Polygenic risk scores for prediction of breast cancer risk in Asian populations

Weang-Kee Ho[1,2*], Mei-Chee Tai[2*], Joe Dennis[3], Xiang Shu[4,5], Jingmei Li[6,7], Peh Joo Ho[7], Iona Y Millwood[8,9], Kuang Lin[8], Yon-Ho Jee[10], Su-Hyun Lee[11], Nasim Mavaddat[3], Manjeet K. Bolla[3], Qin Wang[3], Kyriaki Michailidou[3,12,13], Jirong Long[4], Eldarina Azfar Wijaya[2], Tiara Hassan[2], Kartini Rahmat[14], Veronique Kiak Mien Tan[15,16], Benita Kiat Tee Tan[15,16,17], Su Ming Tan[18], Ern Yu Tan[19], Swee Ho Lim[20], Yu-Tang Gao[21], Ying Zheng[22], Daehee Kang[23,24], Ji-Yeob Choi[24,25,26], Wonshik Han[24,27], Han-Byoel Lee[24,27], Michiki Kubo[28], Yukinori Okada[29-31], Shinichi Namba[32], The Biobank Japan Project[33] , Sue K. Park[23,24,34], Sung-Won Kim[35], Chen-Yang Shen[36], Pei-Ei Wu[36], Boyoung Park[37], Kenneth Muir[38], Artitaya Lophatananon[38], Anna H. Wu[39], Chiu-Chen Tseng[39], Keitaro Matsuo[40,41], Hidemi Ito[42,43], Ava Kwong[44,45,46], Tsun L. Chan[44,47], Esther M. John[48,49], Allison W. Kurian[48,49], Motoki Iwasaki[50], Taiki Yamaji[50], Sun-Seog Kweon[51,52],Kristan J. Aronson[53], Rachel A Murphy[54,55], Woon-Puay Koh[56,57], Chiea-Chuen Khor[58], Jian-Min Yuan[59,60], Rajkumar Dorajoo[56,58], Robin G. Walters[8,9], Zhengming Chen[8,9], Liming Li[61], Jun Lv[61], Keum-Ji Jung[62], Peter Kraft[10,63], Paul D.B. Pharoah[3,64], Alison M. Dunning[64], Jacques Simard[65], Xiao Ou Shu[4], Cheng-Har Yip[66], Aishah Mohd Taib[67], Antonis C. Antoniou[3], Wei Zheng[4], Mikael Hartman[6,57,68], Douglas F. Easton[3,64†], Soo-Hwang Teo[2,67†]

* Joint first authors

† These authors jointly supervised this work

Correspondents: Weang-Kee Ho (WeangKee.Ho@nottingham.edu.my) and Soo-Hwang Teo (soohwang.teo@cancerresearch.my)

[1]School of Mathematical Sciences, Faculty of Science and Engineering, University of Nottingham Malaysia, Jalan Broga, Semenyih, 43500 Selangor, Malaysia.

[2]Cancer Research Malaysia, 1 Jalan SS12/1A, Subang Jaya, 47500 Selangor, Malaysia.

[3]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, CB1 8RN Cambridge, UK.

[4]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University, Medical Center, Nashville, TN37232, USA.

[5]Department of Epidemiology & Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY USA

[6]Department of Surgery, University Surgical Cluster, National University Hospital, 1E Kent Ridge Rd, 119228 Singapore.

[7]Genome Institute of Singapore, Laboratory of Women's Health and Genetics, 60 Biopolis St, 138672 Singapore.

[8]Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK.

[9]MRC Population Health Research Unit, University of Oxford, Oxford OX3 7LF, UK

[10]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

[11]Graduate School of Public Health, Yonsei University, Seoul, Korea

[12]Biostatistics Unit, The Cyprus Institute of Neurology & Genetics, 6 Iroon Avenue, 2371 Ayios Dometios, Cyprus

[13]Cyprus School of Molecular Medicine, The Cyprus Institute of Neurology & Genetics, 6 Iroon Avenue, 2371 Ayios Dometios, Cyprus. 43

[14]Biomedical Imaging Department, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia

[15]Department of Breast Surgery, Singapore General Hospital, Singapore, Singapore

[16]Division of Surgical Oncology, National Cancer Centre Singapore, Singapore, Singapore

[17]Department of General Surgery, Sengkang General Hospital, Singapore, Singapore

[18]Division of Breast Surgery, Changi General Hospital, Singapore, Singapore

[19]Department of General Surgery, Tan Tock Seng Hospital, Singapore 308433, Singapore

[20]KK Breast Department, KK Women's and Children's Hospital, Singapore 229899, Singapore

[21] State Key Laboratory of Oncogene and Related Genes & Department of Epidemiology, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China.

[22]Shanghai Municipal Center for Disease Control and Prevention, Shanghai, China.

[23]Department of Preventive Medicine, Seoul National University College of Medicine, 103 Daehak-ro, Jongno-gu, 03080 Seoul, Korea.

[24]Cancer Research Institute, Seoul National University, 103 Daehak-ro, Jongno-gu, Seoul 03080, Korea.

[25]Department of Biomedical Sciences, Seoul National University Graduate School, 103 Daehak-ro, Jongno-gu, 03080 Seoul, Korea.

[26]Institute of Health Policy and Management, Seoul National University, Medical Research Center, 103 Daehak-ro, Jongno-gu, Seoul, 03080 Korea.

[27]Department of Surgery, Seoul National University College of Medicine, Seoul 03080, South Korea

[28]RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.

[29]Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita 565-0871, Japan.

[30]Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Suita 565-0871, Japan.

[31]Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita 565-0871, Japan.

[32]Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita 565-0871, Japan.

[33]Institute of Medical Science, The University of Tokyo, Tokyo, Japan.
[34]Integrated Major in Innovative Medical Science, Seoul National University College of Medicine, 103 Daehak-ro, Jongno-gu, 03080 Seoul, Korea

[35]Department of Surgery, Daerim Saint Mary's Hospital, 657 Siheung-daero, Daerim-dong, Yeongdeungpo-gu, 07442 Seoul, Korea.

[36]Institute of Biomedical Sciences, Academia Sinica, 115128, Section 2, Academia Rd., Taipei, Taiwan.

[37]Department of Medicine, Hanyang University College of Medicine, Seoul, Korea.

[38]Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, The University of Manchester, Oxford Road, M13 9PL Manchester, UK.

[39]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, 1975 Zonal Ave, Los Angeles, 90033 CA, USA.

[40]Division of Cancer Epidemiology and Prevention, Aichi Cancer Center Research Institute, 1-1 Kanokoden, Chikusa-ku, 464-8681 Nagoya, Japan.

[41]Division of Cancer Epidemiology, Nagoya University Graduate School of Medicine, 65 Tsurumai-cho, Showa-ku, 466-8550 Nagoya, Japan.

[42]Division of Cancer Information and Control, Aichi Cancer Center, Japan

[43]Division of Descriptive Cancer Epidemiology, Nagoya University Graduate School of Medicine, Nagoya University, Japan.

[44]Hong Kong Hereditary Breast Cancer Family Registry, Cancer Genetics Centre, 18 A Kung Ngam Village Road, Happy Valley, Hong Kong.

[45]Department of Surgery, The University of Hong Kong, 102 Pokfulam Road, Pok Fu Lam, Hong Kong.

[46]Department of Surgery, Hong Kong Sanatorium and Hospital, 2 Village Rd, Happy Valley, Hong Kong.

[47]Department of Pathology, Hong Kong Sanatorium and Hospital, 2 Village Rd, Happy Valley, Hong Kong.

[48]Department of Medicine, Division of Oncology, Stanford Cancer Institute, Stanford University School of Medicine, 780 Welch Road, Suite CJ250C, Stanford, 94304 CA, USA.

[49]Department of Epidemiology and Population Health, Stanford University School of Medicine, 259 Campus Drive, Stanford, 94305 CA, USA.

[50]Division of Epidemiology, Center for Public Health Sciences, National Cancer Center, 5-1-1 Tsukiji, Chuo-ku, 104-0045 Tokyo, Japan.

[51]Department of Preventive Medicine, Chonnam National University Medical School, Hwasun, Korea.

[52]Jeonnam Regional Cancer Center, Chonnam National University Hwasun Hospital, Hwasun, Korea.

[53]Department of Public Health Sciences, and Cancer Research Institute, Queen's University, 10 Stuart Street, Kingston, K7L 3N6 ON, Canada.

[54]Cancer Control Research, BC Cancer, 675 West 10th Avenue, Vancouver, BC, Canada.

[55]School of Population and Public Health, University of British Columbia, 2329 West Mall, Vancouver, BC, Canada.

[56]Health Services and Systems Research, Duke-NUS Medical School.

[57]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, 12 Science Drive 2, #10-01, 117549 Singapore.

[58]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore.

[59]Division of Cancer Control and Population Sciences, UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, PA, USA

[60]Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA

[61]Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, Beijing, China

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

[63]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

[64]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, 2 Worts' Causeway, CB1 8RN, Cambridge, UK.

[65]Genomics Center, CHU de Québec-Université Laval Research 2705 Blvd Laurier Québec (Québec) G1V 4G2, Canada.

[66]Sime Darby Medical Centre, 1 Jalan SS12/1A, Subang Jaya, 47500 Selangor, Malaysia.

[67]Department of Surgery, Faculty of Medicine, University of Malaya, Jalan Universiti, Kuala Lumpur 50630 Kuala Lumpur.

[68]Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore.

Abstract

Purpose: Non-European populations are under-represented in genetics studies, hindering clinical implementation of breast cancer polygenic risk scores (PRS). We aimed to develop PRSs using the largest available studies of Asian ancestry and to assess the transferability of PRS across ethnic subgroups. **Methods:** The development dataset comprised 138,309 women from 17 case-control studies. PRSs were generated using clumping+thresholding method,

lasso penalized regression, Empirical Bayes approach, Bayesian polygenic prediction approach or linear combinations of multiple PRSs. These PRSs were evaluated in 89,898 women from three prospective studies (1,592 incident cases). **Results:** The best performing PRS (genome-wide set of single-nucleotide polymorphism (SNPs)) had hazard ratio (HR) per unit standard deviation (SD) of 1.62 (95% CI = 1.46-1.80), and area under the receiver operating curve (AUC) of 0.635 (95%CI = 0.622 – 0.649). Combined Asian and European PRSs (333 SNPs) had HR per SD of 1.53 (95%CI: 1.37-1.71) and AUC of 0.621 (95%CI: 0.608-0.635). The distribution of the later PRS was different across ethnic subgroups, confirming the importance of population-specific calibration for valid estimation of breast cancer risk. **Conclusion:** Leverage information from larger European and smaller Asian datasets can aid the development of PRS, enabling the risk-stratified screening approaches for women of Asian ancestry.

Introduction

Genetic inheritance is an important factor contributing to the likelihood that a woman develops breast cancer[1]. Rare pathogenic variants in several susceptibility genes, including *BRCA1*, *BRCA2* and *PALB2*, confer increased risks of breast cancer[2]; however, much of the genetic variation in risk is polygenic, due to a combination of large numbers of genetic variants each conferring a small increase in risk. The effects of these variants can be summarized as polygenic risk scores (PRS)[3,4] . Mavaddat *et al*[3] developed and validated a 313 variant breast cancer PRS (PRS-313), using data from women of European ancestry in the Breast Cancer Association Consortium (BCAC)[4,5]. The lifetime risk of breast cancer was estimated to be 2.6% for women in the lowest 1% of the PRS-313 distribution and ~32% for women in the highest 1%; the latter group would be classified as at high-risk of developing breast cancer according to the National Institute for Health and Care Excellence (NICE) and other clinical management

guidelines.[3] This demonstrates the potential of PRS to improve quantification of risk and consequently optimize breast cancer screening and prevention strategies[6].

The under-representation of non-European populations in genetic studies, which serve as the foundation of PRS development, has raised questions about the transferability of PRS to other populations. The paucity of data in non-European populations could limit PRS adoption and applicability[7-9], and hence exacerbate health disparities[10]. This is important for ethnic minorities in high income countries, where clinical evaluation of the European 313 variants PRS is already underway, but perhaps more so in low- and middle-income countries, where there is an urgent need to develop breast cancer screening strategies to address rapidly rising breast cancer incidence and high breast cancer mortality[11].

Asians constitute more than half of the world's population and are facing a dramatic increase in breast cancer incidence[12,13], but make up only 15% of the breast cancer genome-wide association studies (GWAS). Efforts to develop breast cancer PRS specifically for Asian populations have so far been limited. In our previous work, we showed that the PRS-313, developed in Europeans, was predictive of breast cancer risk in Asian populations, although the effect size was somewhat smaller than that reported in European populations[14]. This demonstrated that PRS can be used in Asian populations to predict risk. However, an important outstanding question is whether a more predictive PRS utilizing Asian data can be developed. Thus far, the largest study to attempt this involved 23,372 women of Asian ancestry. This study evaluated previously published breast cancer risk single-nucleotide polymorphisms (SNPs) and took forward SNPs that were significantly associated with breast cancer risk in Asians (p-value < 0.05) for PRS derivation, resulting in a 44-SNP PRS (PRS-44)[15]. Although predictive, we have shown in our previous work that the discriminatory power of

the PRS-44 (area under the receiver operating curve (AUC) = 0.586) was much lower than the PRS-313 (AUC = 0.617), derived from European ancestry women, for predicting breast cancer risk in Asian women[14].

In this study, our objectives were twofold: (1) to develop improved breast cancer PRSs utilizing data from Asian populations and to validate their performance in prospective cohorts using the largest available breast cancer genetic study of Asian ancestry; (2) to assess the transferability of PRS across Asian ethnic subgroups.

Materials and Methods

Study populations and genotyping

The study population was divided into training, validation and testing datasets. The training datasets included(a) set 1 - 44,126 (22,013 invasive cases and 22,114 controls) women of East Asian ancestry (where summary statistics from a meta-analysis of genome-wide association studies (GWAS) from 16 studies participating in Breast Cancer Association Consortium (BCAC) and Asia Breast Cancer Consortium (ABCC)), (b) set 2 - 100,094 (16,680 invasive cases and 83,414 controls) women of East Asian ancestry (where summary statistics from a meta-analysis of GWAS from 11 studies participating in BCAC and Biobank Japan (BBJ1)), and (c) set 3 - 228,951 (122,977 invasive cases and 105,974 controls) women of European ancestry participating in BCAC (where summary statistics was available); The validation dataset comprised of (a) 13,030 (6,392 cases invasive cases and 6,638 controls) women of Chinese or Malay (East Asian) ancestry, and (b) 1,603 (585 invasive cases and 1,018 controls) women of Indian (South Asian)ancestry participating in two multi-ethnic case-control studies: Malaysian Breast Cancer Genetics (MyBrCa) study and Singapore Breast Cancer Cohort (SGBCC) study.

The testing dataset comprising 89,898 women (1,595 incident cases) from three prospective cohorts of East Asian ancestry: the Singapore Chinese Health Study (SCHS)[16] , China Kadoorie Biobank (CKB)[17] and the Korean Cancer Prevention Study Biobank (KCPS-II)[18].

Samples in the training dataset were genotyped using one of the six arrays: iCOGS[19], OncoArray[20], Affymetrix Genome-Wide Human SNP Array 6.0 , Illumina Multi-Ethnic Genotyping Array, Illumina HumanOmiExpress[21] and Illumina HumanExome-12v1_A Beadchip[22] (Table S1). Samples in the validation dataset were genotyped using the OncoArray. Samples in SCHS and KCPS-II were genotyped using the Illumina Global Screening Array[16], while samples in CKB were genotyped using custom-designed Affymetrix Axiom arrays[17]. Table S1 summarises the study design and the number of breast cancer cases and controls in each study. See Supplementary Methods for more details.

Genotype calling, quality control procedures and imputation methods have been described previously[4,19,22-25]. All data were imputed using the 1000 Genomes Project (Phase 3) data as the reference panel[26], except for a subset of BioBank Japan, for which the HapMap Phase II (release 22)[27] was used. SNPs with overall minor allele frequency in controls > 0.01 and imputation $r^2 > 0.9$ for OncoArray studies, imputation $r^2 > 0.7$ for BBJ1 and imputation $r^2 > 0.3$ for other studies in the training and validation dataset were included in this analysis. Since all samples in the validation sets were genotyped using OncoArray, a higher threshold was imposed for OncoArray to ensure accurate determination of PRS in the validation datasets.

Ancestry informative principal components were available for Asian ancestry samples in the training dataset and the validation dataset, generated using methods as previously described[20]. Briefly, for the BCAC data, continental ancestry was derived by combining the data with the 1000 Genomes Project reference data. Individuals with >40% estimated East

Asian ancestry were retained. In the second stage, principal components were generated on the Asian ancestry individuals using a subset of uncorrelated SNPs. Similar ancestry informative principal components were generated for the other dataset.

Ethics statement

The Malaysian Breast Cancer Genetic Study (MyBrCa) was approved by the Independent Ethics Committee, Ramsay Sime Darby Health Care (reference no: 201109.4 and 201208.1), and the Medical Ethics Committee, University Malaya Medical Centre (reference no: 842.9). Analyses using CKB data were conducted under research approval 2020-0047. Each study listed was approved by the local institutional ethics committees and review boards, and all participants provided written informed consent.

Statistical methods

Polygenic risk scores (PRS) were derived using the following formula:

$$PRS = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \cdots + \beta_m x_m$$

where x_k is the allele dosage for SNP k, β_k is the corresponding weight for SNP k, and m is the total number of SNPs included in the PRS.

The approaches for subset selection of SNPs and the corresponding weights, β_k to be included in PRS construction can be broadly classified into single-ancestry and multi-ancestry methods. The single-ancestry methods we used were: (a) a clumping and thresholding (C+T) approach, in which SNPs significantly associated with breast cancer risk at a given p-value threshold

were retained; (b) lasso penalised regression, where weights of SNPs with minor contribution to the lasso model were shrunk toward zero[28]. The multi-ancestry methods included: (c) using a linear combination of European PRS with PRSs developed in (a) or (b), similar to the method described by Marquez-luna and group[29] (d) re-weighting of European PRS, where SNPs that have been previously selected for the European PRS were re-weighted using a combination of European and Asian-specific weights and (e) a Bayesian polygenic prediction approach (PRS-CSx[30]), where weights of SNPs were shrunk according to the strength of association in European GWAS and Asian GWAS through a SNP-specific continuous shrinkage prior. Figure 1 summarises the methods and dataset used in this study and Figure S1 gives the detailed schema of analyses. The lists of SNPs and the weights for the PRS computation are given in Table S2 - 4

To enable direct comparison of the performance of different PRSs, we standardised the PRSs to have unit standard deviation (SD) in the control subjects. Logistic regression models were used to estimate odds ratios (ORs) for association between the standardised PRSs and breast cancer risk in the validation set. For all models we adjusted for the first 10 principal components (PC) and study; The two studies in the validation set were genotyped in two batches and hence treated as different strata for the purposes of adjustment. We used a Cox proportional hazard model to evaluate the association between the PRS and breast cancer risk in the test set and to estimate hazard ratios (HRs) per SD. Models were adjusted for the first 2 PCs for SCHS and KCPS-II and the first 12 PCs for CKB. The discrimination of the PRSs were also compared using the receiver operating characteristic curve (AUC).. The HR per SD and AUC were obtained individually for each prospective cohorts and combined using fixed-effect meta-analysis. Test of heterogeneity between studies were obtained using *rma()* command in the *metafor* package in R.

Clumping and thresholding (C+T) method for SNP selection

For the PRS built using C+T or lasso penalized regression, we used training dataset 1 where summary statistics of 20,768 SNPs significantly associated with breast cancer risk at p-values < 0.001 were available. SNPs clumping (within 1Mb windows) was conducted using the software PRSice v2.11[31] to remove highly correlated SNPs (pairwise correlation $r^2 > 0.9$); the SNP with the lowest p-value for association in the correlated pairs was retained, resulting in 3,050 SNPs for subsequent analyses in the C+T method, and lasso penalized regression.

In the C+T approach, SNPs were further clumped within pre-specified clumping window sizes and threshold of a correlation r^2 to obtain nearly uncorrelated SNPs. PRSs were then computed for samples in the validation dataset using the subset of SNPs that were significant at a pre-specified p-value threshold. This threshold was set at $5x10^{-8}$ (genome-wide significance) and then increased in steps of 10^{-10} up to 10^{-3}. The discriminatory accuracy of each PRS for predicting breast cancer risk was evaluated in the validation dataset using the AUC, and the PRS with the highest AUC was selected as the best PRS. The whole process was then repeated for different combinations of clumping size and threshold of correlation r^2. The clumping and derivation of PRSs were done using PRSice v2.11[31], while the AUCs for PRSs were generated using the pROC package in *R*.

To account for the joint effect of SNPs used to derive the best PRS using the C+T method, the SNP weights should ideally be estimated jointly in a single logistic regression model. Since raw genotype data were not available for the training dataset ,we computed the optimal weight, from the summary statistics, for SNP *j* using the following formula:

$$\gamma_j = {\gamma'}_j \Big/ \sqrt{2p_j(1-p_j)} \qquad (1)$$

where $\underline{\gamma'} = R^{-1}\underline{\beta'}$, R is the matrix of correlations between the SNP genotypes, β' is the predicted normalised marginal effect sizes of the SNPs, and p_j is the effect allele frequency of SNP j (see Supplementary Methods). The matrix of correlation was computed using 4,921 control samples from BCAC OncoArray studies where raw genotype data was available.

Lasso penalized regression

For penalized regression, we used the package *lassosum* in R[28] which used the summary statistics of single-SNP association tests based on the training set, genotype data of samples in the validation set for model fitting and genotype data from Asian control samples in BCAC OncoArray studies for calculating linkage disequilibrium among SNPs. All 3,050 SNPs with p-value < 0.001 were included in the lasso. The analyses were repeated for different values of the penalty (lambda) and shrinkage (s) parameters, and the PRS giving the highest correlation between PRS and disease status (default metric in the method)in the validation dataset was selected.

Linear combination of European PRS with Asian PRS

We considered a PRS based on the 313 SNPs developed in European women[3]. Of the 313 SNPs, only 287 SNPs with imputation info score > 0.9 in OncoArray studies were retained for subsequent analyses. The weights reported in *Mavaddat et al (2019)[3]* were used to derive the European PRS in the validation dataset. Asian PRSs generated from C+T method or lasso penalized regression, were linearly combined with European PRS using logistic regression model which takes the form $\alpha_0 + \alpha_1 PRS_1 + \alpha_2 PRS_2$. The constant α_0, weights α_1 and α_2

were estimated using validation dataset and here PRS_1 and PRS_2 represent two different PRSs.

Re-weighting of European-based PRS

For these analyses, we considered two sets of weights for the 287 SNPs: (i) Asian weights estimated from the training dataset 1 alone; and (ii) weights based on a combination of the Asian and European weights, allowing for these weights to differ but be correlated. Other approaches to combine European and Asian-specific weights were also explored, including combining weights using fixed effect meta-analysis, but only the method that gave the best AUC is presented here.

For (i), the optimal weights taking into account the correlation between SNPs were derived using Equation (1). For (ii), we combined Asian and European weights using an Empirical Bayes approach where the optimal weight is given by the following formula

$$\beta_{j,EB} = {\beta_{jA,EB}}\Big/{\sqrt{2p_j(1-p_j)}}.$$

Here, $\beta_{jA,EB}$ is the estimated posterior effect sizes in Asians given the data, obtained through an Expectation-Minimisation algorithm and p_j is the allele frequency for SNP j (see Supplementary Methods). This approach "shrinks" the Asian estimates towards the European estimates, making use of the greater precision in the European estimates but allowing for different Asian weights when the European and Asian estimates differ markedly.

We also considered linear combination of the re-weighted European PRSs with Asian PRSs generated from C+T method or lasso penalized regression in the validation dataset.

Bayesian polygenic prediction approach (PRS-CSx)

For PRS-CSx, we used the summary statistics from Asian GWAS (training dataset 2) and the summary statistics from European GWAS (training dataset 3). Asians and Europeans in the 1000 Genomes Phase 3 project were selected as LD reference panels. Two genome-wide PRSs were generated – one using the European-specific posterior SNP effect size estimates and one using the Asian-specific posterior effect size estimates generated from PRS-CSx. The two PRSs were then linearly combined in the validation dataset. The analyses were repeated across a range of global shrinkage parameter ($\phi = 10^{-6}, 10^{-4}, 10^{-2}, 1$, as suggested by Ruan and collegues[30] ; the ϕ that gave the linear combination of PRSs with the highest AUC in the validation dataset was selected as optimal global shrinkage parameter. Analyses were run using the published Python code based tool published in Github[30].

PRSs for South Asian population

The predictive performance of PRSs developed for women of East Asian ancestry in women of Indian ancestry were assessed using AUC and OR per SD. Given the much smaller sample size for Indian-ancestry women, we did not attempt to generate a South Asian-specific PRS. For linear combinations of multiple PRSs of the form $\alpha_0 + \alpha_1 PRS_1 + \alpha_2 PRS_2$, we considered the predictive performance of these combined PRSs with weights α_0, α_1, and α_2 estimated from East Asian validation dataset as well as South Asian validation dataset. **Absolute risk of breast cancer by PRS percentiles**

The age-specific absolute risks of developing breast cancer in each PRS percentile were obtained by constraining to the incidence of overall population breast cancer incidence (see Supplementary Methods). The details of these methods have been described previously[3]. We calculated lifetime and 10-year absolute risks using Singaporean mortality and breast cancer incidence in 2017[32,33]. For birth-cohort specific incidences, age-specific breast cancer

incidences for the 1960-1969 and 1970-1979 birth cohorts were calculated using data on breast cancer incidence in Singapore from 1968 to 2017[32]. For women born after 1979, incidences could only be calculated up to age 35. Hence, for the >1979 birth cohort, breast cancer incidences were projected by assuming an annual increase in breast cancer incidence of 3.9%[34].

Results

Genetic diversity within Asian populations in our study

We first assessed the overall diversity among the study populations. Figure 1 and Figure S1 summarise the dataset used in this study. The plot of the first two principal components were generated using the validation dataset (annotated by the ethnicities) together with the subset of samples from six countries within the BCAC studies in the training dataset where raw genotype data were available (Figure 2(a)). As expected, the populations are clustered, consistent with geography and population history. The genetic distance is largest between Indian ancestry women and Japanese/Korean women The women of Chinese ancestry in Malaysia, Singapore, mainland China, Hong Kong and Taiwan form a distinct cluster that is genetically closer to Japanese/Koreans women than to women of Indian ancestry. The women from Thailand and women of Malay ancestry from Malaysia and Singapore are genetically closer to women of Chinese ancestry than to women of Indian ancestry. Given the large genetic distance between Indian ancestry women from the other populations, in subsequent analyses, Chinese ancestry and Malay ancestry women were treated as “East Asian” in the validation dataset, and evaluated separately from Indian ancestry women.

PRSs developed using Asian-specific SNPs

We first developed PRSs using Asian-specific SNPs selected using two different methods. For the C+T method, SNPs were removed if they were within 250kb of a SNP already selected and correlated at $r^2 > 0.1$, leaving 1,326 SNPs for analysis. For women of East Asian ancestry in the validation dataset, the best PRS was obtained at a p-value threshold of $p < 5.74 \times 10^{-7}$, resulting in a 46-SNP PRS (PRS_{46}) (Figure S2). The breast cancer odds ratio (OR) per unit standard deviation (SD) for PRS_{46} was 1.35 (95% CI: 1.30 –1.39) and the AUC was 0.586 (Table 1). We searched for better PRSs by using different combinations of clumping size and correlation threshold r^2; however, there was no appreciable difference in terms of AUCs of the resulting PRSs computed (Figure S3).

For lasso penalised regression, the best PRS for women of East Asian ancestry in the validation dataset was obtained at penalty parameter (λ)= 0.014 and shrinkage parameter (s) = 0.9, resulting in a PRS that included 2,985 (out of 3,050) SNPs (Figure S4). The corresponding OR per SD for this 2,985-SNP PRS (PRS_{2985}) was 1.41 (95% CI: 1.36–1.46; AUC = 0.596), slightly more predictive than the best PRS (PRS_{46}) developed by the C+T method (Table 1).

Linear combinations of European PRS and Asian PRSs

We denoted the 287-SNP European PRS computed using the reported weights in women of Europeans ancestry[3] as PRS_{287_EUR}. A linear combination of PRS_{287_EUR} and PRS_{46} significantly (p-values < 0.0001) improved the predictive accuracy in East Asian ancestry women, as compared to using the Asian-specific PRS alone (OR per SD (95% CI) = 1.54 (1.49-1.60) and AUC = 0.623 for PRS_{46}+PRS_{287_EUR} versus OR per SD (95% CI) = 1.37 (1.32-1.47) and AUC = 0.589 for PRS_{46}, Table 1). The improvement using the combined PRS was marginal compared to using PRS_{287_EUR} alone (OR per SD (95% CI) = 1.50 (1.45-1.56) and AUC = 0.615), but PRS_{46}

remained significant in the linear combination model (Table S5). Combination of PRS_{287_EUR} with PRS_{2985} further increased the OR per SD and AUC compared to PRS_{46}+PRS_{287_EUR}, though the differences were not statistically significant.

PRSs developed by integrating Asian weights into the European PRS

We considered two additional version of the 287-SNP PRS, using adjusted weights: (a) PRS_{287_ASN}, which used Asian-specific weights estimated from the training dataset 1 adjusted for linkage disequilibrium; and (b) PRS_{287_EB}, weights based on a combination of the European and Asian-specific weights, using an empirical Bayes procedure. For women of East Asian ancestry, PRS_{287_EB} (OR per SD= 1.53, 95% CI: 1.47–1.58; AUC = 0.620) was slightly more predictive than PRS_{287_ASIAN} (OR per SD = 1.50, 95% CI: 1.45–1.56; AUC = 0.615) and PRS_{287EUR} (OR per SD = 1.50, 95% CI: 1.45–1.56; AUC = 0.614), and significantly more predictive than PRS_{46} and PRS_{2985} (p-value < 0.0001) (Table 1). A linear combination of PRS_{287_EB} with PRS_{46} further improved the PRS performance compared to PRS_{46}+ PRS_{287_EUR}, though the improvement was not significant.

Genome-wide PRSs

We then developed genome-wide PRSs using the Bayesian polygenic prediction approach (PRS-CSx). We denoted the genome-wide PRSs generated using European and Asian posterior weights as PRS_{GW_EUR} and PRS_{GW_ASN}, respectively. The best combined PRS for East Asian ancestry women was obtained at $\phi = 10^{-4}$ (Table S6). The corresponding OR per SD (95% CI) and AUC for PRS_{GW_EUR} + PRS_{GW_ASN} were 1.62 (1.52-1.68) and 0.636, respectively, significantly better than all the PRSs described thus far (Table 1). This significant improvement was mainly driven by the contribution of PRS_{GW_EUR} (OR per SD (95% CI) = 1.59 (1.53-1.65) and AUC =

0.629). The OR per SD (95% CI) and AUC for PRS_{GW_ASN} alone was only 1.44 (1.39-1.49) and 0.601, respectively, only slightly better than PRS_{46} (Table S6).

PRSs for Indian-ancestry population

We evaluated the PRSs developed for women of East Asian ancestry in predicting risk in women of South Asian ancestry. These PRSs were all predictive of risk in South Asian ancestry women in the validation dataset but the effect sizes were reduced compared to East Asian ancestry women (Table 2). As for East Asians, the effect sizes were smallest for Asian-based PRSs alone, larger for combined European and Asian PRSs and the genome-wide PRS (PRS-CSx) was most predictive of breast cancer risk in women of South Asian ancestry. We estimated the weights of the considered linear combinations of PRSs using the South Asian validation dataset (Table S5), there were no appreciable difference in the effect sizes and AUCs compared to estimated weights using East Asian validation dataset (Table S7.

Evaluation of PRSs in prospective cohorts

The combined hazard ratios (HR) per unit SD of Asian PRSs in the prospective cohorts combined were 1.40 (95% CI: 1.25 - 1.56, AUC = 0.600) and 1.45 (95% CI: 1.31-1.61, AUC = 0.608), respectively for PRS_{46} and PRS_{2985}, similar to those observed in the validation set (Table 1 and Figure S5). There was no evidence of heterogeneity in the HRs among studies. Thus, the effect sizes were smallest for PRS based on Asian data alone (PRS_{46} and PRS_{2985}), larger for PRS based on the European PRS (HR per SD = 1.50 (95% CI: 1.35-1.65) for PRS_{287_EB}) and larger still for PRS based on combining the Asian and European PRS (HR per SD = 1.53 (95% CI: 1.37-1.71) for PRS_{46}+PRS_{287_EB}). As in the validation dataset, PRS generating using PRS-CSx showed

the strongest association with breast cancer risk (HR per SD (95% CI) = 1.62 (1.46- 1.80)) and highest AUC (0.635).

Absolute breast cancer risk predictions

To assess the potential of translating PRS into clinical tool for Asian population, we transformed the relative risks across the PRS continuum into absolute risk, using PRS_{46}+PRS_{287_EB} and calendar-specific breast cancer incidence and mortality rates[32] for Chinese women in Singapore as an example. Based on the validation dataset comprised of East Asian ancestry women, the breast cancer odd ratios (ORs) by percentiles of PRS_{46}+PRS_{287_EB} are shown in Table S8. Compared to women in the middle quintile (40-60th), the estimated ORs for developing breast cancer for women in the highest and lowest 1% of the PRS distribution were 0.53 (95% CI = 0.33 -0.82) and 3.01 (95% CI = 2.25-4.06), respectively. The estimated ORs by PRS percentile did not differ from those predicted under a theoretical polygenic model in which the log OR increases linearly with the PRS (Table S8). The corresponding lifetime risks of developing breast cancer by age 80 years, on current incidence rates, were ~2% and ~19% respectively (Figure 3(a)). Assuming that a 10-year absolute risk threshold of 2.3%[35], is used to define women at sufficient risk to justify screening, if we were to offer breast cancer screening to women age 40 years based on 10-year absolute risk as estimated from their polygenic risk profiling, approximately 12% of Chinese women would reach the risk threshold of screening before or at age 40 (Figure 3(b)). However, these are likely underestimates, as young Asian women today are experiencing a dramatic increase in breast cancer incidence[12]. Figure S6 shows the distribution of the 10-year absolute risk at age 40 for women who were born after 1979 using projected breast cancer incidence rates (see Methods). Compared to women who were born in 1960-1969, where

only 11% of Chinese women would reach the risk threshold at age 40, we projected that the proportion would rise to 29% in women age 40 born after 1979.

Generalisability of PRS across Asian ethnic subgroups

Subjects in the validation dataset were recruited from a multi-ethnic population in Southeast Asia comprised of women from three ethnic groups in Asia – Chinese, Malay and Indian ancestry women giving us the opportunity to evaluate the generalisability of PRS across Asian ancestry populations, using PRS_{46}+PRS_{287_EB} developed for women of East Asian ancestry as an example. This combined PRS was predictive of risk in all ethnic groups (Table S9). The effect size and AUC were higher in Chinese ancestry women compared to Malay and Indian ancestry women in Southeast Asia, but the differences were not statistically significant (OR per SD = 1.56; 95% CI: 1.50-1.63; AUC = 0.625 for Chinese versus OR per SD = 1.51; 95% CI: 1.39-1.64; AUC = 0.614 for Malays and OR per SD = 1.49; 95% CI: 1.33-1.66; AUC = 0.610 for Indians, heterogeneity p-value = 0.983 (Figure S7 and Table S9).

The distribution of PRS_{46}+PRS_{287_EB} was, however, different among the three ethnic groups. While there was only a marginal difference in the SD (SD in Chinese, Malay, and Indian controls were 0.439, 0.556, and 0.455, respectively), the means differed markedly, being highest in Chinese and lowest in Indians (mean in Chinese, Malay, and Indian controls were -0.118, -0.197, and -0.328, respectively, p-values of pair-wise comparison of means < 0.0001) (Table S8). To demonstrate the importance of ethnic-specific calibration of PRS, we evaluated the impact of classifying Malay and Indian women according to PRS risk groups as determined by the distribution in Chinese women. Figure 3(c) shows the PRS distribution in control subjects by percentiles across the three ethnic groups and their corresponding cumulative breast cancer risk by age 80, generated using calendar-specific breast cancer incidence and

mortality rates for Chinese, Malay and Indian women in Singapore[36]. For example, the 95th percentile in Indians corresponds, approximately, to the 90th percentile in the Chinese population. The lifetime breast cancer risk of Indian women in the 95th percentile was 11%. If Chinese PRS distribution was used as a reference, these Indian women would be categorised as 90th percentile and hence would be told that their corresponding lifetime risk was 9% instead of 11%. Hence, using the Chinese PRS distribution as a reference would underestimate the risk in Indian women. The differences in PRS distribution was even more apparent when women of European ancestry was used as reference (Figure 3(d)).

The patterns of genetic clusters shown in Figure 2(a) are mirrored in the distribution of the standardised PRS_{46}+PRS_{287_EB} by population (Figure 2b) in the validation and BCAC dataset. The largest differences in the means of the standardised PRS_{46}+PRS_{287_EB} were observed between the Indian-ancestry women and Japanese/Korean women (with Indians being the biggest outlier). The mean PRS_{46}+PRS_{287_EB} of women from Thailand was similar to that of Malay ancestry women, somewhat lower than women of Chinese ancestry.

Discussion

Personalised risk stratification for prevention and early detection of breast cancer has gained increasing interest; however, it is important to recognize the need to study women representing diverse ancestries, to lessen health disparities. Our study provides essential information about the use of PRSs for breast cancer risk prediction in women of Asian ancestry. We developed and validated different PRSs for East Asian ancestry women: the key observations were (a) PRSs generated by integrating information from European ancestry and Asian ancestry GWAS datasets performed better than PRSs based purely on weights derived

from single-ancestry GWAS data, and (b) there were substantial differences in PRS distributions across ethnic groups.

Based on the largest available breast cancer GWAS datasets, the best PRS for women of East Asian ancestry was based on PRS-CSx Bayesian polygenic prediction method [ref] (PRS_{GW_ASN}+PRS_{GW_EUR}). This PRS included ~1 million SNPs and had notably larger effect size than the European PRS (PRS_{287_EUR}) that had previously shown to be the best breast cancer PRS for women of Asian ancestry [ref] (HR per unit SD in prospective cohorts: 1.62 versus 1.46; AUC: 0.635 versus 0.609, Table 1). It is noteworthy that the predictive performance of this PRS was similar to that achieved in European population (HR per SD (95% CI) of 313-SNP PRS: 1.59 (1.54-1.64) as reported in *Mavaddat et al*). However, despite the rapid drop in cost associated with next-generation sequencing, implementation of genome-wide PRS can be practically more challenging compared to the implementation of the European PRS that included only 313 variants.

We showed that adaptions based on this European PRS can improve risk prediction in women of East Asian ancestry . The optimal approach used an empirical Bayes procedure, shrinking the weights from the East Asian data towards those in Europeans. Linearly combining this reweighted European PRS with PRS generated from the C+T method resulting in a PRS (PRS_{46}+PRS_{287_EB}) that had predictive performance that was intermediate between the original European PRS alone and the PRS-CSx approach (HR per unit SD in prospective cohorts: 1.53; AUC: 0.621, Table 1). The 313-SNP PRS is being used in several clinical studies in European populations, including the MyPeBs[8] and WISDOM[7] trials, and the PRS_{46}+PRS_{287_EB} PRS would be relatively easy to implement in clinical settings. This approach is similar to that

described in *Márquez-Luna et al (2017)*, where the authors proposed to linearly combine PRSs generated from two training populations to reduce the gap in prediction accuracy between European and non-European populations[29]. Our results suggest that a similar approach may be effective for adapting PRS developed using data from European ancestry populations to other non-European populations where sample sizes are limited.

The PRS generated for women of East Asian ancestry were also predictive for women South Asian ancestry, but the effect sizes were smaller. When combining East Asian-derived genome-wide PRS with European-derived genome-wide PRS in women of South Asian ancestry using the PRS-CSx approach, it was noticeable that the East Asian component made a smaller contribution to the linear combination (Table S5). These results demonstrate the need for larger studies of women of South Asian ancestry both to optimize the PRS and validate in prospective cohorts.

One of the challenges of moving PRS into clinical implementation is transferability across different ethnic groups. Several studies have evaluated the population-level applicability of European PRSs to non-European populations for various diseases[37-40]. Similar to these studies, we showed that the mean of the PRS distribution differ substantially between European and Asian ethnic subgroups. We showed that if the European PRS (PRS-287_{EUR}) was applied to an Asian population without adjustment, the 60th percentile in Chinese ancestry and Malay ancestry women and 80th percentile in Indian ancestry women corresponds, approximately, to the 90th percentile in the European population, resulting in overestimation of risk in these women (Figure 3(d)). To our knowledge, no studies thus far have looked at the transferability of breast cancer PRS within diverse Asian ethnic subgroups. Our results showed that while the effect sizes appeared to be similar across ethnic groups (Table S8), the mean PRS

distribution differed substantially across Asian populations (Table S7 and Figure 2(b)). For example, even though Japanese, Koreans and Han Chinese are conventionally classified as East Asians in genetic analyses, the means PRS were markedly different between these ethnic groups (Figure 2(b)). The differences are sufficiently large to affect risk classification, so comparing the PRS for an individual woman with the correctly calibrated ethnic-specific distribution is crucial for valid risk prediction. This however can be problematic for admixed individuals, where the genomes composed from multiple ancestries that may be closely or distantly related to the reference population. As more samples of Asian ancestry become available, it may be possible to combine ethnic-specific PRSs with ancestry components to derive better multi-ethnic PRSs[29].

Our work is subject to several limitations that open up future research directions. Firstly, although we have demonstrated that the predictive performance of European PRS can be improved by integrating weights from Asians using an empirical Bayes approach, the absolute increase in predictive accuracies is marginal. This marginal improvement is likely reflective of the similarity of underlying effect sizes between Asians and Europeans for the 287 breast cancer variants. The method may be potentially valuable for adapting European PRS to populations that are not genetically closely related to Europeans. Secondly, our studies focus on developing PRS without using individual-level training data, when such data is available it may be possible to develop PRS with higher accuracy using methods that fit all variants simultaneously, such as the step-wise hard-thresholding method as described in *Mavaddat et al* (2019)[3], or considering subtype-specific disease analyses to retain more informative variants. Morover, our results showed that PRSs developed using Asian-derived GWAS dataset had significantly poorer performance compared to the European PRS indicating that such improvement likely requires much larger discovery cohort of Asian ancestry. Finally, PRSs

were linearly combined using the validation dataset and hence the reported performance is likely subject to overfitting. Although we have shown that performance of the combined PRSs in East Asians were replicated in the prospective cohorts, we did not have a similar independent dataset for South Asian for such replication.

In summary, we have shown that genome-wide PRS derived from trans-ancestry method had significantly higher predictive accuracy for women of Asian ancestry than existing breast cancer PRSs. We also showed that European-based PRS can be improved for use in Asian populations by integrating population-specific weights and combined with Asian-specific PRS. Importantly, the differences in distribution of the same PRS across different ethnic groups (among Asians, or between Asian and Europeans) emphasise the need for ethnic-specific calibration before translating PRS into practice for diverse Asian populations.

Data availability

Summary statistics (odds ratios and confidence limits) for all SNPs used in the analysis are provided in Supplementary Data 1 of the manuscript. Request for access to individual level data on which these analyses were based can be made via the Data Access Coordinating Committee of BCAC (BCAC Coordinator: BCAC@medschl.cam.ac.uk).

References

1. Shiovitz, S. & Korde, L. A. Genetics of breast cancer: a topic in evolution. *Ann Oncol* **26**, 1291-1299, doi:10.1093/annonc/mdv022 (2015).
2. Dorling, L. *et al*. Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. *N Engl J Med*. **4**;384(5):428-439(2021).
3. Mavaddat, N. *et al*. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet* **104**, 21-34, doi:10.1016/j.ajhg.2018.11.002 (2019).
4. Michailidou, K. *et al*. Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92-94, doi:10.1038/nature24284 (2017).
5. Milne, R. L. *et al*. Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. *Nat Genet* **49**, 1767-1778, doi:10.1038/ng.3785 (2017).

6. Pashayan, N. *et al*. Publisher Correction: Personalized early detection and prevention of breast cancer: ENVISION consensus statement. *Nat Rev Clin Oncol* **17**, 716, doi:10.1038/s41571-020-0412-0 (2020).
7. Esserman, L. J., Study, W. & Athena, I. The WISDOM Study: breaking the deadlock in the breast cancer screening debate. *NPJ Breast Cancer* **3**, 34, doi:10.1038/s41523-017-0035-5 (2017).
8. MyPeBS Personalizing Breast Screening, <*https://mypebs.eu/the-project/*>
9. Personalized risk assessment for prevention and early detection of breast cancer: Integration and Implementation, <*https://www.genomecanada.ca/en/personalized-risk-assessment-prevention-and-early-detection-breast-cancer-integration-and*>
10. Martin, A. R. *et al*. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet* **51**, 584-591, doi:10.1038/s41588-019-0379-x (2019).
11. Youlden, D. R., Cramb, S. M., Yip, C. H. & Baade, P. D. Incidence and mortality of female breast cancer in the Asia-Pacific region. *Cancer Biol Med* **11**, 101-115, doi:10.7497/j.issn.2095-3941.2014.02.005 (2014).
12. Bhoo-Pathy, N. et al. Breast cancer research in Asia: adopt or adapt Western knowledge? *Eur J Cancer* **49**, 703-709, doi:10.1016/j.ejca.2012.09.014 (2013).
13. Heer, E. et al. Global burden and trends in premenopausal and postmenopausal breast cancer: a population-based study. *Lancet Glob Health* **8**, e1027-e1037, doi:10.1016/S2214-109X(20)30215-1 (2020).
14. Ho, W. K. et al. European polygenic risk score for prediction of breast cancer shows similar performance in Asian women. *Nat Commun* **11**, 3833, doi:10.1038/s41467-020-17680-w (2020).
15. Wen, W. *et al.* Prediction of breast cancer risk based on common genetic variants in women of East Asian ancestry. *Breast Cancer Res* **18**, 124, doi:10.1186/s13058-016-0786-1 (2016).
16. Hankin, J. H. *et al.* Singapore Chinese Health Study: development, validation, and calibration of the quantitative food frequency questionnaire*. Nutr Cancer* **39**, 187-195, doi:10.1207/S15327914nc392_5 (2001).
17. Chen, Z. *et al.* China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. Int J Epidemiol **40**, 1652-1666, doi:10.1093/ije/dyr120 (2011).
18. Jee, Y. H. *et al*. Cohort Profile: The Korean Cancer Prevention Study-II (KCPS-II) Biobank. *Int J Epidemiol* 47, 385-386f, doi:10.1093/ije/dyx226 (2018).
19. Michailidou, K. *et al*. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* **45**, 353-361, 361e351-352, doi:10.1038/ng.2563 (2013).
20. Amos, C. I. *et al*. The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. *Cancer Epidemiol Biomarkers Prev* **26**, 126-135, doi:10.1158/1055-9965.EPI-16-0106 (2017).
21. Ishigaki, K. *et al*. Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. Nat Genet; 52(7):669-679 (2020).
22. Shu, X. *et al.* Identification of novel breast cancer susceptibility loci in meta-analyses conducted among Asian and European descendants. *Nat Commun* **11**, 1217, doi:10.1038/s41467-020-15046-w (2020).
23. Dorajoo, R. *et al.* Loci for human leukocyte telomere length in the Singaporean Chinese population and trans-ethnic genetic studies. *Nat Commun* **10**, 2491, doi:10.1038/s41467-019-10443-2 (2019).
24. Gan, W. *et al*. Evaluation of type 2 diabetes genetic risk variants in Chinese adults: findings from 93,000 individuals from the China Kadoorie Biobank. *Diabetologia* **59**, 1446-1457, doi:10.1007/s00125-016-3920-9 (2016).
25. Sakaue, S, & Kanai, M., *et al*. A global atlas of genetic assocaition of 220 deep phenotypes. *medRxiv* (2020).

26. Genomes Project, C. *et al*. A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).
27. International HapMap, C. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-861, doi:10.1038/nature06258 (2007).
28. Mak, T. S. H., Porsch, R. M., Choi, S. W., Zhou, X. & Sham, P. C. Polygenic scores via penalized regression on summary statistics. *Genet Epidemiol* **41**, 469-480, doi:10.1002/gepi.22050 (2017).
29. Márquez-Luna, C., et al. Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genet Epidemiol*;**41**(8):811-823 (2017).
30. Ruan, Y., et al., Improving Polygenic Prediction in Ancestrally Diverse Populations. *medRxiv* (2020).
31. Choi, S. W. & O'Reilly, P. F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience* 8, doi:10.1093/gigascience/giz082 (2019).
32. *Age-Specific Death Rates, Annual* (Government of Singapore, 2017).
33. Cancer Incidence in Five Continents. (International Agency for Research on Cancer, Lyon, (2014).
34. Jara-Lazaro, A.R., S. Thilagaratnam, and P.H. Tan, *Breast cancer in Singapore: some perspectives.* Breast Cancer, 2010. **17**(1): p. 23-8.
35. DeSantis, C., Ma, J., Bryan, L. & Jemal, A. Breast cancer statistics, 2013. *Cancer J. Clin*. **64**, 52–62 (2014).
36. SINGAPORE CANCER REGISTRY 50TH ANNIVERSARY MONOGRAPH 1968 – 2017.
37. Martin, A.R., *et al*. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet*;**51(4)**:584-591 (2019).
38. Martin, A.R., *et al*. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am J Hum Genet*; **6**;100(4):635-649 (2017).
39. Reisberg, S., Iljasenko, T., Lall, K., Fischer, K., Vilo, J. Comparing distributions of polygenic risk scores of type 2 diabetes and coronary heart disease within different populations. *PLoS One* ;**12(7)**:e0179238 (2017).
40. Isgut, M., Sun, J., Quyyumi, A.A., Gibson, G. Highly elevated polygenic risk scores are better predictors of myocardial infarction risk early in life than later. *Genome Med*;**13**(1):13 (2021).

Funding and acknowledgements

Funding

This study was supported by grants from Newton-Ungku Omar Fund [grant no: MR/P012930/1] and Wellcome Trust [grant no: v203477/Z/16/Z]. The Malaysian Breast Cancer Genetic Study was established using funds from the Malaysian Ministry of Science, and the Malaysian Ministry of Higher Education High Impact Research Grant [grant no:

UM.C/HIR/MOHE/06]. The Malaysian Mammographic Density Study was established using funds raised through the Sime Darby LPGA tournament and the High Impact Research Grant. Additional funding was received from Yayasan Sime Darby, PETRONAS, Estee Lauder Group of Companies and other donors of Cancer Research Malaysia.

WKH is the recipient of L'Oreal-UNESCO For Women in Science National Fellowship. JL is the recipient of a National Research Foundation Singapore Fellowship (NRF-NRFF2017-02). ACA is supported through Cancer Research - UK (C12292/A20861). JS holds a Canada Research Chair in Oncogenetics.

The PERSPECTIVE I&I project is funded by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the Ministère de l'Économie et de l'Innovation du Québec through Genome Québec, the Quebec Breast Cancer Foundation, the CHU de Quebec Foundation and the Ontario Research Fund.

BCAC is funded by Cancer Research UK [C1287/A16563, C1287/A10118], the European Union's Horizon 2020 Research and Innovation Programme (grant numbers 634935 and 633784 for BRIDGES and B-CAST respectively), and by the European Community´s Seventh Framework Programme under grant agreement number 223175 (grant number HEALTH-F2-2009-223175) (COGS). The EU Horizon 2020 Research and Innovation Programme funding source had no role in study design, data collection, data analysis, data interpretation or writing of the report.

Genotyping of the OncoArray was funded by the NIH Grant U19 CA148065, and Cancer UK Grant C1287/A16563 and the PERSPECTIVE project supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research (grant GPH-129344) and, the Ministère de l'Économie, Science et Innovation du Québec through Genome Québec and the PSRSIIRI-701 grant, and the Quebec Breast Cancer Foundation. Funding for the iCOGS infrastructure came from: the European Community's Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, and Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund. The DRIVE Consortium was funded by U19 CA148065.

MYBRCA is funded by research grants from the Malaysian Ministry of Higher Education (UM.C/HlR/MOHE/06) and Cancer Research Malaysia. MYMAMMO is supported by research grants from Yayasan Sime Darby LPGA Tournament and Malaysian Ministry of Higher Education (RP046B-15HTM). SGBCC is funded by NUS Start Up Grant, National University Cancer Institute Singapore (NCIS) Centre Grant, NMRC Clinical Scientist Award, NMRC Clinician Scientist Award-Senior Investigator, Asian Breast Cancer Research Fund and Breast

Cancer Prevention Programme under Saw Swee Hock School of Public Health. Recruitment of controls were funded by the Biomedical Research Council, (05/1/21/19/425).

The ACP study is funded by the Breast Cancer Research Trust, UK. KM and AL are supported by the NIHR Manchester Biomedical Research Centre, the Allan Turing Institute, and, by the ICEP (Cancer Research UK (C18281/A19169). CBCS is funded by the Canadian Cancer Society (grant # 313404) and the Canadian Institutes of Health Research. The HERPACC was supported by MEXT Kakenhi (No. 170150181 and 26253041) from the Ministry of Education, Science, Sports, Culture and Technology of Japan, by a Grant-in-Aid for the Third Term Comprehensive 10-Year Strategy for Cancer Control from Ministry Health, Labour and Welfare of Japan, by Health and Labour Sciences Research Grants for Research on Applying Health Technology from Ministry Health, Labour and Welfare of Japan, by National Cancer Center Research and Development Fund, and "Practical Research for Innovative Cancer Control (15ck0106177h0001)" from Japan Agency for Medical Research and development, AMED, and Cancer Bio Bank Aichi. The KOHBRA study was partially supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), and the National R&D Program for Cancer Control, Ministry of Health & Welfare, Republic of Korea (HI16C1127; 1020350; 1420190). LAABC is supported by grants (1RB-0287, 3PB-0102, 5PB-0018, 10PB-0098) from the California Breast Cancer Research Program. Incident breast cancer cases were collected by the USC Cancer Surveillance Program (CSP) which is supported under subcontract by the California Department of Health. The CSP is also part of the National Cancer Institute's Division of Cancer Prevention and Control Surveillance, Epidemiology, and End Results Program, under contract number N01CN25403. The Northern

California Breast Cancer Family Registry (NC-BCFR) were supported by grant U01CA164920 from the USA National Cancer Institute of the National Institutes of Health. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR. The NGOBCS was supported by the National Cancer Center Research and Development Fund (Japan). The SBCGS was supported primarily by NIH grants R01CA64277, R01CA148667, UMCA182910, and R37CA70867. Biological sample preparation was conducted the Survey and Biospecimen Shared Resource, which is supported by P30 CA68485. The scientific development and funding of this project were, in part, supported by the Genetic Associations and Mechanisms in Oncology (GAME-ON) Network U19 CA148065. SEBCS was supported by the BRL (Basic Research Laboratory) program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (2012-0000347). The TWBCS is supported by the Taiwan Biobank project of the Institute of Biomedical Sciences, Academia Sinica, Taiwan. the BioBank Japan Project (BBJ1) supported by the Ministry of Education, Culture, Sports, Sciences and Technology from the Japanese Government);the Hwasun Cancer Epidemiology Study-Breast (HCES-Br) supported by the Biobank of Chonnam National University Hwasun Hospital, a member of the Korea Biobank Network.

The SCHS study was supported by grants from the National Medical Research Council, Singapore (NMRC/CIRG/1456/2016) and the National Institutes of Health, USA (R01 CA144034 and UM1 CA182876). China Kadoorie Biobank was supported as follows: Baseline survey and first re-survey: Hong Kong Kadoorie Charitable Foundation; long-term follow-up: UK Wellcome Trust (212946/Z/18/Z, 202922/Z/16/Z, 104085/Z/14/Z, 088158/Z/09/Z),

National Natural Science Foundation of China (91843302), and National Key Research and Development Program of China (2016YFC 0900500, 0900501, 0900504, 1303904). DNA extraction and genotyping: GlaxoSmithKline, UK Medical Research Council (MC_PC_13049, MC-PC-14135). The UK Medical Research Council (MC_UU_00017/1, MC_UU_12026/2 MC_U137686851), Cancer Research UK (C16077/A29186; C500/A16896) and the British Heart Foundation (CH/1996001/9454), provide core funding to the Clinical Trial Service Unit and Epidemiological Studies Unit at Oxford University for the project.

Acknowledgements

We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. The BCAC study would not have been possible without the contributions of the staff of the McGill University and Génome Québec Innovation Centre, Stig E. Bojesen, Sune F. Nielsen, Borge G. Nordestgaard, and the staff of the Copenhagen DNA laboratory, and Julie M. Cunningham, Sharon A. Windebank, Christopher A. Hilker, Jeffrey Meyer and the staff of Mayo Clinic Genotyping Core Facility.

MYBRCA thanks study participants and all research staff at Cancer Research Malaysia, University Malaya, and Sime Darby Medical Centre who assisted in recruitment and interviews (particularly Siti Norhidayu Hassan, Patsy Pei-Sze Ng, Sook-Yee Yoon, Shivani Mariapun, Joanna Lim, Daphne Lee, Norhashimah Hassan and Sze Yee Puah) for their contributions and commitment to this study.

For SGBCC, we want to thank the program manager Jenny Liu, clinical research coordinators/research assistants Siew-Li Tan, Siok-Hoon Yeo, Kimberley Chua, Ting-Ting Koh, Amanda Ong, Jin-Yee Lee, Michelle Mok, Ying-Jia Chew, Jing-Jing Hong and Hui-Min Lau for

their contributions in recruitment, Yen-Shing Yeoh for data preparation and Alexis Khang for processing the DNA samples. We also want to thank all the participants' support to SGBCC.

The ACP study wishes to thank the participants in the Thai Breast Cancer study. Special Thanks also go to the Thai Ministry of Public Health (MOPH), doctors and nurses who helped with the data collection process. Finally, the study would like to thank Dr Prat Boonyawongviroj, the former Permanent Secretary of MOPH and Dr Pornthep Siriwanarungsan, the former Department Director-General of Disease Control who have supported the study throughout. CBCS thanks study participants, co-investigators, collaborators and staff of the Canadian Breast Cancer Study, and project coordinators Agnes Lai and Celine Morissette. HKBCS thanks Hong Kong Sanatorium and Hospital, Dr Ellen Li Charitable Foundation, The Kerry Group Kuok Foundation, National Institute of Health 1R03CA130065 and the North California Cancer Center for support. We thank all investigators of the KOHBRA (Korean Hereditary Breast Cancer) Study. LAABC thanks all the study participants and the entire data collection team, especially Annie Fung and June Yashiki. SBCGS thank study participants and research staff for their contributions and commitment to the studies.

SCHS thank the Singapore Cancer Registry for identification of cancer cases within their cohort. CKB acknowledges the study participants and members of the survey teams in each of the 10 regional centres, and the project development and management teams based at Beijing, Oxford and the 10 regional centres. China Kadoorie Biobank acknowledges the contribution of participants, project staff, and China National Centre for Disease Control and Prevention (CDC) and its regional offices. China's National Health Insurance provided electronic linkage to all hospital treatments.

Authors contribution

Conceptualization: W.K.H., J.Simard, A.C.A., D.F.E., S.H.T.; Project administration: M.C.T., M.K.B, Q.W., E.A.W.; Data curation: J.D., N.M., K.Michailidou, J.Long; Formal analysis: W.K.H., M.C.T., X.S., J.Li, P.J.H., I.Y.M., K.L., Y.H.J., S.H.Lee, D.F.E; Resources: , T.H., K.R., V.K.M.T, B.K.T.T., S.M.T., E.Y.T., S.H.Lim, Y.T.G., Y.Z., D.K., J.Y.C., W.H., H.B.L., M.K., Y.O., N.M., B.B.J., S.K.P., S.W.K., C.Y.S., P.E.W., B.P., K.Muir, A.L., A.H.W., C.C.T., K.Matsuo, H.I., A.K., T.L.C., E.M.J., A.W.K., M.I., T.Y., S-S.K., K.J.A., R.A.M., W.P.K., C.C.K., J.M.Y, R.D., R.G.W., Z.C., L.L., J.Lv, K.J.J, P.K, P.D.P.P., A.M.D., X.O.S., C.H.Y., N.A.M.T., W.Z., M.H., S.H.T. Supervision: D.F.E., S.H.T.; Writing – original draft: W.K.H., M.C.T., A.C.A., D.F.E., S.H.T.; Writing – review & editing: all authors.

Conflict of Interest

The authors confirm that they have no conflict of interests.

Figure legends

Figure 1. Overview of methods for polygenic risk scores (PRS) development. Inputs are summary statistics from meta-analysis of multiple GWAS datasets – BCAC ASN+ABCC denoted training dataset 1, BCAC ASN+BBJ1 denoted training dataset 2 and BCAC-EUR denoted training dataset 3 as described in the method section. LD ref: reference panel for linkage disequilibrium. LD ref: BCAC ASN denoted Oncoarray studies in BCAC Asian studies were used as reference panel; LD ref: BCAC EUR denoted BCAC studies of European ancestries were used as reference panel; 1000G ASN and 1000G EUR denoted the Asian and European samples, respectively, in 1000Genome Project. Figure 1 illustrate methods using East Asian ancestry women (Chinese and Malays) as an example, same methods were applied to South Asian ancestry women in the validation dataset.

Figure 2: Principal components analysis and mean of PRS_{46}+PRS_{287_EB} by country and ethnicity. (a) Principal component plot by country. Principal components analysis of samples genotyped with OncoArray as listed in Table S1. The samples were grouped according to country (Thailand, Taiwan, Hong Kong, China, Korea and Japan). For Malaysia and Singapore (M+S), the samples were further categorized by their self-reported ethnic origin (Chinese, Malay and Indian). (b) Mean of standardised PRS_{46}+PRS_{287_EB} in controls by country. PRS was standardised by the control SDs of each study. Error bars represent 95% CI.

Figure 3: Absolute breast cancer risk by percentiles of PRS and PRS distribution by ancestry. . (a) Lifetime and (b) 10-year absolute risk of developing breast cancer for Chinese women calculated using Singaporean incidence and mortality data and OR per SD of PRS_{46}+PRS_{287_EB} in Chinese (1.56 as reported in Table S5). The gray dashed lines in the (a) and (b) represent the average lifetime risk and absolute 10-year risk, respectively, of Singaporean Chinese women. The red horizontal dashed line (2.3%) in the (b) represents the 10-year absolute risk of a 50-year old European women where screening is recommended; (c) the distribution of PRS_{46}+PRS_{287_EB} in Chinese, Indian and Malay-ancestry women, generated using ethnic-specific mean and standard deviation as reported in Table S5. Area under the curves represent the percentiles of PRS $_{287\_EB}$.The right vertical dashed line represents the 90th percentile cutoff for PRS distribution in Chinese-ancestry women; (d) the distritbution of European PRS (PRS_{287_EUR}) for women of European, Chinese, Malay, or Indian ancestry. The right vertical dashed line represents the 90th percentile cutoff for PRS distribution in European-ancestry women;

.

Table 1. Mean, standard deviation, and the association of polygenic risk scores (PRS) with breast cancer risk in women of East Asian ancestry

Method	PRS	Validation set[a]				Test set[b]			
		Cases Mean (SD)	Control Mean (SD)	OR per SD[†] (95% CI)	AUC	Cases Mean (SD)	Control Mean (SD)	HR per SD[*] (95% CI)	AUC[*]
[1] Clumping and Thresholding	[c]PRS_{46}	-0.387 (0.446)	-0.538 (0.443)	1.37 (1.32-1.42)	0.589	-0.299 (0.433)	-0.444 (0.438)	1.40 (1.25-1.56)	0.600
[2] Penalised regression	[c]PRS_{2985}	0.075 (0.455)	-0.082 (0.452)	1.41 (1.37-1.47)	0.598	0.107 (0.460)	-0.059 (0.458)	1.45 (1.311.61)	0.608
[3] EUR SNPs+ EUR weights	[c]PRS_{287_EUR}	0.865 (0.548)	0.640 (0.549)	1.50 (1.45-1.56)	0.615	0.876 (0.549)	0.679 (0.541)	1.46 (1.34-1.60)	0.609
[4] EUR SNPs +ASN weights	[c]PRS_{287_ASN}	-0.533 (0.445)	-0.714 (0.447)	1.50 (1.45-1.56)	0.614	-0.552 (0.448)	-0.731 (0.441)	1.49 (1.33-1.66)	0.608
[5] EUR SNPs+ EB weights	[c]PRS_{287_EB}	0.343 (0.491)	0.135 (0.492)	1.53 (1.47-1.58)	0.620	0.341 (0.493)	0.153 (0.485)	1.50 (1.35-1.65)	0.609
Combine [1] + [3]	[d]PRS_{46} + PRS_{287_EUR}	0.058 (0.440)	-0.134 (0.437)	1.54 (1.49-1.60)	0.623	0.103 (0.442)	-0.075 (0.436)	1.52 (1.36-1.70)	0.620
Combine [2] + [3]	[d]PRS_{2985} + PRS_{287_EUR}	0.062 (0.447)	-0.139 (0.444)	1.56 (1.50-1.61)	0.626	0.080 (0.454)	-0.106 (0.447)	1.54 (1.38-1.72)	0.622
Combine [1] + [4]	[d]PRS_{46} + PRS_{287_ASN}	0.052 (0.425)	-0.127 (0.423)	1.52 (1.47-1.58)	0.619	0.070 (0.425)	-0.113 (0.421)	1.52 (1.35-1.70)	0.621
Combine [2] + [4]	[d]PRS_{2985} + PRS_{287_ASN}	0.055 (0.430)	-0.130 (0.430)	1.54 (1.48-1.60)	0.621	0.057 (0.435)	-0.135 (0.427)	1.53 (1.37-1.72)	0.623
Combine [1] + [5]	[d]PRS_{46} + PRS_{287_EB}	0.061 (0.446)	-0.137 (0.443)	1.55 (1.50-1.61)	0.625	0.089 (0.447)	-0.089 (0.441)	1.53 (1.37-1.71)	0.621

Combine [2] + [5]	[d]PRS_{2985} + PRS_{287_EB}	0.063 (0.451)	-0.139 (0.449)	1.56 (1.51-1.62)	0.627	0.077 (0.455)	-0.120 (0.447)	1.55 (1.39-1.72)	0.623
[6] PRS-CSx	[d]PRS_{GW_EUR}+PRS_{GW_ASN}	0.082 (0.493)	-0.159 (0.489)	1.62 (1.52-1.68)	0.636	-0.145 (0.511)	-0.388 (0.511)	1.62 (1.46-1.80)	0.635

[a]Validation cohort which consist of 6,392 breast cancer cases and 6,638 control of Chinese- and Malay- ancestry from MyBrCa and SGBCC (Table S1).

[b]Prospective cohorts which consist of 89,898 control and 1,592 breast cancer cases from 3 prospective cohorts, Singapore Chinese Health Study (SCHS), China Kadoorie Biobank (CKB) and Korean Cancer Prevention Study-II Biobank (KCPS-II) (Table S1).

[c]PRSs were derived using 46, 2,985 and 287 selected SNPs respectively as described in the Method section.

[d] Combined PRSs were generated using the formula $C + \alpha_1 PRS_1 + \alpha_2 PRS_2$ where C, α_1 and α_2 are the weights obtained by fitting a logistic regression model with breast cancer as outcome, PRS_1 and PRS_2 as explanatory variables using the validation dataset. The weights for the considered combination of PRSs can be found in Table S5.

[†]Adjusted for first the 10 principal components and study, and standardised to SDs in controls of each PRS.

[*]Fixed effect meta-analysis of three prospective cohorts, SCHS, CKB and KCPS-II. HR per SD and AUC of individual studies can be found in Figure S5.

Table 2. Mean, standard deviation, and the association of polygenic risk scores (PRS) with breast cancer risk in women of South Asian ancestry

Method	PRS developed based on East Asians[a]	Validation set[b]			
		Cases Mean (SD)	Control Mean (SD)	OR per SD[†] (95% CI)	AUC
[1] Clumping and Thresholding	[a]PRS_{46}	-0.490 (0.388)	-0.548 (0.387)	1.18 (1.06-1.31)	0.546
[2] Penalised regression	[a]PRS_{2985}	0.059 (0.381)	-0.048 (0.376)	1.32 (1.19-1.46)	0.581
[3] EUR SNPs+ EUR weights	[a]PRS_{287_EUR}	0.482 (0.570)	0.251 (0.608)	1.49 (1.34-1.67)	0.614
[4] EUR SNPs +ASN weights	[a]PRS_{287_ASN}	-0.552 (0.493)	-0.720 (0.479)	1.43 (1.28-1.58)	0.592
[5] EUR SNPs+ EB weights	[a]PRS_{287_EB}	0.084 (0.521)	-0.127 (0.545)	1.50 (1.35-1.67)	0.613
Combine [1] + [3]	[c]PRS_{46} + PRS_{287_EUR}	-0.212 (0.420)	-0.376 (0.444)	1.48 (1.33-1.65)	0.611
Combine [2] + [3]	[c]PRS_{2985} + PRS_{287_EUR}	-0.166 (0.419)	-0.347 (0.441)	1.53 (1.37-1.71)	0.62
Combine [1] + [4]	[c]PRS_{46} + PRS_{287_ASN}	0.008 (0.431)	-0.135 (0.420)	1.42 (1.28-1.57)	0.591
Combine [2] + [4]	[c]PRS_{2985} + PRS_{287_ASN}	0.036 (0.425)	-0.121 (0.413)	1.46 (1.32-1.62)	0.602
Combine [1] + [5]	[c]PRS_{46} + PRS_{287_EB}	-0.157 (0.438)	-0.328 (0.455)	1.49 (1.33-1.66)	0.610
Combine [2] + [5]	[c]PRS_{2985} + PRS_{287_EB}	-0.119 (0.434)	-0.304 (0.449)	1.52 (1.37-1.70)	0.618
[6] PRS-CSx	[c]PRS_{GW_EUR}+ PRS_{GW_ASN}	-0.308 (0.501)	-0.546 (0.502)	1.62 (1.46-1.81)	0.633

[a] PRSs developed based on Chinese and Malay ancestry women in the validation dataset as described in Table 1. cohort from Chinese- and Malay- ancestry of MyBrCa and SGBCC as in Table 1.

[b]Evaluation of PRSs performance in 585 breast cancer cases and 1,018 controls of Indian ancestry women in the validation dataset (Table S1).

[c] Combined PRSs were generated using the formula $\alpha_0 + \alpha_1 PRS_1 + \alpha_2 PRS_2$ where α_0, α_1 and α_2 are the weights obtained by fitting a logistic regression model with breast cancer as outcome, PRS_1 and PRS_2 as explanatory variables using Chinese- and Malay- ancestries women in the validation dataset. The weights for the considered combination of PRSs can be found in Table S5.

[†]Adjusted for first the 10 principal components and study, and standardised to SDs in controls of each PRS.