Evaluation of three polygenic risk score models for the prediction of breast cancer risk in Singapore Chinese

Claire Hian Tzer Chan¹*, Prabhakaran Munusamy¹*, Sau Yeen Loke¹, Geok Ling Koh¹, Audrey Zhi Yi Yang¹, Hai Yang Law², Chui Sheun Yoon², Chow Yin Wong³, Wei Sean Yong⁴, Nan Soon Wong⁵,⁶, Raymond Chee Hui Ng⁵, Kong Wee Ong⁴, Preetha Madhukumar⁴, Chung Lie Oey⁴, Gay Hui Ho⁴,⁷, Puay Hoon Tan⁸, Min Han Tan⁵,⁹,¹⁰, Peter Ang⁵,⁶, Yoon Sim Yap⁵ and Ann Siew Gek Lee¹,¹¹,¹²

¹Division of Medical Sciences, Humphrey Oei Institute of Cancer Research, National Cancer Centre, Singapore
²DNA Diagnostic and Research Laboratory, KK Women’s and Children’s Hospital, Singapore
³Department of General Surgery, Singapore General Hospital, Singapore
⁴Department of Surgical Oncology, National Cancer Centre, Singapore
⁵Department of Medical Oncology, National Cancer Centre, Singapore
⁶Oncocare Cancer Centre, Gleneagles Medical Centre, Singapore
⁷Koong and Ho Surgery Centre, Singapore
⁸Department of Pathology, Singapore General Hospital, Singapore
⁹Institute of Bioengineering and Nanotechnology, Singapore
¹⁰Lucence Diagnostics Pte Ltd, Singapore
¹¹Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
¹²Office of Clinical and Academic Faculty Affairs, Duke-NUS Graduate Medical School, Singapore

*These authors contributed equally to the work

Correspondence to: Ann Siew Gek Lee, email: dmslg@nccs.com.sg

Keywords: breast cancer; single-nucleotide polymorphism; risk loci; genotyping; polygenic risk score

Received: July 27, 2017 Accepted: January 25, 2018 Published: January 31, 2018

Copyright: Chan et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Genome-wide association studies (GWAS) have proven highly successful in identifying single nucleotide polymorphisms (SNPs) associated with breast cancer (BC) risk. The majority of these studies are on European populations, with limited SNP association data in other populations. We genotyped 51 GWAS-identified SNPs in two independent cohorts of Singaporean Chinese. Cohort 1 comprised 1294 BC cases and 885 controls and was used to determine odds ratios (ORs); Cohort 2 had 301 BC cases and 243 controls for deriving polygenic risk scores (PRS). After age-adjustment, 11 SNPs were found to be significantly associated with BC risk. Five SNPs were present in <1% of Cohort 1 and were excluded from further PRS analysis. To assess the cumulative effect of the remaining 46 SNPs on BC risk, we generated three PRS models: Model-1 included 46 SNPs; Model-2 included 11 statistically significant SNPs; and Model-3 included the SNPs in Model-2 but excluded SNPs that were in strong linkage disequilibrium with the others. Across Models-1, -2 and -3, women in the highest PRS quartile had the greatest ORs of 1.894 (95% CI = 1.157–3.100), 2.013 (95% CI = 1.227–3.302) and 1.751 (95% CI = 1.073–2.856) respectively, suggesting a direct correlation between PRS and BC risk. Given the potential of PRS in BC risk stratification, our findings suggest the need to tailor the selection of SNPs to be included in an ethnic-specific PRS model.
INTRODUCTION

Advances in technology and large collaborative efforts have led to the success of genome-wide association studies (GWAS) in their discovery of multiple breast cancer (BC)-associated risk loci. Researchers are now able to identify regions or genes that were not previously thought to be associated with BC risk. To date, over 100 single nucleotide polymorphisms (SNPs) have been identified. Though many of these SNPs were identified in predominantly Caucasian populations [1–8], there are a handful of SNPs identified in Asian populations as well [9–15]. Many groups have also attempted to replicate these associations in larger cohorts and/or in cohorts of different ethnicities. However, some SNPs have been shown to be ethnic-specific and do not necessarily replicate in other ethnicities [12, 14, 16–23]. Fine-scale mapping has subsequently been carried out to identify functional SNPs associated with BC risk in a particular ethnic group [16, 23]. In more recent years, fine-scale mapping of regions identified by GWAS [24–27] and meta-analysis of existing GWAS [28–33] have also contributed to the growing number of SNPs associated with BC susceptibility. As breast cancer is a highly heterogeneous disease, association studies have also been performed to discover risk loci specific to a particular breast cancer histological type or hormone receptor subtype [3, 4, 8, 17, 28, 30, 33–36].

Though it has been demonstrated that these SNPs are associated with BC risk, the risk that a single variant confers is relatively low. Several groups have attempted to generate polygenic risk scores (PRS) derived from a combination of different selected SNPs to evaluate the cumulative effect of these SNPs [37, 38]. A PRS considers the odds ratio (OR) of each SNP and the total number of risk alleles an individual carries.

As new risk loci have recently been discovered [26, 27, 32, 33], this current study aimed to assess the association of these SNPs with BC risk in Singapore Chinese. Well-established BC risk-associated SNPs as well as 13 recently discovered SNPs that have not been previously genotyped in Asian populations were evaluated to determine if these SNPs are associated with BC risk in our Singapore Chinese population, and combinations of SNPs were used to generate PRS.

RESULTS

Genotyping and association of SNPs with BC risk

Genotyping of the 51 SNPs (Supplementary Table 1) was carried out on 1,670 BC patients and 1,189 healthy controls of Chinese ethnicity. After excluding samples that failed to reach 95% call rate for all assays, samples were further separated into two independent cohorts; Cohort 1 included 1294 cases and 885 controls to determine the association of the SNPs with BC risk, and Cohort 2 included 301 cases and 243 controls to derive PRS models. The demographics and clinico-pathological characteristics of these cases and controls are summarized in Supplementary Table 2. The mean age of cases and controls in Cohort 1 was 50.2 years and 42.7 years respectively, and that of Cohort 2 was 49.9 years and 42.0 years respectively. The differences in age between cases and controls in both cohorts were statistically different ($P < 2.2 \times 10^{-16}$).

All SNP assays had a call rate of more than 95.0% with an average call rate of 99.1%, and did not deviate from Hardy-Weinberg Equilibrium in controls. Five SNPs, rs554219, rs614367, rs75915166, rs78540526, and rs56069439 were present in less than 1% of Cohort 1, and were excluded from further PRS analysis. Associations of the remaining 46 SNPs with BC risk in our Singapore Chinese cohort are reflected in Supplementary Table 3.

Results from logistic regression analysis with and without age-adjustment revealed 10 common SNPs to be statistically significant via an additive model at $P < 0.05$ (Supplementary Table 3). It was also observed that another SNP, rs2981579, which was found to be significant in the analysis without age-adjustment, was no longer significant after age-adjustment. An additional SNP, rs745570, was also found to be significantly associated with BC risk only after age-adjustment.

Development of PRS models and their association with BC risk

PRS were generated based on unadjusted and age-adjusted ORs for 3 models: (1) Model-1 included all 46 SNPs investigated in this study; (2) Model-2 only included 11 statistically significant SNPs; and (3) Model-3 included 9 SNPs, after excluding SNPs in strong linkage disequilibrium (LD) with other SNPs (Supplementary Table 3). The PRS were identified to be statistically significant for BC risk, across all models (Table 1). It was also observed that across all models, the PRS ORs were higher for the 4th quartile when compared to the 1st quartile (Table 1). For instance, when using Model-1 which included unadjusted ORs of 46 SNPs, women in the 4th quartile had a 1.89-fold higher risk of BC compared to the 1st quartile. Similarly, with age-adjusted ORs the increase in BC risk was 1.89-fold.

Finally, the AUC for each of the different PRS models were obtained to evaluate how effective each model was. Model-1 had the highest AUC value among the 3 PRS models for unadjusted and age-adjusted ORs of 0.572 (95% CI = 0.523–0.620) and 0.566 (95% CI = 0.517–0.614) respectively. AUCs of Model-2 and -3 using unadjusted ORs were 0.570 (95% CI = 0.522–0.619) and 0.566 (95% CI = 0.517–0.614) respectively, while that of Model-2 and -3 using age-adjusted ORs were 0.565 (95% CI = 0.516–0.613) and 0.557 (95% CI = 0.508–0.606) respectively.
We assessed the association of 46 GWAS-identified SNPs with BC risk in Singapore Chinese and identified 11 SNPs to be significantly associated with increased BC risk. We also generated a PRS to measure the cumulative effect of variants, and to determine its discriminatory ability by means of AUC. Compared to other studies that have utilized PRS (Supplementary Table 4), this current study has included 7 new SNPs that have not been previously included in any other PRS. We have observed similar AUCs in our study as compared to previous studies, both in European and Asian populations (Table 2).

There has not been a common consensus on whether fewer or a greater number of SNPs would render a better PRS model. In two separate studies conducted in Asians, one obtained an AUC of 0.63 using only 8 SNPs in their PRS [39] while the other obtained an AUC of 0.606 using a 44-SNP PRS [38]. Both Asian studies had evaluated an initial higher number of SNPs but only included SNPs that were found to be statistically significant in their own study cohort for the calculation of their PRS. In comparison, a European study had an AUC of 0.68 obtained from a PRS model which included 76 SNPs [40]. These findings suggest a need to tailor the selection of SNPs to be specific for the populations being studied.

In addition, due to the significant differences in age between cases and controls, we performed age-adjustment for the determination of ORs of SNPs and PRS. We observed similar trends of ORs and PRS for both unadjusted and age-adjusted analysis, suggesting that PRS as a predictor for BC risk is independent of age in our population.

Using age-adjusted ORs, we constructed a PRS using the 11 SNPs found to be significantly associated with BC risk (Model-2) and obtained an AUC of 0.565. As some SNPs were in LD with each other and may thus be over-represented, we constructed a 9-SNP PRS which only included the SNPs with the strongest association in each LD block (Model-3). However, Model-3 had a slightly weaker discriminatory ability with an AUC of 0.557 as compared to Model-2. By generating a PRS with all 46 SNPs studied, a similar AUC was observed at 0.566. Though the remaining 35 SNPs, including 11 out of the 12 SNPs recently discovered by Michailidou et al. [32], were not found to be statistically significant with BC risk in our study, it is possible that some of these SNPs failed to reach statistical significance as our study could have had

| PRS quartile | Controls (n) | No. of SNPs included in the model | PRS derived from unadjusted ORs | PRS derived from age-adjusted ORs |
|--------------|-------------|---------------------------------|--------------------------------|----------------------------------|
|              |             |                                 | Cases (n) OR (95% CI) P-value | Casess (n) OR (95% CI) P-value     |
| Model-1      |             |                                 |                                |                                  |
| 1st          | 61          | 46                              | 53 Ref                         | 51 Ref                            |
| 2nd          | 61          | 72                              | 1.358 (0.823–2.244) 0.231      | 1.412 (0.852–2.338) 0.180         |
| 3rd          | 61          | 78                              | 1.472 (0.895–2.421) 0.128      | 1.627 (0.990–2.677) 0.055         |
| 4th          | 60          | 98                              | 1.880 (1.153–3.064) 0.011      | 1.894 (1.157–3.100) 0.011         |
| Trend        | 243         | 301                             | 1.325 (1.099 to 1.597) 0.003   | 1.301 (1.070–1.587) 0.009         |
| Model-2      |             |                                 |                                |                                  |
| 1st          | 61          | 11                              | 48 Ref                         | 49 Ref                            |
| 2nd          | 61          | 72                              | 1.542 (0.928–2.562) 0.095      | 1.551 (0.936–2.570) 0.088         |
| 3rd          | 61          | 74                              | 1.500 (0.901–2.496) 0.119      | 1.612 (0.975–2.666) 0.063         |
| 4th          | 60          | 107                             | 2.266 (1.384–3.710) 0.001      | 2.013 (1.227–3.302) 0.006         |
| Trend        | 243         | 301                             | 1.312 (1.077–1.600) 0.007      | 1.267 (1.051–1.568) 0.0293        |
| Model-3      |             |                                 |                                |                                  |
| 1st          | 61          | 9                               | 54 Ref                         | 54 Ref                            |
| 2nd          | 61          | 72                              | 1.333 (0.808–2.199) 0.260      | 1.333 (0.808–2.199) 0.260         |
| 3rd          | 61          | 77                              | 1.426 (0.867–2.344) 0.162      | 1.519 (0.927–2.488) 0.097         |
| 4th          | 60          | 98                              | 1.845 (1.134–3.003) 0.014      | 1.751 (1.073–2.856) 0.025         |
| Trend        | 243         | 301                             | 1.608 (1.149–2.250) 0.006      | 1.480 (1.039–2.120) 0.031         |

**DISCUSSION**

We assessed the association of 46 GWAS-identified SNPs with BC risk in Singapore Chinese and identified 11 SNPs to be significantly associated with increased BC risk. We also generated a PRS to measure the cumulative effect of variants, and to determine its discriminatory ability by means of AUC. Compared to other studies that have utilized PRS (Supplementary Table 4), this current study has included 7 new SNPs that have not been previously included in any other PRS. We have observed similar AUCs in our study as compared to previous studies, both in European and Asian populations (Table 2).

There has not been a common consensus on whether fewer or a greater number of SNPs would render a better PRS model. In two separate studies conducted in Asians, one obtained an AUC of 0.63 using only 8 SNPs in their PRS [39] while the other obtained an AUC of 0.606 using a 44-SNP PRS [38]. Both Asian studies had evaluated an initial higher number of SNPs but only included SNPs that were found to be statistically significant in their own study cohort for the calculation of their PRS. In comparison, a European study had an AUC of 0.68 obtained from a PRS model which included 76 SNPs [40]. These findings suggest a need to tailor the selection of SNPs to be specific for the populations being studied.

In addition, due to the significant differences in age between cases and controls, we performed age-adjustment for the determination of ORs of SNPs and PRS. We observed similar trends of ORs and PRS for both unadjusted and age-adjusted analysis, suggesting that PRS as a predictor for BC risk is independent of age in our population.

Using age-adjusted ORs, we constructed a PRS using the 11 SNPs found to be significantly associated with BC risk (Model-2) and obtained an AUC of 0.565. As some SNPs were in LD with each other and may thus be over-represented, we constructed a 9-SNP PRS which only included the SNPs with the strongest association in each LD block (Model-3). However, Model-3 had a slightly weaker discriminatory ability with an AUC of 0.557 as compared to Model-2. By generating a PRS with all 46 SNPs studied, a similar AUC was observed at 0.566. Though the remaining 35 SNPs, including 11 out of the 12 SNPs recently discovered by Michailidou et al. [32], were not found to be statistically significant with BC risk in our study, it is possible that some of these SNPs failed to reach statistical significance as our study could have had...
Table 2: Comparison of the studies on polygenic risk score (PRS) for breast cancer risk

| Study population                  | Our study | Lecarpentier et al., 2017 [44] | Haieb et al., 2017 [26] | Wen et al., 2016 [7] | Movaddat et al., 2015 [6] | Vachon et al., 2015 [13] | Lee et al., 2014 [25] | Zheng et al., 2010 [12] |
|-----------------------------------|-----------|-------------------------------|------------------------|----------------------|-------------------------|------------------------|----------------------|------------------------|
| Cases                             | All       | (Male BRCA1/2 mutation carriers) | Taiwanese              | East Asians          | European                | Caucasian              | Singapore Chinese    | Chinese                |
|                                   | All       |                              | All                    | ER1                  | ER2                     | 446                    | 11,760               | 33,673                 | 1,643                  | 411                    | 5,039                  |
| Controls                          | 243       | 1,469                         | All                    | ER1                  | ER2                     | 514                    | 11,612               | 33,381                 | 2,397                  | 1,212                  | 3,082                  |
| No. of SNPs studied               | Model-1   | (No. of SNPs included in PRS) | 51 (46)                | 102 (88)             | 102 (87)                | 102 (53)               | 13 (6)               | 78 (44)                | 77 (77)                | 76 (76)                | 51 (51)                | 12 (8)                 |
|                                   | Model-2   |                              | 51 (11)                | 102 (88)             | 102 (87)                | 102 (53)               | 13 (6)               | 78 (44)                | 77 (77)                | 76 (76)                | 51 (51)                | 12 (8)                 |
|                                   | Model-3   |                              | 51 (9)                 | 102 (88)             | 102 (87)                | 102 (53)               | 13 (6)               | 78 (44)                | 77 (77)                | 76 (76)                | 51 (51)                | 12 (8)                 |
|                                   | AUC       | 0.566                         | 0.565                  | 0.557                | 0.59                    | 0.59                   | 0.55                 | 0.598                  | 0.606                  | 0.622                  | 0.68                   | -                     | 0.63                  |

Any SNPs found to be associated with ER+ and ER- specific PRS respectively.

Insufficient power to detect the associations and additional studies of Asian ancestry are thus warranted to confirm if these SNPs are significantly associated with BC risk. GWAS and other discovery methods could also be done on Asian populations to further identify novel ethnic-specific SNPs that could have more significant associations with BC risk in Asians [41].

Of the 11 SNPs found to be statistically significant in our study, 4 SNPs were located on 6q25.1 (ESR1). 6q25.1 (ESR1) as a BC susceptible locus was first identified in Chinese [9], and additional SNPs in this region have been found to be associated with BC risk [6, 33, 42]. The SNPs with the strongest association with BC risk identified in our study (rs3757318, rs11155804, rs12662670 and rs2046210) were all located within this locus and each caused an increase in BC risk of about 40%, similar to previous studies carried out on Chinese [9] and South-East Asians [43]. It has been also observed that these variants tend to increase risk by a higher magnitude in Asians as compared to Europeans [42, 43], suggesting the importance of 6q25.1 as a BC susceptible region particularly in Asians. It is noted that the four SNPs exhibited the same statistical tendency and had similar ORs as they were in LD with each other.

Other significant associations identified in our study included variants on 5q11.2-CHD4P2 (rs150816625), 10q22.3-ZMIZ1 (rs704010), 11p15.5-TNNT3 (rs909116), 12p11.2-PTHLH (rs7297051), 16q12.1-TOX3 (rs4784227), and 17q25-CBX8 (rs745570). With the exception of rs745570, all these other SNPs have been previously reported to be significantly associated with BC risk in Asian populations, with similar ORs and direction of effect. Rs745570 which maps to 17q25 (CBX8) was recently identified by Michailidou et al. [32]. Though a recent study has demonstrated that the expression of CBX8 promotes mammary tumorigenesis both in vivo and in vitro [44], information on 17q25 (CBX8) as a breast cancer susceptibility locus is limited. To the best of our knowledge, our study is the first to validate and confirm the association of rs745570 with increased BC risk in an Asian population.

10q21.2 (FGFR2) was one of the first BC susceptibility loci to be identified by early GWAS [1, 2]. Rs11200014, rs1219648, rs2981579 and rs2981582 on 10q21.2 have been found to be associated with BC risk across different ethnicities, and the variant alleles tend to have a slightly greater effect in Europeans (ORs of 1.23 to 1.31) as compared to Asians (ORs of 1.15 to 1.23) [1, 2, 45–47]. Similarly, we observed lower ORs of 1.13 to 1.15 in Singapore Chinese. Though these associations were only found to be of borderline significance, we should not discount the importance of FGFR2 as a BC susceptibility locus in our population.

In addition, our study is the first to investigate the associations of rs554219, rs75915166 and rs78540526, which map to 11q13.3 (CCND1), with BC risk in Asians. We also included an additional SNP at the same locus, rs614367, which has one of the strongest associations with BC risk and was one of the first few risk loci identified by GWAS [6]. The association of rs614367 with BC risk has also been confirmed in Asians [18]. These four SNPs were initially removed from association analysis as they were found in less than 1% of our cohort. Likewise, an earlier study has also demonstrated that the variant alleles of these SNPs are much rarer in Asians as compared to Europeans [48]. Notably, the ORs for these four SNPs at CCND1 ranged from 2.64 to 4.87, and were higher than the other SNPs in this study. As these rare variants are present in low frequencies, sufficiently powered studies of greater sample sizes are needed to further validate these findings.

Though the discriminatory ability of a PRS model has been inadequate for clinical use, it has considerable potential in improving risk modeling. It has been demonstrated that PRS models aid in refining the risk stratification of individuals who are already at an increased risk of developing breast cancer [37, 40, 49, 50]. Some groups have attempted to combine PRS with other BC risk factors, such as breast density [40] or features included in the Gail Model [51], and improvements to AUCs have been observed. In a study by Shieh et al. [52], the addition of a BCSC (Breast Cancer Surveillance System) to a PRS model was shown to improve discriminatory ability significantly.

Though the discriminatory ability of a PRS model has been inadequate for clinical use, it has considerable potential in improving risk modeling. It has been demonstrated that PRS models aid in refining the risk stratification of individuals who are already at an increased risk of developing breast cancer [37, 40, 49, 50]. Some groups have attempted to combine PRS with other BC risk factors, such as breast density [40] or features included in the Gail Model [51], and improvements to AUCs have been observed. In a study by Shieh et al. [52], the addition of a BCSC (Breast Cancer Surveillance System) to a PRS model was shown to improve discriminatory ability significantly.
recent study by Michailidou identified SNPs [13, 26, 27, 56–59], including 12 from the
42, 46, 48, 55] were selected, as well as more recently
BC risk-associated SNPs [1, 2, 5–8, 15, 17, 29, 30, 39,
46, 48, 55] were performed the best. In addition, we performed
logistic regression analysis based on PRS quartiles which showed an overall trend across models and groups, and
the highest quartile predicted to have the highest OR thus implying a direct correlation between PRS and OR. By
improving risk prediction models, we will not only better
stratify individuals according to their risk groups, but we
could potentially also provide more efficient and effective
screening and prevention methods.

MATERIALS AND METHODS

Study cohort

The study utilised DNA from 1,670 patients diagnosed with BC and 1,189 healthy controls with no
known disease upon recruitment. All samples were
obtained from women of Chinese ancestry. Peripheral
blood samples were either obtained from unselected BC patients attending outpatient clinics at the National Cancer Centre and Singapore General Hospital or
were archival frozen peripheral blood samples of BC
patients from the SingHealth Tissue Repository. DNA
was extracted using an optimized in-house method [54].
Control samples comprised of archival DNA acquired
from the DNA Diagnostic and Research Laboratory, KK
Women’s and Children’s Hospital, Singapore. Ethics
approval for the study was given by the SingHealth
Centralized Institutional Review Board (CIRB Ref: 2008/478/B), and written informed consent was taken
from each participant.

SNP selection

The association of 51 SNPs with BC susceptibility
was assessed (Supplementary Table 1). SNPs were
selected based on two criteria: (1) the SNPs were
significantly associated with BC risk at a genome-
wide level ($P$ value $= 5 \times 10^{-8}$); (2) SNPs found to be
monomorphic in Chinese were excluded. Well-established
BC risk-associated SNPs [1, 2, 5–8, 15, 17, 29, 30, 39,
42, 46, 48, 55] were selected, as well as more recently
identified SNPs [13, 26, 27, 56–59], including 12 from the
recent study by Michailidou et al. [32].

SNP genotyping

High-throughput genotyping for the 51 SNPs was
carried out on 192.24 Dynamic Array™ integrated fluidic
circuits (IFC) (Fluidigm, CA, USA). TaqMan® SNP
Genotyping Assays (Applied Biosystems, CA, USA) were
employed, and the BioMark HD (Fluidigm) was used for
thermal cycling and fluorescence detection. Raw intensity
data were converted to genotype calls based on k-means
clustering using the Fluidigm SNP Genotyping Analysis
software.

Statistical analysis

SNP association analysis

A case-control study design was used to determine
the association between the SNPs and BC. Cohort 1
comprised of 1294 cases and 885 controls, and only
samples with a SNP genotype call rate of ≥95% were
included. Using the PLINK toolset [60], logistic regression
analysis was carried out to identify statistically significant
SNPs associated with BC. In addition, we performed
logistic regression analysis using age as a covariate
along with individual SNPs to determine its effect on BC
risk and calculated the age-adjusted ORs along with its
statistical significance. A $P$-value of ≤ 0.05 was considered
statistically significant.

Linkage disequilibrium analysis

Using the PLINK toolset, LD analysis of the SNPs
was performed to determine their non-random association
in our population. The LD pattern between SNPs were
measured using the correlation coefficient, $r^2$, where $r^2 \geq
0.5$ was considered moderate to strong.

Polygenic risk score analysis

An additional independent cohort with 301 cases
and 243 controls (Cohort 2) was used to construct the PRS.
We only considered SNPs with a minor allele frequency
>1% within our cohort from the SNP risk association
analysis to be included in the PRS models. To assess the
cumulative effect of the SNPs, we calculated a PRS by
summing the logOR of the SNP multiplied by the number
of risk alleles of the SNP across all selected SNPs in an
individual [37]. Two different PRS were calculated for
overall BC risk; using unadjusted and age-adjusted ORs.
Further, for each group, we derived three different PRS
models based on varying numbers of SNPs to be included
in the model. Model-1 included 46 SNPs found to be
significantly associated with BC from published studies
(Supplementary Table 1); Model-2 included statistically
significant SNPs ($P$-value ≤ 0.05) associated with BC; Model-3 included statistically significant SNPs ($P$-value
≤ 0.05) but excluded SNPs that were in moderate to strong LD (\(r^2 \geq 0.5\)) with each other.

To investigate the association between BC and PRS, logistic regression analysis was performed with PRS being a continuous variable [37]. In addition, ORs based on logistic regression models were estimated for different PRS quartiles with the first quartile being the reference. Finally, to determine the discriminating ability of the model, the area under the receiver operating characteristic (AUC) was estimated. Statistical analyses were performed using R version 3.4.1 and PASW statistics 18 software.

**Abbreviations**

AUC: area under the curve; BC: Breast Cancer; BCSC: Breast Cancer Surveillance Consortium; CI: confidence interval; GWAS: genome-wide association study; LD: linkage disequilibrium; OR: odds ratio; PRS: polygenic risk score; SNP: single nucleotide polymorphism.

**Author contributions**

A.S.G. Lee designed and supervised the study; H.Y. Law, C.S. Yoon, C.Y. Wong, W.S. Yong, N.S. Wong, R.C.H. Ng, P.H. Tan, M.H. Tan, P. Ang and Y.S. Yap provided samples and clinico-pathological information; C.H.T. Chan, S.Y. Loke, G.L. Koh and A.Z.Y. Yang performed the experiments; C.H.T. Chan, P. Munusamy and A.S.G. Lee analysed and interpreted the data; C.H.T. Chan, P. Munusamy and A.S.G. Lee wrote the manuscript; All authors reviewed and approved the final manuscript.

**ACKNOWLEDGMENTS**

We are grateful to the participants who have volunteered in this study. We also thank the SingHealth Tissue Repository for providing blood samples, and the Department of Clinical Research, SGH, for the use of their BioMark HD (Fluidigm) equipment.

**CONFLICTS OF INTEREST**

The authors declare no conflicts of interests.

**FUNDING**

This work was supported by the National Medical Research Council of Singapore.

**REFERENCES**

1. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struwing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, Healey CS, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature. 2007; 447:1087–1093.
2. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, Wacholder S, Wang Z, Welch R, Hutchinson A, Wang J, Yu K, Chatterjee N, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet. 2007; 39:870–874.
3. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, Masson G, Jakobsdottir M, Thorlacius S, Helgason A, Aben KK, Strobbe LJ, Albers-Akkers MT, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet. 2007; 39:865–869.
4. Stacey SN, Manolescu A, Sulem P, Thorlacius S, Gudjonsson SA, Jonsson GF, Jakobsdottir M, Bergh thorsson JT, Gudmundsson J, Aben KK, Strobbe LJ, Swinkels DW, van Engelenburg KC, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet. 2008; 40:703–706.
5. Ahmed S, Thomas G, Ghous saini M, Healey CS, Humphreys MK, Platte R, Morrison J, Maranian M, Pooley KA, Luben R, Eccles D, Evans DG, Fletcher O, et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. Nat Genet. 2009; 41:585–590.
6. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, Seal S, Ghous saini M, Hines S, Healey CS, Hughes D, Warren-Perry M, Tapper W, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. Nat Genet. 2010; 42:504–507.
7. Fletcher O, Johnson N, Orr N, Hosking FJ, Gibson LJ, Walker K, Zelenika D, Gut I, Heath S, Palles C, Coupland B, Broderick P, Schoemaker M, et al. Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. J Natl Cancer Inst. 2011; 103:425–435.
8. Haiman CA, Chen GK, Vachon CM, Canzian F, Dunning A, Millikan RC, Wang X, Ademuyiwa F, Ahmed S, Ambrosone CB, Baglietto L, Balleine R, Bandera EV, et al. A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. Nat Genet. 2011; 43:1210–1214.
9. Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang Y, Wen W, Levy S, Deming SL, Haines JL, Gu K, Fair AM, Cai Q, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Hum Mol Genet. 2008; 17:328–338.
10. Cai Q, Long J, Lu W, Qu S, Wen W, Kang D, Lee JY, Chen K, Shen H, Shen CY, Sung H, Matsuo K, Haiman CA, et al. Genome-wide association study identifies breast cancer risk variant at 10q21.2: results from the Asia Breast Cancer Consortium. Hum Mol Genet. 2011; 20:4991–4999.
identifies novel susceptibility loci for breast cancer. PLoS Genet. 2012; 8:e1002532.

12. Kim HC, Lee JY, Sung H, Choi JY, Park SK, Lee KM, Kim YJ, Go MJ, Li L, Cho YS, Park M, Kim DJ, Oh JH, et al. A genome-wide association study identifies a breast cancer risk variant in ERBB4 at 2q34: results from the Seoul Breast Cancer Study. Breast Cancer Res. 2012; 14:R56.

13. Cai Q, Zhang B, Sung H, Low SK, Kweon SS, Lu W, Shi J, Long J, Wen W, Choi JY, Noh DY, Shen CY, Matsuo K, et al. Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1. Nat Genet. 2014; 46:886–890.

14. Han MR, Long J, Choi JY, Low SK, Kweon SS, Zheng Y, Cai Q, Shi J, Guo X, Matsuo K, Iwasaki M, Shen CY, Kim MK, et al. Genome-wide association study in East Asians identifies two novel breast cancer susceptibility loci. Hum Mol Genet. 2016; 25:3361–3371.

15. Long J, Cai Q, Shu XO, Qu S, Li C, Zheng Y, Gu K, Wang W, Xiang YB, Cheng J, Chen K, Zhang L, Zheng H, et al. Identification of a Functional Genetic Variant at 16q12.1 for Breast Cancer Risk: Results from the Asia Breast Cancer Consortium. PLoS Genetics. 2010; 6:e1000100.

16. Long J, Shu XO, Cai Q, Gao YT, Zheng Y, Li G, Li C, Gu K, Wen W, Xiang YB, Lu W, Zheng W. Evaluation of breast cancer susceptibility loci in Chinese women. Cancer Epidemiol Biomarkers Prev. 2010; 19:2357–2365.

17. Ghousaini M, Fletcher O, Michailidou K, Turnbull C, Schmidt MK, Dickens E, Dennis J, Wang Q, Humphreys MK, Luccarini C, Baynes C, Conroy D, Maranian M, et al. Genome-wide association analysis identifies three new breast cancer susceptibility loci. Nat Genet. 2012; 44:312–318.

18. Zheng W, Zhang B, Cai Q, Sung H, Michailidou K, Shi J, Choi JY, Long J, Dennis J, Humphreys MK, Wang Q, Lu W, Gao YT, et al. Common genetic determinants of breast-cancer risk in East Asian women: a collaborative study of 23 637 breast cancer cases and 25 579 controls. Human Molecular Genetics. 2013; 22:2539–2550.

19. Li X, Zou W, Liu M, Cao W, Jiang Y, An G, Wang Y, Huang S, Zhao X. Association of multiple genetic variants with breast cancer susceptibility in the Han Chinese population. Oncotarget. 2016; 7:85483–85491. http://doi.org/10.18632/oncotarget.13402.

20. Zhang B, Li Y, Li L, Chen M, Zhang C, Zuo XB, Zhou FS, Liang B, Zhu J, Li P, Huang ZL, Xuan H, Li W, et al. Association study of susceptibility loci with specific breast cancer subtypes in Chinese women. Breast Cancer Res Treat. 2014; 146:503–514.

21. Xu M, Xu Y, Chen M, Li Y, Li W, Zhu J, Zhang M, Chen Z, Zhang X, Liu J, Zhang B. Association study confirms two susceptibility loci for breast cancer in Chinese Han women. Breast Cancer Res Treat. 2016; 159:433–442.

22. Chen Y, Fu F, Lin Y, Qiu L, Lu M, Zhang J, Qiu W, Yang P, Wu N, Huang M, Wang C. The precision relationships between eight GWAS-identified genetic variants and breast cancer in a Chinese population. Oncotarget. 2016; 7:75457–75467. http://doi.org/10.18632/oncotarget.12255.

23. Zheng Y, Ogundiran TO, Falusi AG, Nathanson KL, John EM, Hennis AJ, Amb S, Domchek SM, Rebbeck TR, Simon MS, Nemesure B, Wu SY, Leske MC, et al. Fine mapping of breast cancer genome-wide association studies loci in women of African ancestry identifies novel susceptibility markers. Carcinogenesis. 2013; 34:1520–1528.

24. Meyer KB, O'Reilly M, Michailidou K, Carlebur S, Edwards SL, French JD, Prathalingham R, Dennis J, Bolla MK, Wang Q, de Santiago I, Hopper JL, Tsimiklis H, et al. Fine-scale mapping of the FGFR2 breast cancer risk locus: putative functional variants differentially bind FOXA1 and E2F1. Am J Hum Genet. 2013; 93:1046–1060.

25. Glubb DM, Maranian MJ, Michailidou K, Pooley KA, Meyer KB, Kar S, Carlebur S, O'Reilly M, Betts JA, Hillman KM, Kaufmann S, Beesley J, Canisius S, et al. Fine-scale mapping of the 5q11.2 breast cancer locus reveals at least three independent risk variants regulating MAP3K1. Am J Hum Genet. 2015; 96:5–20.

26. Orr N, Dudbridge F, Dryden N, Maguire S, Novo D, Perrasik E, Johnson N, Ghousaini M, Hopper JL, Southey MC, Apicella C, Stone J, Schmidt MK, et al. Fine-mapping identifies two additional breast cancer susceptibility loci at 9q31.2. Hum Mol Genet. 2015; 24:2966–2984.

27. Shi J, Zhang Y, Zheng W, Michailidou K, Ghousaini M, Bolla MK, Wang Q, Dennis J, Lush M, Milne RL, Shu XO, Beesley J, Kar S, et al. Fine-scale mapping of the 8q24 locus identifies multiple independent risk variants for breast cancer. International Journal of Cancer. 2016; 139:1303–1317.

28. Siddiq A, Couch FJ, Chen GK, Lindstrom S, Eccles D, Millikan RC, Michailidou K, Stram DO, Beckmann L, Rhie SK, Ambrosone CB, Aittomaki K, Amiano P, et al. A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. Hum Mol Genet. 2012; 21:5373–5384.

29. Michailidou K, Hall P, Gonzalez-Neira A, Ghousaini M, Dennis J, Milne RL, Schmidt MK, Chang-Claude J, Bojesen SE, Bolla MK, Wang Q, Dicks E, Lee A, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet. 2013; 45:353–361, 361e351–352.

30. Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN, Orr N, Rhie SK, Riboli E, Feigelson HS, Le Marchand L, Buring JE, Eccles D, et al. Genome-wide association studies loci in women of African ancestry identifies multiple loci for breast cancer. International Journal of Cancer. 2016; 139:1303–1317.

31. Lindstrom S, Thompson DJ, Paterson AD, Li J, Gierach GL, Scott C, Stone J, Douglas JA, dos-Santos-Silva I, Fernandez-Navarro P, Verghese J, Smith P, Brown J, et al. Genome-wide association study identifies multiple loci for breast cancer in a Chinese population. Oncotarget. 2016; 7:75457–75467. http://doi.org/10.18632/oncotarget.12255.
associated with both mammographic density and breast cancer risk. Nat Commun. 2014; 5:5303.

32. Michailidou K, Beesley J, Lindstrom S, Canisius S, Dennis J, Lush MJ, Maranian MJ, Bolla MK, Wang Q, Shah M, Perkins BJ, Czene K, Eriksson M, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. Nat Genet. 2015; 47:373–380.

33. Couch FJ, Kuchenbaecker KB, Michailidou K, Mendoza-Fandino GA, Nord S, Lilyquist J, Olswoold C, Hallberg E, Agata S, Ahsan H, Attomaki K, Ambrosone C, Andrilus JL, et al. Identification of four novel susceptibility loci for oestrogen receptor negative breast cancer. Nat Commun. 2016; 7:11375.

34. Garcia-Closas M, Hall P, Nevanlinna H, Pooley K, Morrison J, Richesson DA, Bojesen SE, Nordestgaard BG, Axelsson CK, Arias JJ, Milne RL, Ribas G, Gonzalez-Neira A, et al. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. PLoS Genet. 2008; 4:e1000054.

35. Stevens KN, Vachon CM, Lee AM, Slager S, Lesnick T, Olswoold C, Fasching PA, Miron P, Eccles D, Carpenter JE, Godwin AK, Ambrosone C, Winqvist R, et al. Common breast cancer susceptibility loci are associated with triple-negative breast cancer. Cancer Res. 2011; 71:6240–6249.

36. Petridis C, Brook MN, Shah V, Kohut K, Gorman P, Caneppele M, Levi D, Papoulis E, Orr N, Cox A, Cross SS, Dosingo GA, Nord S, Lilyquist J, Peto J, et al. Genetic predisposition to ductal carcinoma in situ of the breast. Breast Cancer Res. 2016; 18:22.

37. Mavaddat N, Pharoah PD, Michailidou K, Tyrer J, Brook MN, Bolla MK, Wang Q, Dennis J, Dunning AM, Shah M, Luben R, Brown J, Bojesen SE, et al. Prediction of breast cancer risk based on profiling with common genetic variants. J Natl Cancer Inst. 2015; 107.

38. Wen W, Shu XO, Guo X, Cai Q, Long J, Bolla MK, Michailidou K, Dennis J, Wang Q, Gao YT, Zheng Y, Dunning AM, Garcia-Closas M, et al. Prediction of breast cancer risk based on common genetic variants in women of East Asian ancestry. Breast Cancer Res. 2016; 18:124.

39. Zheng W, Wen W, Gao YT, Shyr Y, Zheng Y, Long J, Li G, Li C, Gu K, Cai Q, Shu XO, Lu W. Genetic and clinical predictors for breast cancer risk assessment and stratification among Chinese women. J Natl Cancer Inst. 2010; 102:972–981.

40. Vachon CM, Pankratz VS, Scott CG, Haeberle L, Ziv E, Jensen MR, Brandt KR, Whaley DH, Olson JE, Heusinger K, Hack CC, Jud SM, Beckmann MW, et al. The contributions of breast density and common genetic variation to breast cancer risk. J Natl Cancer Inst. 2015; 107.

41. Chan CHT, Munusamy P, Loke SY, Koh GL, Wong ESY, Law HY, Yoon CS, Tan MH, Yap YS, Ang P, Lee ASG. Identification of Novel Breast Cancer Risk Loci. Cancer Research. 2017; 77:5428–5437.

42. Stacey SN, Sulem P, Zanon C, Gudjonsson SA, Thorleifsson G, Helgason A, Jonassdotir A, Besenbacher S, Kostic JP, Fackenthal JD, Hsu D, Adebamowo C, Ogundiran T, et al. Ancestry-Shift Refinement Mapping of the C6orf97-ESR1 Breast Cancer Susceptibility Locus. PLoS Genetics. 2010; 6:e1001029.

43. Hein R, Maranian M, Hopper JL, Kapuscinski MK, Southey MC, Park DJ, Schmidt MK, Brooks A, Hogervorst FBL, Bueno-de-Mesquita HB, Muir KR, Lophatananon A, Rattanamongkoul S, et al. Comparison of 6q25 Breast Cancer Hits from Asian and European Genome Wide Association Studies in the Breast Cancer Association Consortium (BCAC). PLOS ONE. 2012; 7:e42380.

44. Chung CY, Sun Z, Mullokandogov G, Bosch A, Qadeer ZA, Cihan E, Rapp Z, Parsons R, Aguirre-Ghiso JA, Farias EF, Brown BD, Gaspar-Maia A, Bernstein E. Cbx8 Acts Non-canonically with Wdr5 to Promote Mammary Tumorigenesis. Cell Rep. 2016; 16:472–486.

45. Sueta A, Itu H, Kawase T, Hirose K, Hosono S, Yatabe Y, Tajima K, Tanaka H, Iwata H, Iwase H, Matsuo K. A genetic risk predictor for breast cancer using a combination of low-penetrance polymorphisms in a Japanese population. Breast Cancer Res Treat. 2012; 132:711–721.

46. Raskin L, Pinchev M, Arad C, Lejbkowicz F, Tamir A, Rennert HS, Rennert G, Gruber SB. FGFR2 is a breast cancer susceptibility gene in Jewish and Arab Israeli populations. Cancer Epidemiol Biomarkers Prev. 2008; 17:1060–1065.

47. Dai J, Hu Z, Jiang Y, Shen H, Dong J, Ma H, Shen H. Breast cancer risk assessment with five independent genetic variants and two risk factors in Chinese women. Breast Cancer Res. 2012; 14:R17.

48. French JD, Ghoussaini M, Edwards SL, Meyer KB, Michailidou K, Ahmed S, Khan S, Maranian MJ, O’Reilly M, Hillman KM, Betts JA, Carroll T, Bailey PJ, et al. Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through long-range enhancers. Am J Hum Genet. 2013; 92:489–503.

49. Evans DG, Brentnall A, Byers H, Harkness E, Stavrinos P, Howell A, Newman WG, Cuzick J. The impact of a panel of 18 SNPs on breast cancer risk in women attending a UK familial screening clinic: a case–control study. Journal of Medical Genetics. 2017; 54:111–113.

50. Cuzick J, Brentnall AR, Segal P, Byers H, Reuter C, Detre S, Lopez-Knowles E, Sestak I, Howell A, Powles TJ, Newman WG, Dowsett M. Impact of a Panel of 88 Single Nucleotide Polymorphisms on the Risk of Breast Cancer in High-Risk Women: Results From Two Randomized Tamoxifen Prevention Trials. J Clin Oncol. 2017; 35:743–750.

51. Lee CP, Irwanto A, Salim A, Yuan JM, Liu J, Kho WP, Hartman M. Breast cancer risk assessment using genetic variants and risk factors in a Singapore Chinese population. Breast Cancer Res. 2014; 16:R64.
52. Shieh Y, Hu D, Ma L, Huntsman S, Gard CC, Leung JW, Tice JA, Vachon CM, Cummings SR, Kerlikowske K, Ziv E. Breast cancer risk prediction using a clinical risk model and polygenic risk score. Breast Cancer Res Treat. 2016; 159:513–525.

53. Hsieh YC, Tu SH, Su CT, Cho EC, Wu CH, Hsieh MC, Lin SY, Liu YR, Hung CS, Chiou HY. A polygenic risk score for breast cancer risk in a Taiwanese population. Breast Cancer Res Treat. 2017; 163:131–138.

54. Chan M, Chan MW, Loh TW, Law HY, Yoon CS, Than SS, Chua JM, Wong CY, Yong WS, Yap YS, Ho GH, Ang P, Lee AS. Evaluation of nanofluidics technology for high-throughput SNP genotyping in a clinical setting. J Mol Diagn. 2011; 13:305–312.

55. Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, Hankinson SE, Hutchinson A, Wang Z, Yu K, Chatterjee N, Garcia-Closas M, Gonzalez-Bosquet J, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nat Genet. 2009; 41:579–584.

56. Ahsan H, Halpern J, Kibriya MG, Pierce BL, Tong L, Gamazon E, McGuire V, Felberg A, Shi J, Jasmine F, Roy S, Brutus R, Argos M, et al. A Genome-wide Association Study of Early-onset Breast Cancer Identifies PFKM as a Novel Breast Cancer Gene and Supports a Common Genetic Spectrum for Breast Cancer at Any Age. Cancer epidemiology, biomarkers & prevention. 2014; 23:658–669.

57. Claussnitzer M, Dankel SN, Kim KH, Quen G, Meuleman W, Haugen C, Glunk V, Sousa IS, Beaudry JL, Puvvindran V, Abdennur NA, Liu J, Svensson PA, et al. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. New England Journal of Medicine. 2015; 373:895–907.

58. Couch FJ, Kuchenbaecker KB, Michailidou K, Mendoza-Fandino GA, Nord S, Lilyquist J, Olswold C, Hallberg E, Agata S, Ahsan H, Aittomäki K, Ambrosone C, Andrulis IL, et al. Identification of four novel susceptibility loci for oestrogen receptor negative breast cancer. Nature Communications. 2016; 7:11375.

59. Lei J, Rudolph A, Moysich KB, Behrens S, Goode EL, Bolla MK, Dennis J, Dunning AM, Easton DF, Wang Q, Benitez J, Hopper JL, Southey MC, et al. Genetic variation in the immunosuppression pathway genes and breast cancer susceptibility: a pooled analysis of 42,510 cases and 40,577 controls from the Breast Cancer Association Consortium. Human Genetics. 2016; 135:137–154.

60. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira Manuel AR, Bender D, Maller J, Sklar P, de Bakker Paul IW, Daly Mark J, Sham Pak C. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. American Journal of Human Genetics. 2007; 81:559–575.