Internal Medicine, University of Utah, Salt Lake City, Utah, USA.

Department of Medicine and Surgery, LUM University, Camassima, Italy
86 Gastrointestinal Genetics Lab, CIC bioGUNE-BRTA, Derio, Spain

87 Department of Environmental Genomics, School of Public Health, Nanjing Medical University, Nanjing, China.
88 Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany.

89 Memorial University of Newfoundland, Discipline of Genetics, St. John's, Canada.

90 Departments of Epidemiology and Nutrition, Harvard TH Chan School of Public Health, Boston, Massachusetts, USA.

91 Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA.

92 Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA.

93 Division of Epidemiology, National Cancer Center Institute for Cancer Control, National Cancer Center, Tokyo, Japan

94 Division of Cohort Research, National Cancer Center Institute for Cancer Control, National Cancer Center, Tokyo, Japan

95 Department of Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, China.

96 Department of Environmental Health, Harvard T.H. Chan School of Public Health, MA, USA.

97 Division of Cohort Research, National Cancer Center Institute for Cancer Control, National Cancer Center, Tokyo, Japan

98 Department of Molecular Biology of Cancer, Institute of Experimental Medicine of the Czech Academy of Sciences, Prague, Czech Republic.

99 Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University, Prague, Czech Republic.
Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY, USA.

Division of Epidemiology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville Tennessee, USA.

Division of Human Genetics, Department of Internal Medicine, The Ohio State University Comprehensive Cancer Center, Columbus, Ohio, USA.

Department of Laboratory Medicine and Pathology, Mayo Clinic Arizona, Scottsdale, Arizona, USA.

Department of Medicine and Epidemiology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania, USA.

Samuel Oschin Comprehensive Cancer Institute, CEDARS-SINAI, Los Angeles, California, USA.

Department of Public Health, Erasmus University Medical Center, Rotterdam, The Netherlands

Nantes Université, CHU Nantes, Service de Génétique Médicale, F-44000 Nantes, France.

Gastroenterology Department, Hospital Clínic, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), University of Barcelona, Barcelona, Spain.

Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA.

National University Cancer Institute, Singapore.

Center for Cancer Research, Medical University Vienna, Vienna, Austria.
112 Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA.

113 Department of Epidemiology and Health Promotion, Graduate School of Public Health, Yonsei University, Seoul, Korea.

114 Department of Preventive Medicine, Chonnam National University Medical School, Gwangju, Korea.

115 Jeonnam Regional Cancer Center, Chonnam National University Hwasun Hospital, Hwasun, Korea.

116 Division of Epidemiology, National Cancer Center Institute for Cancer Control, National Cancer Center, Tokyo, Japan

117 Center for Gastrointestinal Biology and Disease, University of North Carolina, Chapel Hill, North Carolina, USA.

118 Department of Molecular Biology of Cancer, Institute of Experimental Medicine of the Czech Academy of Sciences, Prague, Czech Republic.

119 Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University, Prague, Czech Republic.

120 Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany.

121 State Key Laboratory of Oncology in South China, Cancer Center, Sun Yat-sen University, Guangzhou, China.

122 Vanderbilt University Medical Center, Nashville Tennessee, USA.

123 Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, South Korea.

124 Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York, USA.
Department of Radiation Sciences, Oncology Unit, Umeå University, Umeå, Sweden.

Wallenberg Centre for Molecular Medicine, Umeå University, Umeå, Sweden.

Huntsman Cancer Institute and Department of Population Health Sciences, University of Utah, Salt Lake City, Utah, USA.

Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA.

Department of Public Health, Erasmus MC University Medical Center, Rotterdam, the Netherlands.

Oncology Data Analytics Program, Catalan Institute of Oncology-IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain.

CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain.

Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, Barcelona, Spain.

ONCOBEL Program, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Barcelona, Spain.

Department of Medicine, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA.

Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, California, USA.

Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, USA.

Department of Public Health, Erasmus MC, University Medical Center, Rotterdam, The Netherlands.
138 Department of Medical Oncology & Therapeutics Research, City of Hope National Medical Center, Duarte CA, USA.

139 Division of Epidemiology, Department of Population Health, New York University School of Medicine, New York, New York, USA.

140 Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.

141 Division of Epidemiology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, Tennessee, USA.

142 Department of Gastroenterology, Kaiser Permanente Medical Center, San Francisco, California, USA.

143 Department of Biostatistics, University of Washington, Seattle, Washington, USA

**Corresponding Authors:**

Li Hsu: 1100 Fairview Ave. N., M2-B500, Fred Hutchinson Cancer Center, Seattle, WA 98109. Email: lih@fredhutch.org; (P) 206-604-5136.

Ulrike Peters: 1100 Fairview Ave. N., M4-B402, Fred Hutchinson Cancer Center, Seattle, WA 98109. Email: upeters@fredhutch.org; (P) 206-667-2450
Abstract

Polygenic risk scores (PRS) have great potential to guide precision colorectal cancer (CRC) prevention by identifying those at higher risk to undertake targeted screening. However, current PRS using European ancestry data have sub-optimal performance in non-European ancestry populations, limiting their utility among these populations. Towards addressing this deficiency, we expanded PRS development for CRC by incorporating Asian ancestry data (21,731 cases; 47,444 controls) into European ancestry training datasets (78,473 cases; 107,143 controls). The AUC estimates (95% CI) of PRS were 0.63(0.62-0.64), 0.59(0.57-0.61), 0.62(0.60-0.63), and 0.65(0.63-0.66) in independent datasets including 1,681-3,651 cases and 8,696-115,105 controls of Asian, Black/African American, Latinx/Hispanic, and non-Hispanic White, respectively. They were significantly better than the European-centric PRS in all four major US racial and ethnic groups (p-values<0.05). Further inclusion of non-European ancestry populations, especially Black/African American and Latinx/Hispanic, is needed to improve the risk prediction and enhance equity in applying PRS in clinical practice.

Introduction

Colorectal cancer (CRC) is a leading cause of cancer death, yet it is among the most preventable cancers via screening\(^1\). Together with detection of CRC at early stages, which dramatically improves prognosis, optimal screening has the potential for major impact on CRC mortality. However, current screening programs are primarily age and family-history based and more
refinement through risk-based screening recommendations could be instrumental in improving their effectiveness.

Genetics plays a key role in the CRC development and, as for most cancers and other common diseases, the risk is polygenic\(^2\). As such, we can utilize the polygenic risk structure to develop a polygenic risk score (PRS) to quantify an individual’s inherited risk of developing CRC. As the predictive performance improves, a PRS can become clinically useful as a risk stratification tool for targeted screening and chemoprevention. However, PRS built based on European ancestry data have sub-optimal performance in other ancestral populations\(^3\) because of differential linkage disequilibrium (LD) patterns and allele frequencies across racial and ethnic groups for disease risk variants of CRC\(^4\)–\(^9\). The poor transferability of PRS across racial and ethnic groups has raised concern regarding whether its application in clinical practice may exacerbate existing health disparities\(^7\). As a result, there is a need to improve the accuracy of polygenic prediction across different racial and ethnic groups to maximize the clinical and public-health translational potential of PRS and enhance equity in precision medicine.

Developing ancestry-specific PRS requires sufficient sample sizes for each ancestral group; however, the sample sizes for non-European ancestry groups, while increasing, remain only a fraction of the sample size for European ancestry. Existing studies suggest that leveraging information from other ancestries can improve ancestry-specific PRS \(^10,11\). As an alternative to developing ancestry-specific PRS, one may develop a single cross-ancestry PRS based on meta-analysis of genome-wide association studies (GWAS) across all available ancestral groups\(^12\)–\(^14\). To our knowledge, there is no study of PRS for non-European ancestral populations for CRC. Here we consider two different approaches to PRS development, (1) ancestry-specific PRS using PRS-
CSx\textsuperscript{15} based on ancestry-specific GWAS while leveraging cross-ancestry information and (2) single cross-ancestry Asian-European PRS using LDpred\textsuperscript{216} based on combined meta-analysis summary statistics and LD matrices across Asian and European ancestries. Using independent racially and ethnically diverse datasets, we evaluated the performance of these two PRS and compared them with a genome-wide PRS built using European-only GWAS data\textsuperscript{3} and a PRS based on 204 known CRC loci\textsuperscript{17–20}. To facilitate understanding of its clinical utility, we used decision-curve analyses\textsuperscript{21} to assess the standardized net benefit for the model based on family-history and PRS and compared to the family-history-only model, as the latter is currently used to decide at what age screening starts.

**Results**

For developing PRSs, we used GWAS summary statistics of 1,020,293 SNPs based on 21,731 cases and 47,444 controls of Asian and 78,473 cases and 107,143 controls of European ancestries. We evaluated the performance of the PRS in independent validation individual-level data sets including 12,025 Asian (2,420 cases; 9,605 controls), 13,823 Black/African-American (1,954 cases; 11,869 controls), 10,378 Latinx/Hispanic (1,682 cases; 8,696 controls) and 118,756 non-Hispanic White (3,651 cases; 115,105 controls) participants. More details about study participant characteristics for training and validation data sets are included in Table 1, Supplemental Table 1, and Supplemental Material and Methods.

*Discriminatory accuracy of Asian-European PRS*

The single cross-ancestry Asian-European PRS derived using the combined Asian-European GWAS meta-analysis summary statistics and LD matrices with LDpred2 improved the
discriminatory accuracy in the Asian population compared to the European-centric PRS (AUC = 0.63 vs. 0.59, p-value < 4.5e-09, Table 2). It also improved the AUC significantly in the non-Hispanic White population (AUC = 0.65 vs. 0.63, p-value = 6.0e-03). Despite lack of Black/African American and Hispanic individuals in deriving the PRS, the Asian-European PRS improved the AUC for Black/African American (AUC = 0.59 vs. 0.58, p-value = 0.05) and Hispanic individuals (AUC = 0.62 vs. 0.59, p-value = 5.0e-03). The Asian-European PRS improved the AUC in all racial and ethnic groups compared to the known-loci PRS (all p-values < 0.05).

The ancestry-specific PRS derived using PRS-CSx improved the discriminatory accuracy in the Asian population compared to the European-centric PRS (AUC = 0.64 vs. 0.59), though not statistically significant with p-value 0.06 (Table 2). The AUC for the ancestry-specific non-Hispanic White-specific PRS was also not statistically different from the European-centric PRS (p-value = 0.15) in the non-Hispanic White population; however, it was significantly higher than the known-loci PRS (p-value = 1.8e-05). The ancestry-specific PRS-CSx is not relevant for Black/African American and Hispanic groups, because there were no GWAS for these groups included in the training datasets.

There was little variation in AUC estimates across studies (Supplemental Table 2). Among these two approaches, the Asian-European PRS using the combined Asian-European summary statistics in LDpred2 had greater discriminatory accuracy than the ancestry-specific non-Hispanic White-specific PRS from PRS-CSx with p-value = 3.0e-03. However, we did not observe statistically significant differences in Asian individuals (p-value = 0.75). Taken together, the single cross-ancestry Asian-European PRS using LDpred2 performs among the best in terms of AUC but with
much narrower confidence intervals; hereafter we focus only on the single cross-ancestry Asian-European PRS. The ROC curves for the cross ancestry Asian-European PRS showed a similar pattern to the AUC for Asian, Black/African American, Hispanic, and non-Hispanic White participants (Supplemental Figure 1).

**PRS distribution across racial and ethnic groups**

As expected, the PRS distributions varied across the racial and ethnic groups (Figure 1(A) and Supplemental Figure 2). After trans-ancestry correction, the PRS distributions largely overlapped except for the MG-JPN study (Figure 1(B) and Supplemental Figure 3). This may be due to the use of the imputation reference panel of only Asian individuals from the 1000 Genomes Projects for MG-JPN; this differs from all other studies, which used all 1000 Genome Project samples in the reference panel. We thus performed an additional mean adjustment to the PRS for the MG-JPN study. After this adjustment, all PRS distributions overlapped (Figure 1(C)).

Cases had higher mean PRS than controls across all racial and ethnic groups (Supplemental Figure 4). The OR estimates per SD of PRS (95% CI) were 1.64 (1.55-1.74), 1.39 (1.31-1.47), 1.62 (1.51-1.73) and 1.67 (1.60-1.75) for Asian, Black/African American, Latinx/Hispanic, and non-Hispanic White participants, respectively, with p-value < 2.0e-18 for all four groups (Figure 1(D) and Table 3).

Compared to the mean risk, the relative risks of PRS at any given percentile were similar for all racial and ethnic groups except for Black/African American participants for whom it was attenuated (Figure 2). The relative risk at the 90th percentile of the PRS distribution compared to
mean was 1.67, 1.44, 1.65, and 1.69 for Asian, Black/African American, Latinx/African American, and non-Hispanic participants, respectively.

The model-based relative risk was calibrated well across the PRS range in all racial and ethnic groups (Figure 3).

**Odds ratios (ORs) for PRS stratified by family-history and age**

Across all racial and ethnic groups, the ORs for the PRS were higher in those without a family-history than those with a family-history with p-values 0.21, 0.01, 3.0e-3, and 0.11 for Asian, Black/African American, Latinx/Hispanic, and non-Hispanic White participants respectively (Table 3). The estimates were consistent across studies (Supplemental Table 3).

The strength of association estimates for PRS in relation to CRC decreased over strata of increased age in each racial and ethnic group with trend test p-values of 0.07, 0.11, 2.8e-4, and 1.2e-03 for Asian, Black/African American, Latinx/Hispanic, and non-Hispanic White, participants, respectively. The ORs, 95% CI and trend p-value for each racial and ethnic group are given in Table 3. The estimates were consistent across studies (Supplemental Table 3).

**Clinical utility for model based on PRS and family-history**

We calculated the standardized net benefit (sNB) to assess the clinical utility of using a model based on PRS and family-history to recommend an intervention (such as screening) for participants < 50 years of age. We used the average 10-year risk of developing CRC at age 45 as the risk threshold, because the current CRC-screening guidelines recommend that an average-risk individual start screening at age 45 years old. Using the GERA cohort, we estimated the 10-year
risk to be 0.29% across all racial and ethnic groups. At this risk threshold, the risk model based on PRS, and family-history achieved 37.3% (95% CI: 23.8%-50.8%) of the maximum possible achievable utility. This was greater than the model based on family-history alone (sNB = 21.7%, 95% CI: 12.4%-33%, p-value 0.02) and hypothetically intervening on all or no people (Figure 4), a pattern that generally holds for each racial and ethnic group (Supplemental Figure 5).

We observed a similar pattern for participants between the ages of 50 and 60 years (Supplemental Figure 6). We also used the 10-year risk 0.39% at age 50 and 0.49% at 55 years as the risk thresholds. The risk model based on PRS, and family history achieved greater sNB (sNB = 24.8% and 21.6%, respectively) than the model based on family history-alone (sNB = 19.3% and 15.9%, respectively).

**Discussion**

Using large-scale Asian and European GWAS data, we demonstrate that combining Asian and European summary statistics in deriving PRS led to statistically significant improvement in discriminatory accuracy across Asian, Black/African American, Latinx/Hispanic and non-Hispanic White groups, although the improvement was less marked in Latinx/Hispanic and Black/African American participants. We further show that across all groups, the PRS has stronger associations with CRC-risk in younger individuals and in those without a family-history of CRC, which will likely increase the possible clinical utility of the PRS given the rising young-onset CRC incidence rates in recent decades, mostly in individuals without a known family-history. This is supported by our decision-curve analysis demonstrating that adding PRS improves the maximum achievable clinical utility over the model based on family-history only for ages 40-60 years.
A challenging factor of moving PRS to clinical implementation is ensuring that the PRS is equally applicable to individuals across all racial and ethnic groups to prevent an increase in health disparities. Relevant to this objective, we evaluated two broad categories of approaches (ancestry-specific PRS while leveraging cross ancestry information and single cross-ancestry PRS based on the combined cross-ancestry GWAS) for improving the prediction in under-represented groups, and our observation of the performance of these approaches could be generalized to other traits besides CRC. We found that both approaches performed similarly in Asian and non-Hispanic European individuals. Further, the cross-ancestry Asian-European PRS also improved risk prediction in Black/African American and Hispanic groups, though to a lesser extent. We also show that we can correct this raw PRS for genetic ancestry and create a common distribution that can be used across racial and ethnic groups, avoiding the potential difficulty of using ancestry-specific PRS in admixed populations. Accordingly, our cross-ancestry Asian-European PRS has the potential to reduce health disparities between non-European ancestry populations and the European ancestry population.

As there is growing interest in clinical use of PRS, it is important to point out that the purpose of PRS is not to identify CRC, but rather stratify individuals into different risk strata for which different levels of cancer preventive interventions may be devised. Their performance should thus be compared with risk factors currently used for risk stratification such as family-history in terms of cost effectiveness. In this paper, we performed a decision-curve analysis that has been used in cancer research for assessing the potential population impact of incorporating a risk prediction model into clinical practice. The risk model that incorporates both the PRS and family-history achieves 37.3% of the maximum possible achievable utility for those 40-49 years old, significantly greater than 21.7% under the family-history-only model. Recently the US
Preventive Services Task Force recommended lowering the age at screening initiation to 45 years for individuals at average risk. However, given the substantial burden of additional approximately 22 million people becoming eligible for screening and the fact that CRC remains a rare event in younger individuals, there has been critique of the universal change to the initial screening age that, instead, emphasizes the importance of targeted screening based on an individual’s risk factors. The results from the decision-curve analysis suggest that there is clinical utility to adding a PRS to the family-history-only model in risk stratification for CRC prevention. To fully evaluate the effectiveness of including PRS as part of risk stratification, a full decision analytic modeling that incorporates other aspects such as different screening methods, implementation factors, behavioral factors, and corresponding costs are warranted.

Recent efforts in clinical implementation of PRS shows the potential of PRS to effectively stratify the risk of diseases development and guide screening. BOADICEA v5 (as implemented in the CanRisk tool) already implements a 313-variant PRS of breast cancer and currently supports hundreds of thousands of women, doctors, and genetic counselors annually in >90 countries making treatment decisions. PRS-guided mammographic screening is also being tested in the WISDOM and PERSPECTIVE I&I studies. GenoVA Study is a clinical trial in which patients and their primary care physicians receive a clinical PRS laboratory report on five diseases including CRC. MyOme implements a cross-ancestry risk score for breast cancer risk stratification. As CRC has an effective screening intervention, it would be of great interest to explore implementation of PRS for guiding personal screening recommendations.

This study has several strengths. We brought together most of the globally available GWAS of CRC for Asian and European ancestry populations as our training data, which is an important
factor for the improved performance of the proposed PRS. Further, we used multiple independent evaluation data sets that were not part of our training data nor GWAS discovery, providing an unbiased evaluation of the developed models. Moreover, the single cross-ancestral PRS derived in this study makes it easy to implement in any admixed population.

The results of this investigation should be interpreted in the context of its limitations. The discriminatory accuracy remains lower in Latinx/Hispanic and particularly in Black/African American individuals due to their limited sample sizes in training data. Future studies more inclusive of these individuals are warranted for deriving PRS to enhance the discriminatory accuracy. Furthermore, we have not been able to evaluate the performance of these models in other racial and ethnic groups, including Alaskan Native, Native American and Pacific Islander individuals. Lastly, we expect to further improve risk prediction by combining the PRS with non-genetic risk factors such as obesity, diet, and aspirin use, as previously shown\textsuperscript{23,36}.

Advances in PRS development have promoted the use of PRS-enhanced models to determine and stratify disease risk, which could improve disease prevention and management through screening and early detection. Our cross-ancestry Asian-European PRS, built upon data on both Asian and European ancestry individuals, improves the PRS performance in Asian, Black/African, and Latinx/Hispanic individuals considerably. Combining PRS and other CRC-associated risk factors such as lifestyle/environmental risk factors and high penetrance genes will likely further improve the prediction performance\textsuperscript{36}. We anticipate that the continuous expansion of PRS development and validation to include more diverse populations and prospective evaluation of PRS-enhanced risk prediction model in clinical trials along with decreasing genotyping cost and adaptation of
health care systems to accommodate genetic data and prediction algorithm will bring closer the implementation of PRS in clinical practice.

**Methods and Materials**

**Training Data Sets**

To develop polygenic risk scores (PRS) across population, we used the genome-wide association study (GWAS) summary statistics of 1,020,293 SNPs based on 78,473 cases and 107,143 controls of European (EUR) and 21,731 cases and 47,444 controls of Asian ancestries from previously published GWAS (Supplemental Table 1)\(^{17–19}\). For this we group participants into analytical units by study or genotyping platform as consistent with the original reports\(^{17–20,37,38}\). Ancestry was determined by the genetic principal component analysis. Studies that contributed to more than one prior genome-wide association analyses were analyzed only once. In total, there were 31 analytical units (17 from EUR descent populations and 14 from Asian descent populations), totaling 100,204 CRC cases and 154,587 controls. Comprehensive details on the participants, genotyping and standard quality control (QC) procedures have been previously reported and are also summarized in Supplemental Table 1. All study protocols were approved by the relevant Institutional Review Boards, and informed consent was obtained from all study participants in accordance with the Helsinki accord.

**Independent Validation Data Sets**

We evaluated the performance of each of the developed PRS in the Genetic Epidemiology Research on Adult Health and Aging Cohort (GERA) cohort; Minority GWAS Japanese study (MG-JPN)\(^{39}\); Minority GWAS African American study (MG-AA)\(^{40}\); Hispanic Colorectal Cancer
Study (HCCS); Multiethnic Cohort study (MEC); Cancer Prevention Study II (CPSII); Basque-colon cohort (BCC); and Electronic Medical Records and Genomics (eMERGE) study. Racial and ethnic identification in these studies were self-reported. In total, there were 12,025 Asian (2,420 cases; 9,605 controls), 13,823 Black/African-American (1,954 cases; 11,869 controls), 10,378 Latinx/Hispanic (1,682 cases; 8,696 controls) and 118,756 non-Hispanic White (3,651 cases; 115,105 controls) participants. None of these samples was included in the training data sets for model building. More details about study participant characteristics are included in Table 1.

CRC status (Yes/No) was determined from cancer-registry data. Family-history of CRC (>=1 first-degree relatives with CRC), was ascertained through baseline study questionnaire or electronic medical records at study entry.

**Approaches for Deriving PRS**

We compared two different approaches for PRS development using 1) ancestry-specific PRS using PRS-CSx that integrates genome-wide Asian and European summary statistics and LD matrices; 2) single cross-ancestry PRS using LDpred2 that combine genome-wide Asian and European summary statistics and a weighted LD matrix with weight defined as the proportion of participants from each ancestry in the summary statistics. Figure 5 depicts the summary of these PRS derivations.

PRS-CSx\textsuperscript{15} derives ancestry-specific PRS while leveraging GWAS summary statistics from other ancestral groups. We first obtained ancestry-specific PRS using ancestry-specific GWAS summary statistics and LD matrix for Asian and non-Hispanic White participants based on ~1M genome-wide SNPs, respectively, while leveraging GWAS from the other ancestral group. We
denoted these PRS by PRS$_{\text{Asian}}$ and PRS$_{\text{European}}$, respectively. We then improved ancestry-specific PRS by taking a weighted sum of these PRSs to predict CRC of respective ancestral group. To derive PRS for the Asian population, we calculated a weighted sum of PRS$_{\text{Asian}}$ and PRS$_{\text{European}}$ ($\alpha_1$ PRS$_{\text{European}}$ + $\beta_1$ PRS$_{\text{Asian}}$) and obtained $\alpha_1$ and $\beta_1$ from a logistic regression model using the MG-JPN study. Similarly, to derive PRS for the European population, we calculated a weighted sum of PRS$_{\text{Asian}}$ and PRS$_{\text{European}}$ ($\alpha_2$ PRS$_{\text{European}}$ + $\beta_2$ PRS$_{\text{Asian}}$), where $\alpha_2$ and $\beta_2$ were obtained based on the pooled BCC and CPSII studies.

To derive the single cross-ancestry PRS using LDpred2$^{16}$, we combined the summary statistics from the Asian and European GWAS using the inverse variance weighted estimator$^{43}$ and combined the LD matrices, as the weighted sum of the Asian and European-specific LD matrices with the weights proportional to the sample sizes of the Asian and European individuals in the combined summary statistics.

We compared ancestry-specific and single cross-ancestry PRS from PRS-CSx and LDpred2 with a previously published European-centric genome-wide PRS$^3$ and a known-loci PRS consisting of 204 independently CRC-associated variants based on GWAS of European and Asian ancestries$^{17-20}$ (Supplemental Table 4).

Evaluation of Model Performance

*The Area Under the Receiver Operating Characteristics curve (AUC)*

We evaluated the predictive performance of the PRS by the area under the receiver operating characteristics curve (AUC) in each of the racial and ethnic groups$^{44}$. We calculated the adjusted AUC of PRS for each study using the ROCt R package$^{45}$, adjusting for covariates age, sex and
four PCs. We emphasize that the AUC estimate was for PRS only and the covariates were not part of prediction along PRS. These covariates were included as potential confounders. We then combined the AUC estimates of PRS across studies for each ancestry using the inverse variance weighted estimator.

We obtained the bootstrapped-based standard error (se), 95% confidence intervals (CI) (1.96* se) and two-sided p-values for comparisons across various subgroups using 500 bootstrap samples.

Ancestry Adjustment of PRS Distribution

As the PRS distributions were different across racial and ethnic groups due to different allele frequencies, we used a modified trans-ancestry adjustment of PRS to align the PRS distributions

We used the 1000 Genome dataset to estimate the ancestry adjustment following the approach in Khera et al. (2019).

Specifically, we derived principal components (PCs) based on 343,662 ancestry informative SNPs with little overlapped (0.3%) with SNPs used in PRS development. To correct for the mean and variance differences between ancestry groups, we fit two linear regression models to predict the mean and variance of PRS based on the first four PCs. To correct for the raw PRS distribution in our data set, we first calculated the PCs using the same loadings for the top 4 PCs from the 1000 Genome data set. We then obtained the ancestry-adjusted PRS for each individual by subtracting the predicted mean based on the 4 PCs from the individual’s raw PRS and then divided it by the predicted standard deviation based on the 4 PCs. Additional adjustments are needed for data sets with different imputation panels. The ancestry adjusted PRS is computed as given below:
\[ PRS_{adjusted} = \frac{PRS_{sample} - (\alpha_o + \sum \alpha_i PC_i)}{\sqrt{\beta_o + \sum \beta_i PC_i}} \]

**Odds Ratio (OR) Estimates**

We estimated the OR and 95% CI of CRC-risk associated per SD change in PRS by logistic regression model, overall and stratified by family history and age. For each racial and ethnic group, we estimated the AUC and OR by study and combined the estimates using the inverse variance weighted estimator. In addition, we estimated OR stratified by family history of 1st degree relative with CRC (yes, no) and age (<50, 50-59, 60-69, 70-79, and >80). All analyses were adjusted for age, sex, and top 4 principal components of ancestry.

**Relative Risk Calibration of PRS**

We binned PRS into 5% strata and defined the reference group as PRS in the 40%-60% stratum. The expected OR for a PRS stratum is the ratio of the within-stratum geometric average of individuals’ model-based OR, defined as exponent of individuals’ PRS times log (OR), between that stratum and the reference stratum. We estimated the observed OR estimates and its 95% CI by fitting a logistic regression model with CRC disease status as outcome and a binary variable with 1 indicating a specific stratum and 0 indicating the reference stratum, adjusting for age, sex, and first four principal components.

**Decision Curve Analysis**

The decision-curve analysis was performed by calculating the standardized net benefit (sNB), defined as the net benefit divided by the maximum possible net benefit, to assess the potential
clinical impact of the risk prediction models on recommended interventions (i.e., screening). For a given risk threshold, the NB was defined as sensitivity × p – (1 – specificity) × (1 – p) × w, where w was the odds at the threshold, sensitivity was the proportion of cases above the risk threshold based on the model, specificity was the proportion of controls below the risk threshold based on the model, and p was the disease probability at the landmark time. As it was difficult to interpret NB itself, we followed the approach proposed by Kerr et al.²¹ to calculate sNB, i.e., dividing NB by the maximum NB, which is achieved when sensitivity = 1 and specificity = 1. Hence, the sNB was equal to sensitivity – (1 – specificity) × (1 – p)/p × w. It provided some sense of magnitude of sNB on a percent scale and was interpreted as the relative utility that has maximum value of 1. For example, if sNB = 0.4, it means that the risk model achieves 40% of the maximum possible achievable utility.

We compared the model based on PRS and family history with the model based on family history alone, as well as two hypothetical extreme scenarios: intervention (e.g., screening) for all and intervention for none. We calculated the sNB under the competing risks framework⁴⁷, where the observational time is the minimum of time to CRC, time to death, and time at last observation, and the disease status is 1 if the study participant had CRC, 2 if the participant died (competing event), and 0 otherwise. We used GERA to calculate sNB, as this was the only cohort study among our independent validation data sets. We plotted decision-curves of sNB at the 10-year landmark time vs. risk threshold for age at study entry 40-49 and 50-59 years old, because average-risk individuals in these age groups are recommended to start CRC screening.

We performed the analyses using R packages²⁴,⁴⁵,⁴⁸–⁵⁰. A two-sided p-value < 0.05 is considered statistically significant.
Figures

**Figure 1:** Distribution of PRS: A) PRS distributions varied across racial/ethnic groups, B) PRS distribution after ancestry adjustment, C) Additional mean adjustment for the Asian MG (Minor GWAS Japanese Study) study that has a different imputation panel, and D) forest plot by racial and ethnic group for OR estimates of PRS per SD. PRS is based on single cross-ancestry Asian-European PRS.

**Figure 2:** The relative risk of individuals at different percentiles of the single cross-ancestral Asian-European PRS compared to a population average odds ratio, stratified by race and ethnicity.

**Figure 3:** The relative risk calibration of PRS, stratified by race and ethnicity. The x-axis is the log-transformed predicted RR values and the y-axis is the log-transformed observed RR with the middle bin (40-60) as the reference group.

**Figure 4:** Decision-curves for the model based on family-history and PRS, model based on family-history alone, intervene for all, and intervene for none, in GERA study participants with age 40-49 (22,628 participants with 149 cases). X-axis is the range of threshold probabilities (CRC-risk). The legends of the lines are solid black: intervene for none (none); solid gray: intervene for all (All); dash green: model based on family-history (Famhx); dash blue: model based on family-history and PRS (Famhx + PRS).

**Figure 5:** Different approaches for deriving polygenic risk scores (PRS) for colorectal cancer.
Tables

**Table 1: Characteristics of the validation studies**

<table>
<thead>
<tr>
<th>Study</th>
<th>Racial and Ethnic Group</th>
<th>Total N</th>
<th>CRC No. (%)</th>
<th>Mean age (range)</th>
<th>Female No. (%)</th>
<th>Family-History</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Yes No. (%)</td>
</tr>
<tr>
<td>GERA</td>
<td>Asian</td>
<td>7,370</td>
<td>96 (1.0%)</td>
<td>64 (19-95)</td>
<td>4,152 (56.3%)</td>
<td>643 (8.7%)</td>
</tr>
<tr>
<td></td>
<td>Black or African American</td>
<td>3,159</td>
<td>56 (2.0%)</td>
<td>66 (20-95)</td>
<td>1,811 (57.3%)</td>
<td>313 (9.9%)</td>
</tr>
<tr>
<td></td>
<td>Latinx or Hispanic</td>
<td>6,660</td>
<td>70 (1.0%)</td>
<td>63 (19-95)</td>
<td>4,081 (61.3%)</td>
<td>543 (8.2%)</td>
</tr>
<tr>
<td></td>
<td>non-Hispanic White</td>
<td>77,012</td>
<td>1,401 (1.8%)</td>
<td>70 (19-95)</td>
<td>44,125 (57.2%)</td>
<td>7,423 (9.6%)</td>
</tr>
<tr>
<td>MG</td>
<td>Japanese</td>
<td>4,655</td>
<td>2,324 (50.0%)</td>
<td>65 (20-90)</td>
<td>2,007 (43.1%)</td>
<td>407 (8.7%)</td>
</tr>
<tr>
<td></td>
<td>Black or African American</td>
<td>6,597</td>
<td>1,856 (28.1%)</td>
<td>66 (20-91)</td>
<td>2,581 (39.1%)</td>
<td>721 (11.0%)</td>
</tr>
<tr>
<td>Hispanic GWAS</td>
<td>Latinx or Hispanic</td>
<td>3,717</td>
<td>1,611 (43.3%)</td>
<td>65 (21-90)</td>
<td>1,790 (48.1%)</td>
<td>260 (7.0%)</td>
</tr>
<tr>
<td></td>
<td>non-Hispanic White</td>
<td>1,712</td>
<td>804 (46.9%)</td>
<td>71 (54-90)</td>
<td>769 (44.9%)</td>
<td>204 (11.9%)</td>
</tr>
<tr>
<td>CPSII</td>
<td>non-Hispanic White</td>
<td>1,818</td>
<td>873 (48.0%)</td>
<td>57 (8-99)</td>
<td>627 (34.4%)</td>
<td>0</td>
</tr>
<tr>
<td>BCC</td>
<td>non-Hispanic White</td>
<td>3,821</td>
<td>573 (1.5%)</td>
<td>65 (18-90)</td>
<td>20,543 (54.0%)</td>
<td>965 (2.5%)</td>
</tr>
<tr>
<td>eMERGE</td>
<td>Black or African American</td>
<td>4,067</td>
<td>42 (1.0%)</td>
<td>52 (18-90)</td>
<td>2,946 (72.0%)</td>
<td>105 (2.6%)</td>
</tr>
<tr>
<td></td>
<td>non-Hispanic White</td>
<td>38,214</td>
<td>573 (1.5%)</td>
<td>65 (18-90)</td>
<td>20,543 (54.0%)</td>
<td>965 (2.5%)</td>
</tr>
</tbody>
</table>

**Table 2: AUC estimates (95% confidence interval) for European-centric PRS, known loci PRS, PRS-CSx and LDpred2.**

<table>
<thead>
<tr>
<th>Race and Ethnicity</th>
<th>Cases/ controls</th>
<th>European-centric PRS</th>
<th>Known Loci PRS</th>
<th>PRS-CSx</th>
<th>LDpred2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asian</td>
<td>2,420/9,605</td>
<td>0.59 (0.57-0.60)</td>
<td>0.60 (0.59-0.62)</td>
<td>p-val*: 0.24</td>
<td>0.63 (0.62-0.64)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>1,954/11,869</td>
<td>0.58 (0.56-0.59)</td>
<td>0.58 (0.56-0.59)</td>
<td>p-val*: 0.92</td>
<td>0.59 (0.57-0.61)</td>
</tr>
</tbody>
</table>
### Table 3. Odds ratios (OR), 95% confidence interval (95% CI) and p-values for PRS per SD and PRS stratified by family-history and age.

<table>
<thead>
<tr>
<th></th>
<th>Asian</th>
<th>Black or African American</th>
<th>Latinx or Hispanic</th>
<th>Non-Hispanic White</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PRS per SD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OR (95% CI)</td>
<td>p-value</td>
<td>OR (95% CI)</td>
<td>p-value</td>
</tr>
<tr>
<td></td>
<td>&lt;2.00e-18</td>
<td></td>
<td>&lt;2.00e-18</td>
<td></td>
</tr>
<tr>
<td><strong>Family-History</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1.62 (1.46-1.80)</td>
<td>&lt;2.0e-18</td>
<td>1.53 (1.26-1.85)</td>
<td>1.21e-5</td>
</tr>
<tr>
<td>Yes</td>
<td>1.45 (1.19-1.76)</td>
<td>1.22 (1.03-1.44)</td>
<td>1.15 (0.91-1.46)</td>
<td>0.245</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 50</td>
<td>1.88 (1.50-2.35)</td>
<td>1.51 (1.17-1.94)</td>
<td>1.17 (0.68-2.00)</td>
<td>1.85 (1.48-2.31)</td>
</tr>
<tr>
<td>50 to 60</td>
<td>1.85 (1.62-2.12)</td>
<td>1.53 (1.36-1.72)</td>
<td>2.15 (1.77-2.562)</td>
<td>1.75 (1.567-1.96)</td>
</tr>
<tr>
<td>60 to 70</td>
<td>1.58 (1.43-1.74)</td>
<td>1.41 (1.28-1.55)</td>
<td>1.58 (1.42-1.76)</td>
<td>1.88 (1.73-2.04)</td>
</tr>
<tr>
<td>70 to 80</td>
<td>1.57 (1.41-1.75)</td>
<td>1.31 (1.17-1.45)</td>
<td>1.47 (1.30-1.67)</td>
<td>1.61 (1.49-1.74)</td>
</tr>
<tr>
<td>&gt; 80</td>
<td>1.65 (1.28-2.13)</td>
<td>1.32 (1.03-1.69)</td>
<td>1.58 (1.211-2.05)</td>
<td>1.43 (1.30-1.58)</td>
</tr>
<tr>
<td><strong>Trend p-value (Age)</strong></td>
<td>7.0e-2</td>
<td>0.11</td>
<td>2.8e-4</td>
<td>1.2e-3</td>
</tr>
</tbody>
</table>

*p-value comparison of PRS with the European-centric PRS.
+p-value comparison of PRS with the known-loci PRS.
**p-value comparison of PRS with the PRS-CSx.
** All AUC estimates were adjusted for age, sex, and top 4 principal components.
REFERENCES


This graph illustrates the standardized net benefit across different threshold probabilities for various combinations of factors:

- **None**
- **All**
- **Famhx**
- **Famhx+PRS**

The x-axis represents the threshold probability, while the y-axis shows the standardized net benefit. The lines depict the net benefit for each combination, allowing for a visual comparison of their effectiveness as the threshold probability changes.
**Summary Statistics**
- 204 published CRC loci

**LD Matrix**
- Overlapping SNPs with HapMap3 SNPs as suggested in PMID: 33326037
- (N ~ 1 million SNPs)

**Model Development**
- EAS: \( \text{outcome} \sim b_0 + \alpha_1 \text{PRS}_{w} + \beta_1 \text{PRS}_{r} + \text{covariates} \)
- EUR: \( \text{outcome} \sim b_0 + \alpha_2 \text{PRS}_{w} + \beta_2 \text{PRS}_{r} + \text{covariates} \)

**Validation Data**
- GERA, BCC, CPSII, Hispanic GWAS, MG-JPN, MG-AA and eMERGE: 9,706 cancer cases and 145,275 controls;
- Cancer cases: 2,420 Asian, 1,954 Black or African American, 1,681 Latinx or Hispanic and 3,651 Non-Hispanic White participants

**PRS**
- \( \text{PRS} = \sum_{i=1}^{K} \frac{x_i}{1-x_i} \)

**Test Data Sets**
- EAS: 2324 Cases and 2331 Controls
- EUR: 1,677 Cases and 1,853 Controls

**Leveraging PRS from another ancestry (Approach 1)**
- EAS: PRS\(_{w}, \text{PRS}_{r}\)
- EUR: PRS\(_{w}, \text{PRS}_{r}\)