

Published in final edited form as:

Genet Med. 2022 March 01; 24(3): 586–600. doi:10.1016/j.gim.2021.11.008.

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

A full list of authors and affiliations appears at the end of the article.

Abstract

Purpose—Non-European populations are under-represented in genetics studies, hindering clinical implementation of breast cancer polygenic risk scores (PRSs). 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 data set comprised 138,309 women from 17 case-control studies. PRSs were generated using a clumping and thresholding method, lasso penalized regression, an Empirical Bayes approach, a Bayesian polygenic prediction approach, or linear combinations of multiple PRSs. These PRSs were evaluated in 89,898 women from 3 prospective studies (1592 incident cases).

Results—The best performing PRS (genome-wide set of single-nucleotide variations [formerly single-nucleotide polymorphism]) had a hazard ratio per unit SD of 1.62 (95% CI = 1.46-1.80) and an area under the receiver operating curve of 0.635 (95% CI = 0.622-0.649). Combined Asian and

This work is licensed under a [CC BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

*Correspondence and requests for materials should be addressed to Weang-Kee Ho, School of Mathematical Sciences, Faculty of Science and Engineering, University of Nottingham Malaysia, Jalan Broga, Semenyih, 43500 Selangor, Malaysia. WeangKee.Ho@nottingham.edu.my . **Soo-Hwang Teo, Cancer Research Malaysia, 1 Jalan SS12/1A, Subang Jaya, 47500 Selangor, Malaysia. soohwang.teo@cancerresearch.my . Weang-Kee Ho and Mei-Chee Tai are co-first authors. Douglas F. Easton and Soo-Hwang Teo jointly supervised this work.

Author Information

Conceptualization: W.-K.H., J.S., A.C.A., D.F.E., S.-H.T.; Data Curation: J.D., N.M., K.Mi., J.Lo.; 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.L., D.F.E.; Project Administration: M.-C.T., M.K.B., Q.W., E.A.W.; Resources: T.H., K.R., V.K.M.T., B.K.T.T., S.M.T., E.Y.T., S.H.L., 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.R.Mu., A.L., A.H.W., C.-C.T., K.Ma., 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.B.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 and editing: W.-K.H., M.-C.T., J.D., X.S., J.Li., P.J.H., I.Y.M., K.L., Y.-H.J., S.-H.L., N.M., M.K.B., Q.W., K.Mi., J.Lo., E.A.W., T.H., K.R., V.K.M.T., B.K.T.T., S.M.T., E.Y.T., S.H.L., Y.-T.G., Y.Z., D.K., J.-Y.C., W.H., H.-B.L., M.K., Y.O., S.N., B.B.J., S.K.P., S.-W.K., C.-Y.S., P.-E.W., B.P., K.Mu., A.L., A.H.W., C.-C.T., K.Ma., 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.B.P., A.M.D., J.S., X.-O.S., C.-H.Y., A.M.T., A.C.A., W.Z., M.H., D.F.E., S.-H.T.

Ethics Declaration

The Malaysian Breast Cancer Genetic Study 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 China Kadoorie Biobank 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.

Conflict of Interest

The authors declare no conflicts of interest.

Additional Information

The online version of this article (<https://doi.org/10.1016/j.gim.2021.11.008>) contains supplementary material, which is available to authorized users.

European PRSs (333 single-nucleotide variations) had a hazard ratio per SD of 1.53 (95% CI = 1.37-1.71) and an area under the receiver operating curve of 0.621 (95% CI = 0.608-0.635). The distribution of the latter PRS was different across ethnic subgroups, confirming the importance of population-specific calibration for valid estimation of breast cancer risk.

Conclusion—PRSs developed in this study, from association data from multiple ancestries, can enhance risk stratification for women of Asian ancestry.

Keywords

Breast cancer; Genetic; Polygenic risk score; Risk prediction

Introduction

Genetic inheritance is an important risk factor for breast cancer.¹ Rare pathogenic variants in several susceptibility genes, including *BRCA1*, *BRCA2* and *PALB2*, confer increased risks of breast cancer²; however, a majority of the genetic variations in risk is polygenic owing to the fact that a large number of genetic variants combine, in which each genetic variant confers a small increase in risk. The effects of these variants can be summarized as polygenic risk scores (PRSs).^{3,4} Mavaddat et al³ developed and validated a 313 variant breast cancer PRS (PRS-313), using data from European-ancestry women 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 approximately 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 and other clinical management guidelines.³ This shows the potential of PRS to improve quantification of risk and consequently optimize breast cancer screening and prevention strategies.⁶

Non-European populations are under-represented in genetic studies, and this could limit PRS adoption and applicability⁷⁻⁹ and exacerbate health disparities.¹⁰ This is important for ethnic minorities in high income countries, where clinical evaluation of the European PRS-313 is already underway but perhaps more so in low- and middleincome 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 participants in 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 PRS-313, developed for Europeans, was predictive of breast cancer risk in Asian populations, although the effect size was somewhat smaller than that reported in European populations.¹⁴ However, an important outstanding question is whether a more predictive PRS using 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 variations (SNVs) (formerly singlenucleotide polymorphisms) and took forward SNVs that were significantly associated with breast cancer risk in Asians ($P <$

.05) for PRS derivation, resulting in a 44-SNV PRS.¹⁵ Although predictive, we have shown in our previous work that the discriminatory power of 44-SNV PRS (area under the receiver operating curve [AUC] = 0.586) was much lower than 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 using data from Asian populations and to validate their performance in prospective cohorts using the largest available breast cancer genetic study of Asian ancestry and (2) to assess the transferability of PRSs across Asian ethnic subgroups.

Materials and Methods

Study populations

The study population was divided into training, validation, and testing data sets. The training data sets included (1) set 1, which comprised 22,013 invasive cases and 22,114 controls of East Asian ancestry from studies participating in BCAC and Asia Breast Cancer Consortium (where GWAS summary statistics of SNVs significant up to $P < .0001$ were available); (2) set 2, which comprised 16,680 invasive cases and 83,414 controls of East Asian ancestry from studies participating in BCAC together with BioBank Japan (where GWAS summary statistics were available); and (3) set 3, which comprised 122,977 invasive cases and 105,974 controls of European ancestry participating in BCAC⁴ (where GWAS summary statistics were available). The validation data set comprised (1) 6392 invasive cases and 6638 controls of Chinese or Malay ancestry and (2) 585 invasive cases and 1018 controls of Indian ancestry participating in 2 multiethnic case-control studies: the Malaysian Breast Cancer Genetics study and the Singapore Breast Cancer Cohort study. The testing data set comprised 89,898 women (1595 incident cases) from 3 prospective cohorts of East Asian ancestry: the Singapore Chinese Health Study,¹⁶ the China Kadoorie Biobank,¹⁷ and the Korean Cancer Prevention Study Biobank.¹⁸ Supplemental Table 1 summarizes the study design, genotyping arrays, and the sample size in each study. Genotype calling, quality control procedures, and imputation methods have been described previously.^{4,19–23} Ancestry informative principal components (PCs) were available for Asian ancestry samples in the BCAC and validation data sets generated using methods as previously described.²⁴ See Supplemental Methods for more details.

All studies were approved by the relevant institutional ethics committees and review boards, and all participants provided written informed consent.

Statistical methods

PRSs were given by the following equation:

$$\text{PRS} = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \dots + \beta_m x_m$$

where x_k is the allele dosage for SNV k , β_k is the corresponding weight, and m is the total number of SNVs. PRSs were standardized to have unit SD in the control subjects. Logistic regression models, adjusted for the first 10 PCs and study, were used to estimate

odds ratios (ORs) for association between the standardized PRSs and breast cancer risk in the validation set. The studies in the validation set were genotyped in 2 batches and hence treated as different strata for the purposes of adjustment. Cox proportional hazard model, adjusted for the first 2 PCs for Singapore Chinese Health Study (SCHS) and Korean Cancer Prevention Study-II and the first 12 PCs for China Kadoorie Biobank, was used to estimate hazard ratios per SD (HR_{perSD}) for the association between the PRS and breast cancer risk in the test set. The discrimination of PRS was assessed using AUC. The HR_{perSD} and AUC were obtained individually for each study and combined using a fixed-effect meta-analysis. Test of heterogeneity between studies were obtained using *rma()* command in the *metafor* package in R version 3.6.1.²⁵

The approaches for SNVs selection to be included in PRS and the corresponding weights are described in subsequent sections. Figure 1 and Supplemental Figure 1 summarizes the methods and data set. The lists of SNVs and the weights for the PRS computation are given in Supplemental Tables 2 to 4.

Clumping and thresholding approach

Training data set 1 was used in these analyses. SNVs clumping (within 1 megabase pair windows) was conducted to remove highly correlated SNVs (pairwise correlation $r^2 > 0.9$); the SNV with the lowest *P* value for association in the correlated pairs was retained, resulting in 3050 SNVs. SNVs were further clumped within prespecified clumping window sizes and threshold of a correlation r^2 . PRSs were then computed using the subset of SNVs that were significant at a prespecified *P* value threshold (set at 5×10^{-8} and then increased in steps of 10^{-10} up to 10^{-3}). The PRS with the highest AUC in the validation data set was selected as the best PRS. The clumping and derivation of PRSs were performed using PRSice v2.11,²⁶ whereas the AUCs for PRSs were generated using the pROC package in R version 3.6.1.

To account for the joint effect of SNVs used to derive the best PRS, we computed the optimal weight, from the summary statistics, for SNV *j* using the following formula:

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

where $\underline{\gamma}' = R^{-1} \underline{\beta}'$, *R* is the correlation matrix between the SNV genotypes, $\underline{\beta}'$ is the predicted normalized marginal effect sizes of the SNVs, and p_j is the effect allele frequency of SNV *j* (see Supplemental Methods).

Lasso penalized regression

All 3050 SNVs described in clumping and thresholding (C + T) section were included in these analyses, together with genotype data from Asian controls in BCAC OncoArray studies for calculating linkage disequilibrium among SNVs. The analyses were run using the package *lassosum* in R²⁷ across different values of the penalty and shrinkage parameters, and the PRS giving the highest correlation between PRS and the disease status (default metric in the method) in the validation data set was selected.

Linear combination of European PRS with Asian PRS

Of the 313 SNVs included in PRS developed for European women,³ only 287 SNVs with imputation info score > 0.9 in validation data set were retained for subsequent analyses. Reported weights³ were used to derive the European PRS (hereafter denoted as PRS_{287_EUR}). Asian PRSs generated from C + T or lasso penalized regression were linearly combined with PRS_{287_EUR}. The relative contribution of each PRS were estimated by logistic regression using the validation data set.

Reweighting of European-based PRS

We considered 2 sets of weights for PRS derivation using the 287 SNVs: (1) Asian weights estimated from the training data set 1, taking into account the correlation between SNVs using equation (1) (hereafter denoted as PRS_{287_ASN}), and (2) weights based on a combination of the Asian and European weights using an Empirical Bayes (EB) approach (hereafter denoted as PRS_{287_EB}), where the optimal weight is given by the following equation:

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

Here, $\beta_{jA, EB}$ is the estimated posterior effect sizes in Asians given the data and p_j is the allele frequency for SNV j (see Supplemental Methods). Other approaches to combine European- and Asian-specific weights were also explored, including fixed effect meta-analysis, but only the method that gave the best AUC is presented in this study.

We also considered linear combinations of the reweighted European PRSs with Asian PRSs generated from C + T method or lasso penalized regression (as described before).

Bayesian polygenic prediction approach (PRS-CSx)

Training sets 2 and 3 were used as training data sets for PRS-CSx²⁸ together with Asians and Europeans in the 1000 Genomes Phase 3 project as linkage disequilibrium reference panels.²⁹ PRSs generated using European- (hereafter denoted as PRS_{GW_EUR}) and Asian-specific posterior weights (hereafter denoted as PRS_{GW_ASN}) were linearly combined (hereafter denoted as PRS_{GW_EUR} + PRS_{GW_ASN}) in the validation data set. The analyses were repeated across a range of global shrinkage parameter (ϕ), and the ϕ that gave the linear combination of PRSs with the highest AUC in the validation data set was selected as the optimal ϕ . Analyses were run using the published Python code-based tool in Github.²⁷

PRSs for the South Asian population

The predictive performance of PRSs developed for East Asian-ancestry women in Indian-ancestry women were assessed using AUC and OR per SD (OR_{perSD}). Given the much smaller sample size of Indian-ancestry women, we did not attempt to generate a South Asian-specific PRS, but we considered estimating the weights in the linear combinations of multiple PRSs using the South Asian validation data set.

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 Supplemental 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.^{30,31} For birth-cohort specific incidences, age-specific breast cancer incidences for the 1960-1969 and 1970-1979 birth cohorts were calculated using the data on breast cancer incidence in Singapore from 1968 to 2017.³⁰ For women born between 1980 and 1989, incidences could only be calculated up to age 35, and hence, breast cancer incidences were projected by assuming an annual increase in breast cancer incidence of 3.9%.³²

Results

Genetic diversity within Asian populations

Figure 1 summarizes the data set and methods used in this study. The populations are clustered, consistent with geography and population history, with the Chinese-ancestry women (Malaysia/Singapore/mainland China/Hong Kong/Taiwan) forming a distinct cluster that is genetically closer to Japanese/Koreans women than to Indian-ancestry women (Figure 2A). The Malay-ancestry women from Malaysia/Singapore are genetically closer to Chinese-ancestry women than to Indian-ancestry women. Given the large genetic distance between Indian-ancestry women from the other populations, the primary validation data set was based on Chinese-ancestry and Malay-ancestry women, and Indian-ancestry women were evaluated separately.

PRSs developed using Asian-specific SNVs

For C + T, SNVs were removed if they were within 250 kb of an SNV already selected and correlated at $r^2 > 0.1$, leaving 1326 SNVs for analysis. For East Asian-ancestry women, the best PRS was obtained at a P value threshold of 5.74×10^{-7} , resulting in a 46-SNV PRS (PRS₄₆) (Supplemental Figure 2), with OR_{perSD} (95% CI) of 1.35 (1.30-1.39; AUC = 0.586) (Table 1). Other combinations of clumping size and correlation threshold r^2 did not result in PRSs that showed appreciable improvement (Supplemental Figure 3).

For lasso penalized regression, the best PRS was obtained at penalty parameter (λ) = 0.014 and shrinkage parameter (s) = 0.9, resulting in a PRS that included 2985 SNVs (PRS₂₉₈₅) (Supplemental Figure 4), with OR_{perSD} (95% CI) of 1.41 (1.36-1.46; AUC = 0.596), slightly more predictive than the PRS₄₆ (Table 1).

Linear combinations of European and Asian PRSs

Combining PRS_{287_EUR} and PRS₄₆ (OR_{perSD} [95% CI] = 1.54 [1.49-1.60]; AUC = 0.623) yielded markedly higher predictive accuracy in East Asian-ancestry women than that achieved using the Asian-specific PRSs alone (Table 1). The improvement was marginal when compared with the predictive accuracy obtained using PRS_{287_EUR} alone (OR_{perSD} [95% CI] = 1.50 [1.45-1.56]; AUC = 0.615), but relative contribution of PRS₄₆ to the linear combination model was approximately 30% (Supplemental Table 5). Compared with PRS₄₆

+ PRS287_EUR, combining PRS287_EUR and PRS2985 further increased the OR_{perSD} and AUC.

PRSs developed by integrating Asian weights into the European PRS

For East Asian-ancestry women, PRS_{287_EB} (OR_{perSD} [95% CI] = 1.53 [1.47–1.58]; AUC = 0.620) was slightly more predictive than PRS_{287_ASN} (OR_{perSD} [95% CI] = 1.50 [1.45–1.56]; AUC = 0.615) and PRS_{287_EUR} (OR_{perSD} [95% CI] = 1.50 [1.45–1.56]; AUC = 0.614) and markedly more predictive than PRS₄₆ and PRS₂₉₈₅ (Table 1). Compared with PRS₄₆ + PRS_{287_EUR}, a linear combination of PRS_{287_EB} with PRS₄₆ further improved the PRS performance.

Continuous shrinkage PRSs (PRS-CSx)

The best combined PRS for East Asian-ancestry women was obtained at $\phi = 10^{-4}$ (Supplemental Table 6), with OR_{perSD} (95% CI) of 1.62 (1.52–1.68) and AUC of 0.636 for PRSGW_EUR + PRSGW_ASN, markedly better than all the PRSs described thus far (Table 1). This improvement was mainly driven by the contribution of PRSGW_EUR (OR_{perSD} [95% CI] = 1.59 [1.53–1.65]; AUC = 0.629). The OR_{perSD} (95% CI) and AUC for PRSGW_ASN alone was 1.44 (1.39–1.49) and 0.601, respectively, only slightly better than PRS₄₆ (Supplemental Table 6).

PRSs for Indian-ancestry population

The PRSs derived from East Asian-ancestry women (as shown in Table 1) were all predictive of risk in South Asian-ancestry women but the OR_{perSD} were less than that of East Asian-ancestry women. Although linear combination of Asian-based and European-based PRSs improved the PRS performance compared with individual PRSs in East Asians, the improvement of PRS performance in women of South Asian ancestry was observed only when PRS₂₉₈₅ was considered in the linear combination (Table 2). There was no improvement in the effect sizes when European-based PRS was combined with PRS₄₆. Whereas incorporating Asian weights via the EB approach improved the performance of PRSs in East Asians, there was no improvement in performance in women of South Asian ancestry. Re-estimating the weights of the combined models using South Asian-ancestry women in the validation data set did not lead to an appreciable difference in predictive performance (Supplemental Table 7).

Evaluation of PRSs in prospective cohorts

The predictive performance of PRSs in the East Asian-ancestry women was replicated in the prospective cohorts (Table 1). Thus, the effect size was smallest for PRS based on Asian data alone (HR_{perSD} [95% CI] = 1.40 [1.25–1.56] for PRS₄₆ and 1.45 [1.31–1.61] for PRS₂₉₈₅), larger for PRS based on the European PRS (HR_{perSD} [95% CI] = 1.50 [1.35–1.65] for PRS_{287_EB}), and still larger for PRS based on combining the Asian and European PRS (HR_{perSD} [95% CI] = 1.53 [1.37–1.71] for PRS₄₆ + PRS_{287_EB}). As in the validation data set, PRS generated using PRS-CSx showed the strongest association with breast cancer risk (HR_{perSD} [95% CI] = 1.62 [1.46–1.80]) and highest AUC (0.635). There was no evidence of heterogeneity in the hazard ratios among studies for any PRS (Supplemental Figure 5).

Absolute breast cancer risk predictions

We used $PRS_{46} + PRS_{287_EB}$ to show the potential of translating PRS into clinical tool for Asian population. Based on East Asian-ancestry women in the validation data set, the estimated breast cancer ORs (95% CI) for women in the lowest 1% and highest 1% of the PRS distribution were 0.53 (0.33-0.82) and 3.01 (2.25-4.06), respectively, compared with middle quintile. The estimated ORs did not differ from those predicted under a theoretical polygenic model in which the log OR increases linearly with the PRS (Supplemental Table 8). The corresponding lifetime risks of developing breast cancer by age 80 years, on current incidence rates, were approximately 2% and approximately 19% for women in the lowest 1% and highest 1% of the PRS distribution, respectively, respectively (Figure 3A). Assuming that a 10-year absolute risk threshold of 2.3%³³ is used to define women at sufficient risk to justify screening, approximately 12% of Chinese women would reach the risk threshold before or at age 40 (Figure 3B). Supplemental Figure 6 shows the distribution of the 10-year absolute risk at age 40 for women who were born between 1980 and 1989 using projected incidence rates (see Methods). It is projected that the proportion of women who would reach the risk threshold would rise to 29%.

Generalizability of PRS across Asian ethnic subgroups

We showed the generalizability of PRS across Asian ancestry population using the 3 ethnic groups in the validation set and $PRS_{46} + PRS_{287_EB}$ as an example. This combined PRS was predictive of risk in all ethnic groups, with the effect size being higher in Chinese-ancestry women than in Malay and Indian-ancestry women (OR_{perSD} [95% CI] = 1.56 [1.50-1.63] for Chinese vs 1.51 [1.39-1.64] for Malays and 1.49 [1.33-1.66] for Indians, heterogeneity $P = .983$) (Supplemental Figure 7, Supplemental Table 9). The PRS distribution was, however, different among the 3 ethnic groups. Although there was only a marginal difference in the SD, the means differed markedly, being highest in Chinese and lowest in Indians (mean [SD] in Chinese, Malay, and Indian controls were -0.118 [0.439], -0.197 [0.556], and -0.328 [0.455], respectively, P -values for pair-wise comparison of means $< .0001$) (Supplemental Table 9). Figure 3C shows that if the Chinese PRS distribution was applied to Indians without adjustment, the 95th percentile in Indians corresponds, approximately, to the 90th percentile in the Chinese population, resulting in underestimation of risk in Indian women. The difference in the PRS distributions is even more apparent when women of European ancestry is used as reference (Figure 3D).

The patterns of PRS distribution by population (Figure 2B) are mirrored in the genetic clusters shown in Figure 2A. The largest differences in the means of the standardized $PRS_{46} + PRS_{287_EB}$ were observed between the Indian-ancestry women and Japanese/Korean women (with Indians being the biggest outlier).

Discussion

Personalized 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 utility 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 (1) PRSs generated by integrating information from European ancestry and Asian ancestry GWAS data sets performed better than PRSs based purely on weights derived from single-ancestry GWAS data, and (2) there were substantial differences in PRS distributions across ethnic groups.

Based on the largest available breast cancer GWAS data sets, the best PRS for East Asian–ancestry women was based on PRS-CSx approach²⁸ ($PRS_{GW_EUR} + PRS_{GW_ASN}$). This PRS had a notably larger effect size than the European PRS (PRS_{287_EUR}) that we had previously shown to be the best breast cancer PRS for women of Asian ancestry¹⁴ (HR_{perSD} in prospective cohorts: 1.62 vs 1.46; Table 1). It is noteworthy that the predictive performance of this PRS was similar to that achieved in European populations (HR_{perSD} [95% CI] of 313-SNV PRS: 1.59 [1.54-1.64] as reported in Mavaddat et al³). However, despite the rapid drop in cost associated with next-generation sequencing, implementation of PRS comprising approximately 1 million SNVs can be practically more challenging than the implementation of the European PRS that included only 313 variants.

We showed that adaptations based on the European 313-SNV PRS can improve risk prediction in women of East Asian ancestry. First, incorporating SNVs identified in the Asian populations (PRS_{46}) improved predictive power. This approach of linearly combining PRSs may reduce the gap in prediction accuracy between European and non-European populations as described previously.³⁴ Second, incorporating Asian weights further improved predictive power ($PRS_{46} + PRS_{287_EB}$) but to a lesser extent. The 313-SNV PRS is being used in several clinical studies in European populations, including the MyPeBs (My Personal Breast Screening)⁸ and WISDOM (Women Informed to Screen Depending On Measures of risk)⁷ trials, and the $PRS_{46} + PRS_{287_EB}$ PRS would be relatively easy to implement in clinical settings.

The PRS generated for women of East Asian ancestry were also predictive for women of 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 (relative contribution of approximately 14%, Supplemental Table 5). These results suggest 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.^{10,35–37} 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 correspond, approximately, to the 90th percentile in the European population, resulting in overestimation of risk in these women (Figure 3D). To our knowledge, no

studies thus far have considered the transferability of breast cancer PRS within diverse Asian ethnic subgroups. Our results showed that although the effect sizes appeared to be similar across ethnic groups (Supplemental Table 8), the mean PRS distribution differed substantially across Asian populations (Supplemental Table 7, Figure 2B). For example, although Japanese, Koreans, and Han Chinese are conventionally classified as East Asians in genetic analyses, the mean PRSs were markedly different between these ethnic groups (Figure 2B). The differences are sufficiently large to affect risk classification, and thus, 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 multiethnic PRSs.³²

Our work is subject to several limitations. First, although we have showed that the predictive performance of European PRS can be improved by integrating weights from Asians using an EB approach, the absolute increase in predictive accuracies is marginal. Second, our studies focus on developing PRS without using individual-level training data. When such data are 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.³ or considering subtype-specific disease analyses to retain more informative variants. Third, our results showed that PRSs developed using Asian-derived GWAS data set showed significantly poorer performance than the European PRS, indicating that further improvement is likely to require much larger Asian discovery data set. Finally, PRSs were linearly combined using the validation data set, 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 data set for South Asian women 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, and between Asian and Europeans) emphasize the need for ethnic-specific calibration before translating PRS into practice for diverse Asian populations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Weang-Kee Ho^{1,2,*}, Mei-Chee Tai², Joe Dennis³, Xiang Shu^{4,5}, Jingmei Li^{6,7}, Peh Joo Ho⁷, Iona Y. Millwood^{8,9}, Kuang Lin⁸, Yon-Ho Jee¹⁰, Su-Hyun Lee¹¹, Nasim Mavaddat³, Manjeet K. Bolla³, Qin Wang³, Kyriaki Michailidou^{3,12,13}, Jirong Long⁴,

Eldarina Azfar Wijaya², Tiara Hassan², Kartini Rahmat¹⁴, Veronique Kiak Mien Tan^{15,16}, Benita Kiat Tee Tan^{15,16,17}, Su Ming Tan¹⁸, Ern Yu Tan¹⁹, Swee Ho Lim²⁰, Yu-Tang Gao²¹, Ying Zheng²², Daehee Kang^{23,24}, Ji-Yeob Choi^{24,25,26}, Wonshik Han^{24,27}, Han-Byoel Lee^{24,27}, Michiki Kubo²⁸, Yukinori Okada^{29,30,31}, Shinichi Namba²⁹, The BioBank Japan Project³², Sue K. Park^{23,24,33}, Sung-Won Kim³⁴, Chen-Yang Shen³⁵, Pei-Ei Wu³⁵, Boyoung Park³⁶, Kenneth R. Muir³⁷, Artitaya Lophatananon³⁷, Anna H. Wu³⁸, Chiu-Chen Tseng³⁸, Keitaro Matsuo^{39,40}, Hidemi Ito^{41,42}, Ava Kwong^{43,44,45}, Tsun L. Chan^{43,46}, Esther M. John^{47,48}, Allison W. Kurian^{47,48}, Motoki Iwasaki⁴⁹, Taiki Yamaji⁴⁹, Sun-Seog Kweon^{50,51}, Kristan J. Aronson⁵², Rachel A. Murphy^{53,54}, Woon-Puay Koh^{55,56}, Chiea-Chuen Khor⁵⁷, Jian-Min Yuan^{58,59}, Rajkumar Dorajoo^{57,60}, Robin G. Walters^{8,9}, Zhengming Chen^{8,9}, Liming Li⁶¹, Jun Lv⁶¹, Keum-Ji Jung⁶², Peter Kraft^{10,63}, Paul D.B. Pharoah^{3,64}, Alison M. Dunning⁶⁴, Jacques Simard⁶⁵, Xiao-Ou Shu⁴, Cheng-Har Yip⁶⁶, Nur Aishah Mohd Taib⁶⁷, Antonis C. Antoniou³, Wei Zheng⁴, Mikael Hartman^{6,68,69}, Douglas F. Easton^{3,64}, Soo-Hwang Teo^{2,67,**}

Affiliations

¹School of Mathematical Sciences, Faculty of Science and Engineering, University of Nottingham Malaysia, Selangor, Malaysia

²Cancer Research Malaysia, Selangor, Malaysia

³Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom

⁴Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, TN

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

⁶Department of Surgery, University Surgical Cluster, National University Hospital, Singapore, Singapore

⁷Genome Institute of Singapore, Laboratory of Women's Health and Genetics, Singapore, Singapore

⁸Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom

⁹MRC Population Health Research Unit, University of Oxford, Oxford, United Kingdom

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

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

¹²Biostatistics Unit, The Cyprus Institute of Neurology & Genetics, Ayios Dometios, Cyprus

- ¹³Cyprus School of Molecular Medicine, The Cyprus Institute of Neurology & Genetics, Ayios Dometios, Cyprus
- ¹⁴Biomedical Imaging Department, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia
- ¹⁵Department of Breast Surgery, Singapore General Hospital, Singapore, Singapore
- ¹⁶Division of Surgery and Surgical Oncology, National Cancer Center Singapore, Singapore, Singapore
- ¹⁷Department of General Surgery, Sengkang General Hospital, Singapore, Singapore
- ¹⁸Division of Breast Surgery, Changi General Hospital, Singapore, Singapore
- ¹⁹Department of General Surgery, Tan Tock Seng Hospital, Singapore, Singapore
- ²⁰KK Breast Department, KK Women's and Children's Hospital, Singapore, Singapore
- ²¹State Key Laboratory of Oncogene and Related Genes & Department of Epidemiology, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China
- ²²Shanghai Municipal Center for Disease Control and Prevention, Shanghai, China
- ²³Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea
- ²⁴Cancer Research Institute, Seoul National University, Seoul, Korea
- ²⁵Department of Biomedical Sciences, Seoul National University Graduate School, Seoul, Korea
- ²⁶Institute of Health Policy and Management, Medical Research Center, Seoul National University, Seoul, Korea
- ²⁷Department of Surgery, Seoul National University College of Medicine, Seoul, South Korea
- ²⁸RIKEN Center for Integrative Medical Sciences (IMS), Yokohama, Japan
- ²⁹Department of Statistical Genetics, Graduate School of Medicine, Faculty of Medicine, Osaka University, Suita, Japan
- ³⁰Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Suita, Japan
- ³¹Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan
- ³²Institute of Medical Science, The University of Tokyo, Tokyo, Japan
- ³³Integrated Major in Innovative Medical Science, Seoul National University College of Medicine, Seoul, Korea

- ³⁴Department of Surgery, Daerim Saint Mary's Hospital, Seoul, Korea
- ³⁵Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan
- ³⁶Department of Medicine, Hanyang University College of Medicine, Seoul, Korea
- ³⁷Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, The University of Manchester, Manchester, United Kingdom
- ³⁸Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA
- ³⁹Division of Cancer Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan
- ⁴⁰Division of Cancer Epidemiology, Nagoya University Graduate School of Medicine, Nagoya, Japan
- ⁴¹Division of Cancer Information and Control, Aichi Cancer Center Research Institute, Nagoya, Japan
- ⁴²Division of Descriptive Cancer Epidemiology, Nagoya University Graduate School of Medicine, Nagoya University, Nagoya, Japan
- ⁴³Hong Kong Hereditary Breast Cancer Family Registry, Cancer Genetics Centre, Happy Valley, Hong Kong
- ⁴⁴Department of Surgery, The University of Hong Kong, Pok Fu Lam, Hong Kong
- ⁴⁵Department of Surgery, Hong Kong Sanatorium & Hospital, Happy Valley, Hong Kong
- ⁴⁶Department of Pathology, Hong Kong Sanatorium & Hospital, Happy Valley, Hong Kong
- ⁴⁷Division of Oncology, Department of Medicine, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA
- ⁴⁸Department of Epidemiology & Population Health, Stanford University School of Medicine, Stanford, CA
- ⁴⁹Division of Epidemiology, Center for Public Health Sciences, National Cancer Center, Tokyo, Japan
- ⁵⁰Department of Preventive Medicine, Chonnam National University Medical School, Hwasun, Korea
- ⁵¹Jeonnam Regional Cancer Center, Chonnam National University Hwasun Hospital, Hwasun, Korea
- ⁵²Department of Public Health Sciences, and Cancer Research Institute, School of Medicine, Queen's University, Kingston, Ontario, Canada
- ⁵³Cancer Control Research, BC Cancer, Vancouver, British Columbia, Canada

- ⁵⁴School of Population and Public Health, Faculty of Medicine, The University of British Columbia, Vancouver, British Columbia, Canada
- ⁵⁵Healthy Longevity Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
- ⁵⁶Singapore Institute for Clinical Sciences, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore
- ⁵⁷Genome Institute of Singapore, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore
- ⁵⁸Division of Cancer Control and Population Sciences, UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, PA
- ⁵⁹Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA
- ⁶⁰Health Services and Systems Research, Duke-NUS Medical School, Singapore, Singapore
- ⁶¹Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, Beijing, China
- ⁶²Institute for Health Promotion, Graduate School of Public Health, Yonsei University, Seoul, Korea
- ⁶³Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA
- ⁶⁴Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, United Kingdom
- ⁶⁵Genomics Center, CHU de Québec-UniversitéLaval Research Center, Quebec City, Quebec, Canada
- ⁶⁶Subang Jaya Medical Centre, Selangor, Malaysia
- ⁶⁷Department of Surgery, Faculty of Medicine, University of Malaya Centre, UM Cancer Research Institute, Kuala Lumpur, Malaysia
- ⁶⁸Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore
- ⁶⁹Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore

Acknowledgments

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 Breast Cancer Association Consortium 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, the staff of the Copenhagen DNA laboratory, Julie M. Cunningham, and the staff of Mayo Clinic Genotyping Core Facility.

Malaysian Breast Cancer Genetic Study 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, Shivaani Mariapun, and Joanna Lim) for their contributions and commitment to this study.

For Singapore Breast Cancer Cohort, we want to thank the program manager Jenny Liu and clinical research co-ordinators/research assistants Siew-Li Tan, Siok-Hoon Yeo, 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 Khng for processing the DNA samples. We also want to thank all the participants' support to Singapore Breast Cancer Cohort.

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, doctors, and nurses who helped with the data collection process. CBCS thanks study participants, coinvestigators, 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. SBCGS thank study participants and research staff for their contributions and commitment to the studies.

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

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). For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission. 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 number: 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.

W.-K.H. is the recipient of L'Oreal-UNESCO For Women in Science National Fellowship. J.Li. is the recipient of a National Research Foundation Singapore Fellowship (NRF-NRFF2017-02). A.C.A., J.D. and N.M. are supported through Cancer Research UK grants (PPRPGM-Nov20/100002 and C12292/A20861). J.S. 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.

Breast Cancer Association Consortium is funded by Cancer Research UK (C1287/A16563, C1287/A10118), by 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 European Union's 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 National Institutes of Health (NIH) Grant U19 CA148065 and Cancer Research UK Grant C1287/A16563 and the PERSPECTIVE I&I 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 following grants: the European Community's Seventh Framework Programme under grant agreement number 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 NIH (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the Genetic Associations and Mechanisms in Oncology initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research for the Canadian Institutes of Health Research Team in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund. The DRIVE Consortium was funded by U19CA148065.

Malaysian Breast Cancer Genetic Study is funded by research grants from the Malaysian Ministry of Higher Education (UM.C/HIR/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). Singapore Breast Cancer Cohort is funded by NUS Start Up Grant, National University Cancer Institute Singapore Centre Grant, National Medical Research Council (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, United Kingdom. K.Mu. and A.L. are supported by the NIHR Manchester Biomedical Research Centre, by the Allan Turing Institute, and by the ICEP (Cancer Research UK [C18281/A19169]). CBCS is funded by the Canadian Cancer Society (grant number 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, 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 University of Southern California Cancer Surveillance Program that is supported under subcontract by the California Department of Health. The University of Southern California Cancer Surveillance Program 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 (BCFR) was supported by grant U01CA164920 from the US National Cancer Institute of the NIH. 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 BCFR nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR. The NCOBCS was supported by the National Cancer Center Research and Development Fund (Japan). The SBCGS was supported primarily by NIH grants R01CA64277, R01CA148667, UM1CA182910, R01CA235553 and R37CA70867. Biological sample preparation was performed by the Survey and Biospecimen Shared Resource that 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 Network U19 CA148065. SEBCS was supported by the 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 is supported by the Ministry of Education, Culture, Sports, Sciences and Technology from the Japanese Government); the Hwasun Cancer Epidemiology Study-Breast is supported by the Biobank of Chonnam National University Hwasun Hospital, a member of the Korea Biobank Network.

The Singapore Chinese Health Study was supported by grants from the National Medical Research Council, Singapore (NMRC/CIRG/1456/2016) and the National Institutes of Health, United States (R01 CA144034 and UM1 CA182876). China Kadoorie Biobank was supported as follows: Baseline survey and first resurvey: 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.

Data Availability

Summary statistics (odds ratios and confidence limits) for all single-nucleotide variations used in derivation of various polygenic risk scores are provided in Supplemental Tables 2 to 4 of the manuscript. Summary statistics of European breast cancer genome-wide association studies analysis used in this study can be accessed via Breast Cancer Association Consortium (BCAC) website (<http://bcac.ccge.medschl.cam.ac.uk/bcacdata/oncoarray/oncoarray-and-combined-summary-result/>). Summary statistics of genome-wide association studies analyses from The BioBank Japan Project can be accessed via The

BioBank Japan Project website (<https://pheweb.jp/pheno/BrC>). Request for access to individual level data from BCAC studies can be made via the Data Access Coordinating Committee of BCAC (BCAC Coordinator: BCAC@medschl.cam.ac.uk). Request for access to the Asia Breast Cancer Consortium data could be requested by submission of an inquiry to Wei Zheng (wei.zheng@vanderbilt.edu).

References

- Shiovitz S, Korde LA. Genetics of breast cancer: a topic in evolution. *Ann Oncol.* 2015; 26 (7) 1291–1299. DOI: 10.1093/annonc/mdv022 [PubMed: 25605744]
- Breast Cancer Association Consortium. Dorling L, Carvalho S, et al. Breast cancer risk genes - association analysis in more than 113,000 women. *N Engl J Med.* 2021; 384 (5) 428–439. DOI: 10.1056/NEJMoa1913948 [PubMed: 33471991]
- Mavaddat N, Michailidou K, Dennis J, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am J Hum Genet.* 2019; 104 (1) 21–34. DOI: 10.1016/j.ajhg.2018.11.002 [PubMed: 30554720]
- Michailidou K, Lindström S, Dennis J, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature.* 2017; 551 (7678) 92–94. DOI: 10.1038/nature24284 [PubMed: 29059683]
- Milne RL, Kuchenbaecker KB, Michailidou K, et al. Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. *Nat Genet.* 2017; 49 (12) 1767–1778. DOI: 10.1038/ng.3785 [PubMed: 29058716]
- Pashayan N, Antoniou AC, Ivanus U, et al. Publisher correction: personalized early detection and prevention of breast cancer: ENVISION consensus statement. *Nat Rev Clin Oncol.* 2020; 17 (11) 716. doi: 10.1038/s41571-020-0412-0 [PubMed: 32601456]
- Esserman LJ. WISDOM Study and Athena Investigators. The WISDOM Study: breaking the deadlock in the breast cancer screening debate. *NPJ Breast Cancer.* 2017; 3: 34. doi: 10.1038/s41523-017-0035-5 [PubMed: 28944288]
- MyPeBS personalizing breast screening. Accessed December 1, 2021 <https://mypebs.eu>
- Brooks JD, Nabi HH, Andrulis IL, et al. Personalized Risk Assessment for Prevention and Early Detection of Breast Cancer: Integration and Implementation (PERSPECTIVE I&I). *J Pers Med.* 2021; 11: 511. doi: 10.3390/jpm11060511 [PubMed: 34199804]
- Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet.* 2019; 51 (4) 584–591. DOI: 10.1038/s41588-019-0379-x [PubMed: 30926966]
- Youlten DR, Cramb SM, Yip CH, Baade PD. Incidence and mortality of female breast cancer in the Asia-Pacific region. *Cancer Biol Med.* 2014; 11 (2) 101–115. DOI: 10.7497/j.issn.2095-3941.2014.02.005 [PubMed: 25009752]
- Bhoo-Pathy N, Yip CH, Hartman M, et al. Breast cancer research in Asia: adopt or adapt Western knowledge? *Eur J Cancer.* 2013; 49 (3) 703–709. DOI: 10.1016/j.ejca.2012.09.014 [PubMed: 23040889]
- Heer E, Harper A, Escandor N, Sung H, McCormack V, Fidler-Benaoudia MM. Global burden and trends in premenopausal and postmenopausal breast cancer: a population-based study. *Lancet Glob Health.* 2020; 8 (8) e1027–e1037. DOI: 10.1016/S2214-109X(20)30215-1 [PubMed: 32710860]
- Ho WK, Tan MM, Mavaddat N, et al. European polygenic risk score for prediction of breast cancer shows similar performance in Asian women. *Nat Commun.* 2020; 11 (1) 3833 doi: 10.1038/s41467-020-17680-w [PubMed: 32737321]
- Wen W, Shu XO, Guo X, et al. Prediction of breast cancer risk based on common genetic variants in women of East Asian ancestry. *Breast Cancer Res.* 2016; 18 (1) 124. doi: 10.1186/s13058-016-0786-1 [PubMed: 27931260]
- Hankin JH, Stram DO, Arakawa K, et al. Singapore Chinese Health Study: development, validation, and calibration of the quantitative food frequency questionnaire. *Nutr Cancer.* 2001; 39 (2) 187–195. DOI: 10.1207/S15327914nc392_5 [PubMed: 11759279]

17. Chen Z, Chen J, Collins R, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol.* 2011; 40 (6) 1652–1666. DOI: 10.1093/ije/dyr120 [PubMed: 22158673]
18. Jee YH, Emberson J, Jung KJ, et al. Cohort profile: the Korean Cancer Prevention Study-II (KCPS-II) Biobank. *Int J Epidemiol.* 2018; 47 (2) 385–386f. DOI: 10.1093/ije/dyx226 [PubMed: 29186422]
19. Michailidou K, Hall P, Gonzalez-Neira A, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet.* 2013; 45 (4) 353–361. e3612 doi: 10.1038/ng.2563 [PubMed: 23535729]
20. Shu X, Long J, Cai Q, et al. Identification of novel breast cancer susceptibility loci in meta-analyses conducted among Asian and European descendants. *Nat Commun.* 2020; 11 (1) 1217 doi: 10.1038/s41467-020-15046-w [PubMed: 32139696]
21. Dorajoo R, Chang X, Gurung RL, et al. Loci for human leukocyte telomere length in the Singaporean Chinese population and trans-ethnic genetic studies. *Nat Commun.* 2019; 10 (1) 2491 doi: 10.1038/s41467-019-10443-2 [PubMed: 31171785]
22. Gan W, Walters RG, Holmes MV, et al. Evaluation of type 2 diabetes genetic risk variants in Chinese adults: findings from 93,000 individuals from the China Kadoorie Biobank. *Diabetologia.* 2016; 59 (7) 1446–1457. DOI: 10.1007/s00125-016-3920-9 [PubMed: 27053236]
23. The biobank Japan Project website. BioBank Japan PheWeb. Accessed December 1, 2021 <https://pheweb.jp/pheno/BrC>
24. Amos CI, Dennis J, Wang Z, et al. The OncoArray consortium: a network for understanding the genetic architecture of common cancers. *Cancer Epidemiol Biomarkers Prev.* 2017; 26 (1) 126–135. DOI: 10.1158/1055-9965.EPI-16-0106 [PubMed: 27697780]
25. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw.* 2010; 36 (3) 1–48. DOI: 10.18637/jss.v036.i03
26. Choi SW, O'Reilly PF. PRSice-2: polygenic risk score software for biobank-scale data. *GigaScience.* 2019; 8 (7) giz082 doi: 10.1093/gigascience/giz082 [PubMed: 31307061]
27. Mak TSH, Porsch RM, Choi SW, Zhou X, Sham PC. Polygenic scores via penalized regression on summary statistics. *Genet Epidemiol.* 2017; 41 (6) 469–480. DOI: 10.1002/gepi.22050 [PubMed: 28480976]
28. GitHub repository: PRS-CSx. GitHub, Inc; <https://github.com/getian107/PRS-CSx> [Accessed December 1, 2021]
29. 1000 Genomes Project Consortium. Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature.* 2015; 526 (7571) 68–74. DOI: 10.1038/nature15393 [PubMed: 26432245]
30. Government of Singapore. Age-specific death rates, annual. Government of Singapore; Updated August 2019. <https://data.gov.sg/dataset/age-specific-death-rates-annual> [Accessed January 31, 2021]
31. Forman, D, Bray, F, Brewster, DH. , et al. *Cancer Incidence in Five Continents. Vol. X.* International Agency for Research on Cancer; 2014.
32. Jara-Lazaro AR, Thilagaratnam S, Tan PH. Breast cancer in Singapore: some perspectives. *Breast Cancer.* 2010; 17 (1) 23–28. DOI: 10.1007/s12282-009-0155-3 [PubMed: 19701678]
33. DeSantis C, Ma J, Bryan L, Jemal A. Breast cancer statistics, 2013. *Cancer J Clin.* 2014; 64: 52–62.
34. Maárquez-Luna C, Loh PR. South Asian Type 2 Diabetes (SAT2D) Consortium, SIGMA Type 2 Diabetes Consortium, Price AL. Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genet Epidemiol.* 2017; 41 (8) 811–823. DOI: 10.1002/gepi.22083 [PubMed: 29110330]
35. Martin AR, Gignoux CR, Walters RK, et al. Human demographic history impacts genetic risk prediction across diverse populations. *Am J Hum Genet.* 2017; 100 (4) 635–649. DOI: 10.1016/j.ajhg.2017.03.004 [PubMed: 28366442]
36. Reisberg S, Ijjasenko T, Läll K, Fischer K, Vilo J. Comparing distributions of polygenic risk scores of type 2 diabetes and coronary heart disease within different populations. *PLoS One.* 2017; 12 (7) e0179238 doi: 10.1371/journal.pone.0179238 [PubMed: 28678847]

37. Isgut M, Sun J, Quyyumi AA, Gibson G. Highly elevated polygenic risk scores are better predictors of myocardial infarction risk early in life than later. *Genome Med.* 2021; 13 (1) 13. doi: 10.1186/s13073-021-00828-8 [PubMed: 33509272]

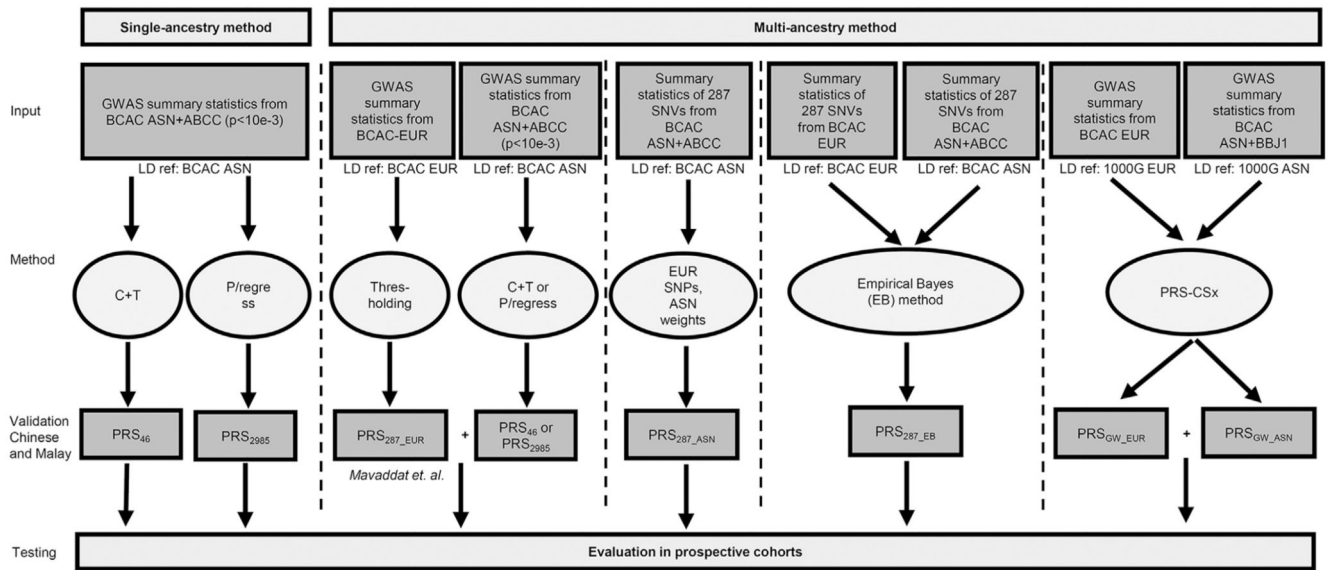


Figure 1. Overview of methods for PRSs development.

Inputs are summary statistics from the meta-analysis of multiple GWAS data sets—BCAC ASN + ABCC denotes training data set 1, BCAC ASN + BBJ denotes training data set 2, and BCAC-EUR denotes training data set 3 as described in the method section. LD ref: BCAC ASN denotes OncoArray studies in which BCAC Asian studies were used as reference panel; LD ref: BCAC EUR denotes BCAC studies in which European ancestries were used as reference panel; 1000G ASN and 1000G EUR denote the Asian and European samples, respectively, in 1000 Genomes Project. Figure 1 shows methods using East Asian–ancestry women (Chinese and Malays), and as an example, same methods were applied to South Asian–ancestry women in the validation data set. ABCC, Asia Breast Cancer Consortium; ASN, Asian; BBJ, The BioBank Japan Project; BCAC, Breast Cancer Association Consortium; C + T, clumping and thresholding; EUR, European; GWAS, genome-wide association study; LD ref, reference panel for linkage disequilibrium; PRS, polygenic risk score; SNV, single-nucleotide variation.

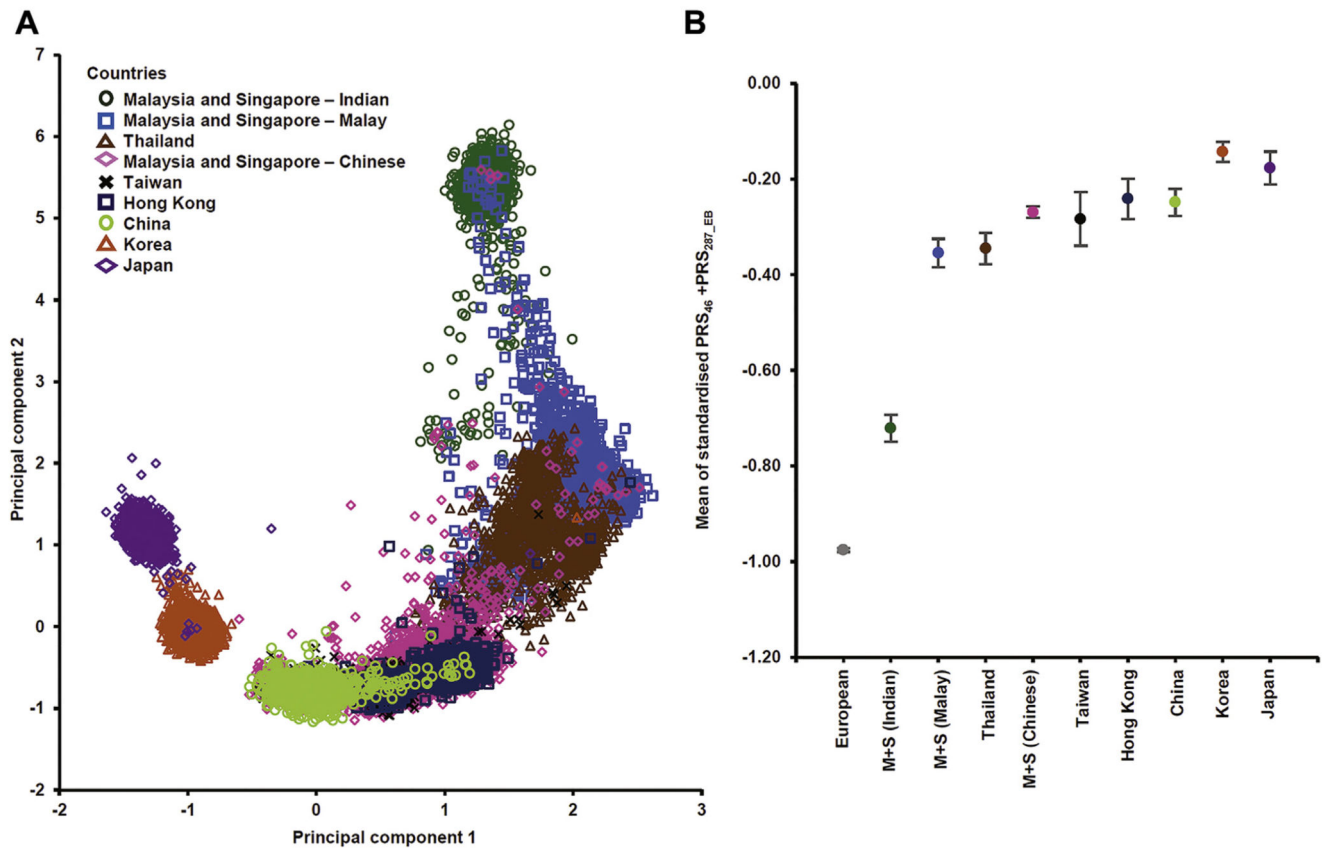


Figure 2. Principal components analysis and mean of PRS₄₆ + PRS_{287_EB} according to country and ethnicity.

(A) PC plotted according to country. PCs analysis of samples genotyped with OncoArray as listed in Supplemental Table 1. The samples were grouped according to country (Thailand, Taiwan, Hong Kong, China, Korea, and Japan). For M + S, the samples were further categorized by their self-reported ethnic origin (Chinese, Malay, and Indian). (B) Mean of standardized PRS₄₆ + PRS_{287_EB} in controls according to country. PRS was standardized according to the control SDs of each study. Error bars represent 95% CI. The mean of standardized PRS₄₆ + PRS_{287_EB} in European controls were included for reference. EB, Empirical Bayes; M + S, Malaysia and Singapore; PC, principal component; PRS, polygenic risk score.

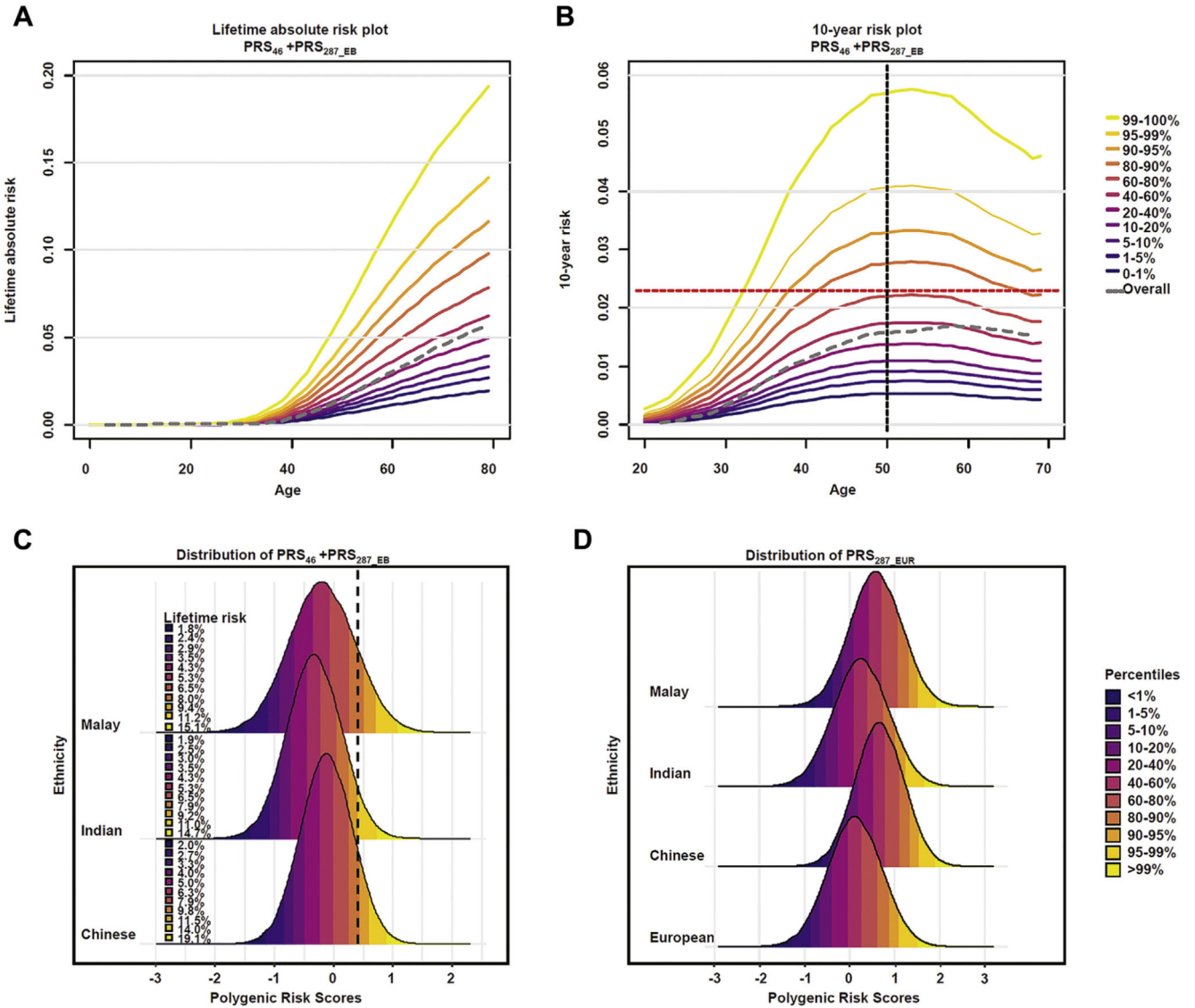


Figure 3. Absolute breast cancer risk by percentiles of PRS and PRS distribution by ancestry. (A) Lifetime and (B) 10-year absolute risk of developing breast cancer for Chinese women calculated using Singaporean incidence and mortality data and odds ratio per SD of PRS₄₆ + PRS_{287_EB} in Chinese (1.56 as reported in Supplemental Table 9). The gray dashed lines in the (A) and (B) represent the average lifetime risk and absolute 10-year risk, respectively, for Singaporean Chinese women. The red horizontal dashed line (2.3%) in the (B) represents the 10-year absolute risk for a 50-year old EUR women where screening is recommended; (C) the distribution of PRS₄₆ + PRS_{287_EB} in Chinese-ancestry, Indian-ancestry and Malay-ancestry women, generated using ethnic-specific mean and SD of controls as reported in Supplemental Table 9, and the 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.³⁰ 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; eg, the 95th percentile in Indians (lifetime risk

= 11%) corresponds, approximately, to the 90th percentile in the Chinese population. If Chinese PRS distribution was used as a reference, these Indian women would be categorized as 90th percentile and hence would be told that their corresponding lifetime risk was 9% instead of 11%; (D) the distribution of EUR PRS (PRS_{287_EUR}) for women of EUR ancestry, Chinese ancestry, Malay ancestry, or Indian ancestry. The right vertical dashed line represents the 90th percentile cutoff for PRS distribution in EUR-ancestry women. EB, Empirical Bayes; EUR, European; PRS, polygenic risk score.

Table 1
Mean, SD, and the association of PRSs with breast cancer risk in women of East Asian ancestry

Method	PRS	Validation Set ^a				Test Set ^b			
		Cases	Control	OR Per SD ^c	AUC	Cases	Control	HR Per SD ^d	AUC ^d
		Mean (SD)	Mean (SD)	(95% CI)		Mean (SD)	Mean (SD)	(95% CI)	
(1) Clumping and thresholding	$e_{PRS_{46}}$	-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) Penalized regression	$e_{PRS_{2985}}$	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 (1.31-1.61)	0.608
(3) EUR SNVs + EUR weights	$e_{PRS_{287_EUR}}$	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 SNVs + ASN weights	$e_{PRS_{287_ASN}}$	-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 SNVs + EB weights	$e_{PRS_{287_EB}}$	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)	$f_{PRS_{46} + PRS_{287_EUR}}$	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)	$f_{PRS_{2985} + PRS_{287_EUR}}$	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)	$f_{PRS_{46} + PRS_{287_ASN}}$	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)	$f_{PRS_{2985} + PRS_{287_ASN}}$	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)	$f_{PRS_{46} + PRS_{287_EB}}$	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)	$f_{PRS_{2985} + PRS_{287_EB}}$	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	$f_{PRS_{GW_EUR} + PRS_{GW_ASN}}$	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

ASN, Asian; AUC, area under the receiver operating curve; CKB, China Kadoorie Biobank; EB, Empirical Bayes; EUR, European; HR, hazard ratio; KCPS-II, Korean Cancer Prevention Study-II Biobank; MYBRCA, Malaysian Breast Cancer Genetic Study; OR, odds ratio; PRS, polygenic risk score; SCHS, Singapore Chinese Health Study; SGBCC, Singapore Breast Cancer Cohort; SNV, single-nucleotide variation.

^aValidation cohort that consisted of 6392 breast cancer cases and 6638 control of Chinese- and Malay-ancestry from MYBRCA and SGBCC (Supplemental Table 1).

^bProspective cohorts that consisted of 89,898 control and 1592 breast cancer cases from 3 prospective cohorts, SCHS, China Kadoorie Biobank (CKB), and KCPS-II (Supplemental Table 1).

^cAdjusted for the first 10 principal components and study, and standardized to SDs in controls of each PRS.

^dFixed effect meta-analysis of 3 prospective cohorts, SCHS, CKB and KCPS-II. HR per SD and AUC of individual studies can be found in Supplemental Figure 5.

^ePRSs were derived using 46, 2985 and 287 selected SNVs respectively as described in the Method section.

f Combined PRSs were generated using the formula $a_0 + a_1 PRS_1 + a_2 PRS_2$ where a_0 , a_1 and a_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 data set. The weights for the considered combination of PRSs can be found in Supplemental Table 5.

Table 2
Mean, SD, and the association of PRSs with breast cancer risk in women of South Asian ancestry

Method	PRS Developed on the basis of East Asians' data set ^a	Validation Set ^b			
		Cases	Control	OR Per SD ^c	
		Mean (SD)	Mean (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) Penalized regression	^a PRS ₂₉₈₅	0.059 (0.381)	-0.048 (0.376)	1.32 (1.19-1.46)	0.581
(3) EUR SNVs + EUR weights	^a PRS _{287_EUR}	0.482 (0.570)	0.251 (0.608)	1.49 (1.34-1.67)	0.614
(4) EUR SNVs + ASN weights	^a PRS _{287_ASN}	-0.552 (0.493)	-0.720 (0.479)	1.43 (1.28-1.58)	0.592
(5) EUR SNVs + 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)	^d PRS ₄₆ + PRS _{287_EUR}	-0.212 (0.420)	-0.376 (0.444)	1.48 (1.33-1.65)	0.611
Combine (2) + (3)	^d PRS ₂₉₈₅ + PRS _{287_EUR}	-0.166 (0.419)	-0.347 (0.441)	1.53 (1.37-1.71)	0.620
Combine (1) + (4)	^d PRS ₄₆ + PRS _{287_ASN}	0.008 (0.431)	-0.135 (0.420)	1.42 (1.28-1.57)	0.591
Combine (2) + (4)	^d PRS ₂₉₈₅ + PRS _{287_ASN}	0.036 (0.425)	-0.121 (0.413)	1.46 (1.32-1.62)	0.602
Combine (1) + (5)	^d PRS ₄₆ + PRS _{287_EB}	-0.157 (0.438)	-0.328 (0.455)	1.49 (1.33-1.66)	0.610
Combine (2) + (5)	^d PRS ₂₉₈₅ + PRS _{287_EB}	-0.119 (0.434)	-0.304 (0.449)	1.52 (1.37-1.70)	0.618
(6) PRS-CSx	^d PRS _{GW_EUR} + PRS _{GW_ASN}	-0.308 (0.501)	-0.546 (0.502)	1.62 (1.46-1.81)	0.633

ASN, Asian; AUC, area under the receiver operating curve; EB, Empirical Bayes; EUR, European; MYBRCA, Malaysian Breast Cancer Genetic Study; OR, odds ratio; PRS, polygenic risk score; SGBCC, Singapore Breast Cancer Cohort; SNV, single-nucleotide variation.

^aPRSs developed on the basis of Chinese and Malay-ancestry women in the validation data set as described in Table 1. Cohort from Chinese- and Malay-ancestry of MYBRCA and SGBCC as in Table 1.

^bEvaluation of PRSs performance in 585 breast cancer cases and 1018 controls of Indian-ancestry women in the validation dataset (Supplemental Table 1).

^cAdjusted for the first 10 principal components and study, and standardized to SDs in controls of each PRS.

^dCombined 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 estimated from East Asian ancestry women as described in Table 1. The weights for the considered combination of PRSs can be found in Supplemental Table 5.