


RESEARCH ARTICLE

Open Access



Overlap of high-risk individuals predicted by family history, and genetic and non-genetic breast cancer risk prediction models: implications for risk stratification

Peh Joo Ho^{1,2,3}, Weang Kee Ho^{4,5}, Alexis J. Khng¹, Yen Shing Yeoh³, Benita Kiat-Tee Tan^{6,7,8}, Ern Yu Tan^{9,10,11}, Geok Hoon Lim¹², Su-Ming Tan¹³, Veronique Kiak Mien Tan^{7,8}, Cheng-Har Yip¹⁴, Nur-Aishah Mohd-Taib^{15,16}, Fuh Yong Wong¹⁷, Elaine Hsuen Lim¹⁸, Joanne Ngeow^{10,18,19}, Wen Yee Chay¹⁸, Lester Chee Hao Leong²⁰, Wei Sean Yong¹⁷, Chin Mui Seah¹³, Siau Wei Tang²¹, Celene Wei Qi Ng²¹, Zhiyan Yan¹², Jung Ah Lee¹², Kartini Rahmat²², Tania Islam^{15,16}, Tiara Hassan⁴, Mei-Chee Tai⁴, Chiea Chuen Khor¹, Jian-Min Yuan^{23,24}, Woon-Puay Koh^{22,25}, Xueling Sim², Alison M. Dunning²⁶, Manjeet K. Bolla²⁷, Antonis C. Antoniou²⁷, Soo-Hwang Teo^{4,28†}, Jingmei Li^{1,3*†}  and Mikael Hartman^{2,3,21†}

Abstract

Background: Family history, and genetic and non-genetic risk factors can stratify women according to their individual risk of developing breast cancer. The extent of overlap between these risk predictors is not clear.

Methods: In this case-only analysis involving 7600 Asian breast cancer patients diagnosed between age 30 and 75 years, we examined identification of high-risk patients based on positive family history, the Gail model 5-year absolute risk [5yAR] above 1.3%, breast cancer predisposition genes (protein-truncating variants [PTV] in *ATM*, *BRCA1*, *BRCA2*, *CHEK2*, *PALB2*, *BARD1*, *RAD51C*, *RAD51D*, or *TP53*), and polygenic risk score (PRS) 5yAR above 1.3%.

Results: Correlation between 5yAR (at age of diagnosis) predicted by PRS and the Gail model was low ($r=0.27$). Fifty-three percent of breast cancer patients ($n=4041$) were considered high risk by one or more classification criteria. Positive family history, PTV carriership, PRS, or the Gail model identified 1247 (16%), 385 (5%), 2774 (36%), and 1592 (21%) patients who were considered at high risk, respectively. In a subset of 3227 women aged below 50 years, the four models studied identified 470 (15%), 213 (7%), 769 (24%), and 325 (10%) unique patients who were considered at high risk, respectively. For younger women, PRS and PTVs together identified 745 (59% of 1276) high-risk individuals who were not identified by the Gail model or family history.

Conclusions: Family history and genetic and non-genetic risk stratification tools have the potential to complement one another to identify women at high risk.

Keywords: Breast cancer, Polygenic risk score, Gail model, Protein-truncating variants, Risk-based screening

*Correspondence: lijm1@gis.a-star.edu.sg

†Soo-Hwang Teo, Jingmei Li and Mikael Hartman contributed equally to this work.

¹ Genome Institute of Singapore, Human Genetics, Singapore, Singapore
Full list of author information is available at the end of the article

Background

Multiple randomized controlled trials have shown that screening mammography reduces mortality from breast cancer for women who are over 50 years old [1].



Screening programs have been set up in different countries since the 1970s with recommendations on the screening interval and the age to start screening [2]. There is a strong consensus that women above 50 should attend routine screening. Recommendations for this group of women have remained largely unchanged for over 40 years [1, 2].

The benefit of screening mammography for younger women aged 40 to 49 years is less clear [1]. Recommendations for screening mammography across different countries and time periods are inconsistent and subject to change [2, 3]. In many cases, a woman is to make a personal choice based on risk factors such as personal and family history of the disease, often with the help of professional advice from a doctor [1].

With advances made in disease prediction, the approach to breast screening is now leaning towards a tailored, individual risk-based approach. For example, mammographic breast density, which refers to the proportion of fibroglandular breast tissue compared to fat seen on mammography, is a risk factor for breast cancer and is increasingly communicated to screening participants [4, 5]. Women with dense breasts are informed of their increased risk of breast cancer development and reduced sensitivity of mammography to detect breast cancer so that they can make a better-informed decision as to whether they should undergo supplemental imaging screening adjunct to mammography [5].

The risk of breast cancer is multifactorial. Apart from mammographic density, other known conventional risk factors include family history, menarche age, menopause age, height, body mass index, age at first childbirth, menopausal hormone therapy, and benign breast disease [6]. Many of these factors have been incorporated into prediction models to estimate the personal risk of developing breast cancer [7].

Breast cancer has a significant genetic component. It has been estimated that 27–31% of breast cancer risk may be explained by heritable factors [8, 9]. Frequently described breast cancer predisposition genes that are highly penetrant include *ATM*, *BRCA1*, *BRCA2*, *CHEK2*, *PALB2*, *BARD1*, *RAD51C*, *RAD51D*, and *TP53* [10]. However, pathogenic mutations in these genes are rare in the population. Polygenic risk scores (PRS) computed from another class of genetic variants that are of smaller effect sizes individually but more common in the population have shown promise to add information to better stratify individuals with different breast cancer risks as compared to age-based screening programs [11–13].

Non-genetic risk prediction models are attractive as they are non-invasive and are easier to implement in a general population or primary care screening setting. Currently, breast cancer risk prediction is predominantly

based on information on age, family history, lifestyle, and reproductive factors. These data can be collected at a low cost. There is evidence that genetics contribute to risk prediction but the data generation will incur additional costs to individuals or the health system. Hence, there is a need to evaluate how much information genetics can add to the identification of high-risk individuals over non-genetic risk factors. In this case-only analysis involving 7600 Asian breast cancer patients, we look at the overlap of individuals with a family history of breast cancer and those identified to be at high risk based on family history, the Gail model, breast cancer predisposition genes, and polygenic risk score.

Methods

Study populations

Breast cancer patients from two multi-ethnic populations recruited in Singapore and Malaysia, with ethnic groups of Chinese, Malay, and Indian descent were included in this study. These patients were recruited as part of the Singapore Breast Cancer Cohort (SGBCC) [14] and the Malaysian Breast Cancer Genetic Study (MyBrCa) [15]. Controls were recruited from the Singapore Multi-Ethnic Cohort (MEC) study [16] and the Malaysian Mammography Study (MyMammo) [15] for the calculation of the mean PRS and SD for the standard population. A prospective cohort study of healthy individuals—the Singapore Chinese Health Study (SCHS)—was used as a validation cohort [17]. Full details on each study and their respective DNA isolation and genotyping protocols are available in Additional file 1 [14–23].

Carriership of protein-truncating variants in nine breast cancer predisposition genes

Target-enriched sequencing libraries of germline DNA for the breast cancer cases (SGBCC and MyBrCa) were prepared at the Centre for Cancer Genetic Epidemiology (University of Cambridge) as part of a larger effort (Breast Cancer Risk after Diagnostic Gene Sequencing) [10]. Details of the library preparation, sequencing, variant calling, and quality control methods are given in Dorling et al. [10]. Protein-truncating variants (PTVs) include nonsense single-nucleotide variants (SNVs), frameshift insertions or deletions (indels), and splice-disrupting SNVs. PTVs occurring in the last exon of each gene were excluded to avoid including variants that do not lead to nonsense-mediated decay. Here, we studied nine genes found to be relevant for breast cancer risk as reported in Dorling et al. [10]—*ATM*, *BRCA1*, *BRCA2*, *CHEK2*, *PALB2*, *BARD1*, *RAD51C*, *RAD51D*, and *TP53*—in 192 and 193 breast cancer patients from

SGBCC and MyBrCa, respectively (Additional file 2 – Tables S1 and S2) [10].

Polygenic risk score (PRS)

PRS is estimated as the weighted sum of effect alleles in 313 single nucleotide polymorphisms (SNPs) found to be associated with breast cancer, using plink (version 3) with the *scoresum* option (full details in Additional file 1) [19].

Gail model

Data on breast cancer risk factors (age at menarche, age at first live birth, ever had a biopsy, and family history of breast cancer) were obtained from structured questionnaires. Family history of breast cancer was available as a binary variable (yes/no). Weights (logistic regression coefficients derived from the Gail model) and attributable risks of Asian-Americans (“Asian.AABCS”, *BCRA* package in R) were used in the calculation of the Gail model absolute risk [22].

Five-year absolute risk

Five-year absolute risk at the age of breast cancer diagnosis, for both PRS and the Gail model relative risk, was estimated for breast cancer patients aged between 30 and 75 from SGBCC and MyBrCa. The absolute risk was based on ethnic-specific or overall breast cancer incidence rates (period of 2013 to 2017) for Singapore Citizens and mortality rates (the year 2016) in Singapore (Additional file 3) [20, 23]. Both incidence and mortality rates were recorded in 5-year intervals. The 5-year absolute risk based on PRS was estimated using an iterative method detailed by Mavaddat et al. [13]. The 5-year absolute risk predicted by the Gail model was estimated using the method in the *BCRA* package in R [22]. Details on the calculation of 5-year absolute risks are available in Additional file 1.

Statistical analysis

Individual risk of developing breast cancer over 5 years based on PRS and Gail model

A comparison between the 5-year absolute risks predicted by PRS and the Gail model was examined using the Wilcoxon signed-rank test. Spearman’s correlation coefficient was estimated.

Classification of breast cancer patients into high- or low-risk groups

To illustrate a potential screening program where only high-risk individuals are screened, individuals were classified into high or low breast cancer risk groups. The following criteria were used to define high-risk

groups: (1) at least one first degree relative diagnosed with breast cancer or ovarian cancer (effect of family history), (2) 5-year absolute risk above 1.3% estimated by PRS (effect of common genetic variants), (3) 5-year absolute risk above 1.3% estimated by Gail risk model (effect of non-genetic variants), and (4) carrier-ship of PTV in *ATM*, *BRCA1*, *BRCA2*, *CHEK2*, *PALB2*, *BARD1*, *RAD51C*, *RAD51D*, or *TP53* (effect of rare genetic variants). The current recommendation is for women aged between 40 and 49 years to start screening when their individual 5-year risk is the same as or exceeds that of an average 50-year-old woman [24]. The threshold of 1.3% is equivalent to the 5-year absolute risk of developing breast cancer for an average Caucasian woman aged 50 years [24]. Cohen’s kappa was used to test pairwise concordance between the classification of breast cancer patients based on PRS and the Gail model [25].

Agreement between criteria

The agreement between pairs of different criteria to identify breast cancer patients at high risk was estimated using kappa scores. The incremental proportions of breast cancer patients identified as being at high risk are presented for 5-year age intervals.

Analyses were performed on a combined dataset of SGBCC and MyBrCa breast cancer patients and repeated for each cohort separately. In addition, the analyses were repeated for the prospective SCHS study, but without PTV.

Analysis was performed in R version 4.0.3.

Results

Description of SGBCC and MyBrCa breast cancer patients diagnosed between ages 30 and 75

Table 1 describes the summary characteristics of the combined 7600 breast cancer patients included in the case-only analysis from SGBCC ($n=4284$) and MyBrCa ($n=3316$). The median age at diagnosis was 53 years (interquartile range [IQR]: 45 to 59). Fifteen percent of the patients had first-degree relatives with breast ($n=1133$) and 2% with ovarian cancer ($n=151$). Five percent ($n=385$) of our breast cancer patients were PTV carriers with one or more of the nine known breast cancer predisposition genes [10] (Additional file 2 – Table S3). Additional file 2 – Table S4 describes the attributes of the breast cancer patients of three participating studies separately.

Low correlation between breast cancer absolute risk by PRS and the Gail model by study

The median 5-year absolute risk was 1.1 (IQR: 0.7 to 1.6) by PRS and 0.9 (IQR: 0.7 to 1.2) by the Gail model

Table 1 Description of 7600 breast cancer patients diagnosed between ages 30 and 70. Patients were recruited as part of the Singapore Breast Cancer Cohort (SGBCC) and the Malaysian Breast Cancer Genetic Study (MyBrCa). The description of the study is in Additional file Table S4. IQR interquartile range

Variable	Statistic
Median age at diagnosis (IQR)	52 (45–59)
Study	
SGBCC	4284 (56%)
MyBrCa	3316 (44%)
Case-type	
Incidence (enrolled within one year of diagnosis)	4511 (59%)
Prevalence	3087 (41%)
Missing	2 (0%)
Ethnicity	
Chinese	5724 (75%)
Malay	1145 (15%)
Indian	645 (8%)
Other	82 (1%)
Unknown	4 (0%)
Age at menarche, years, n (%)	
≥ 14	2212 (29%)
12 to 13	3968 (52%)
< 12	803 (11%)
Unknown	617 (8%)
Age at first birth, years, n (%)	
< 20	335 (4%)
20 to 25	1479 (19%)
25 to 30	2289 (30%)
≥ 30	1792 (24%)
Nulliparous	1294 (17%)
Unknown	411 (5%)
Family history of breast cancer	
No	6364 (84%)
Yes	1132 (15%)
Number of first degree relatives with ovarian cancer	
No	7444 (98%)
Yes	151 (2%)
Carriers of breast cancer predisposition genes, n (%)	
Non-carrier	7215 (95%)
Carrier	385 (5%)
Median 5-year absolute risk (IQR)	
Gail model relative risk	0.9 (0.7–1.2)
Polygenic risk score (PRS)	1.1 (0.7–1.6)

(Table 1). Spearman's correlation coefficient between the 5-year absolute risks by PRS and the Gail model was low ($r=0.27$) (Fig. 1A). The median difference between the 5-year absolute risks estimated by PRS and the Gail model was -0.097 (IQR: -0.591 to 0.296 , Wilcoxon

signed-rank test p value <0.001) (Fig. 1B). Comparison by cohorts (SGBCC and MyBrCa) are shown in Additional file 2 – Fig. S1.

PRS and the Gail model identified larger proportions of unique high-risk individuals among breast cancer patients compared to PTV carriership and family history

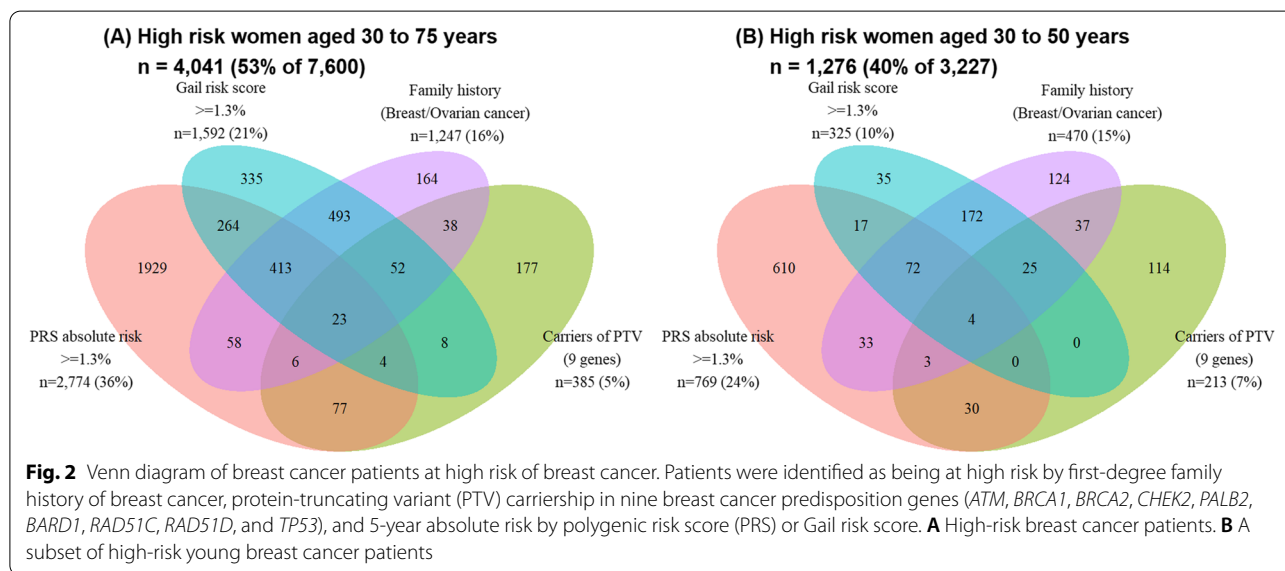
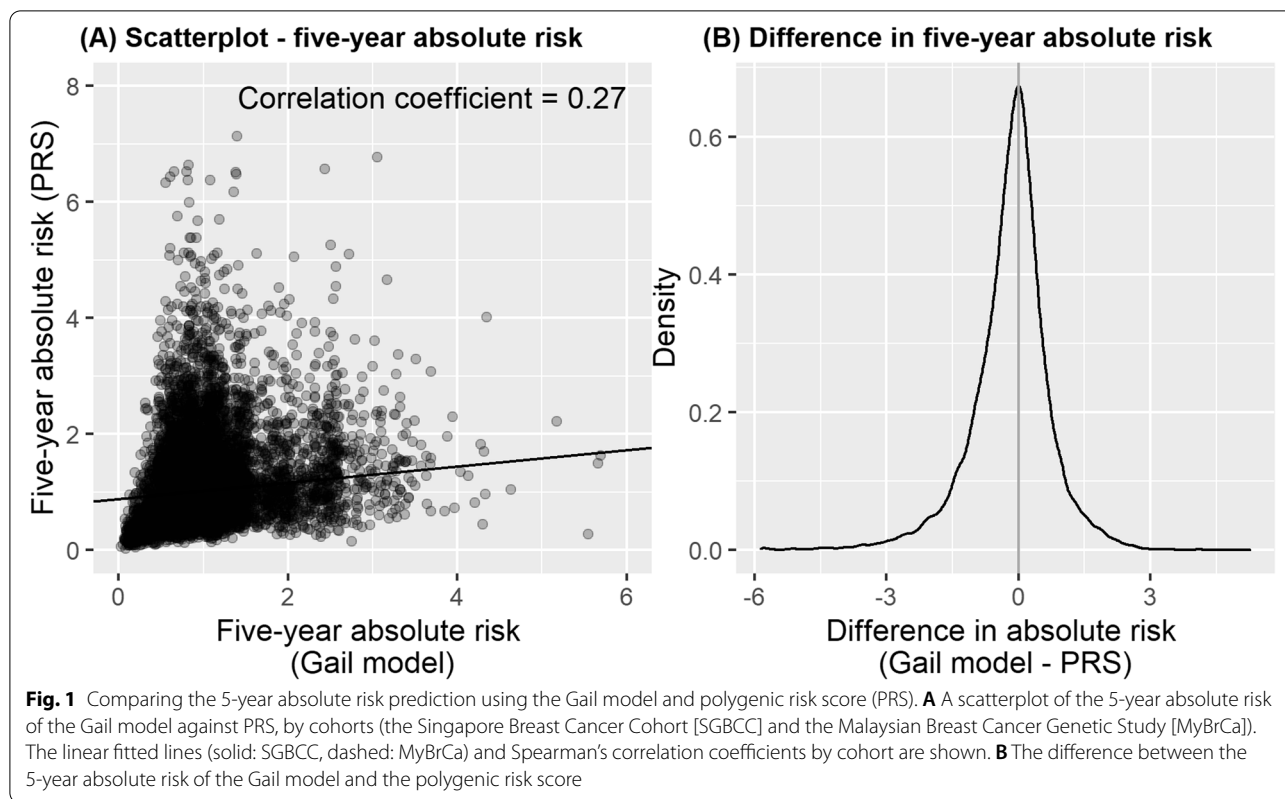
Approximately half of all (53%, $n=4041$) breast cancer patients were considered high risk by any of the four risk classification criteria studied (Fig. 2A). PRS (5-year absolute risk $\geq 1.3\%$) alone identified the largest proportion of high-risk patients ($n=2774$ (36%)) (Fig. 2A). This was followed by Gail model (5-year absolute risk $\geq 1.3\%$, $n=1592$ (21%)), positive family history ($n=1247$ (16%)), and PTV carriership ($n=385$ (5%)) (Fig. 2A). Among 385 PTV carriers, 110 (28%) were considered high risk by PRS. We observed poor or slight concordance between each pair of criteria in classifying patients as high risk; Cohen's kappa ranged from -0.025 to 0.095 with the exception of 0.621 for the Gail model with a family history of breast or ovarian cancer (Additional file 2 – Table S5).

Family history and genetic risk stratification models identified much younger breast cancer patients at high risk

For women below standard mammography screening entry age of 50 years (age 30 to 50 years; $n=3227$), the breast cancer risk stratification tools studied identified 40% ($n=1276$) of the breast cancer patients to be high risk. Risk stratification by positive family history, PTV carriership, and 5-year absolute risk $\geq 1.3\%$ by PRS or the Gail model identified 470 (15%), 213 (7%), 769 (24%), and 325 (10%) unique breast cancer patients who were considered at high risk of breast cancer, respectively (Fig. 2B). The genetic risk stratification models, PTV carriership and PRS, identified 114 and 610 additional high-risk individuals that were not identified by family history and the Gail model. Thirty-seven individuals were considered high risk based on both PTV carriership and PRS. Slight concordance was observed between the criteria based on the 5-year absolute risks by PRS and the Gail model in classifying young patients as high risk (Cohen's kappa: 0.052 , $p<0.001$, Additional file 2 – Table S6). Additional file 2 – Fig. S2 presents the classification of high-risk patients by study.

Proportion of breast cancer patients identified as being at high risk within 5-year age groups

As an individual criterion, the 5-year absolute risk, by PRS or the Gail model, identified the largest proportion of high-risk patients in the breast cancer screening age group (50 to 65 years) (Fig. 3). The 5-year absolute risk by PRS identified the largest proportion of high-risk patients



in age groups including ages of 40 years and above, with the highest proportion (50%) in the age group 60 to 64 years (Fig. 3). For both criteria (PRS or the Gail model), the proportion identified as at high risk decreases in the younger age groups, with less than 10% being identified in the youngest age group 30 to 34 years. Family history

remained at the level of less than 20% for all age groups younger than 70 years.

More than 50% in the older age groups (50 years and above) of breast cancer were identified as at high risk when we added information of each criterion sequentially (starting with family history \rightarrow above 1.3% 5-year

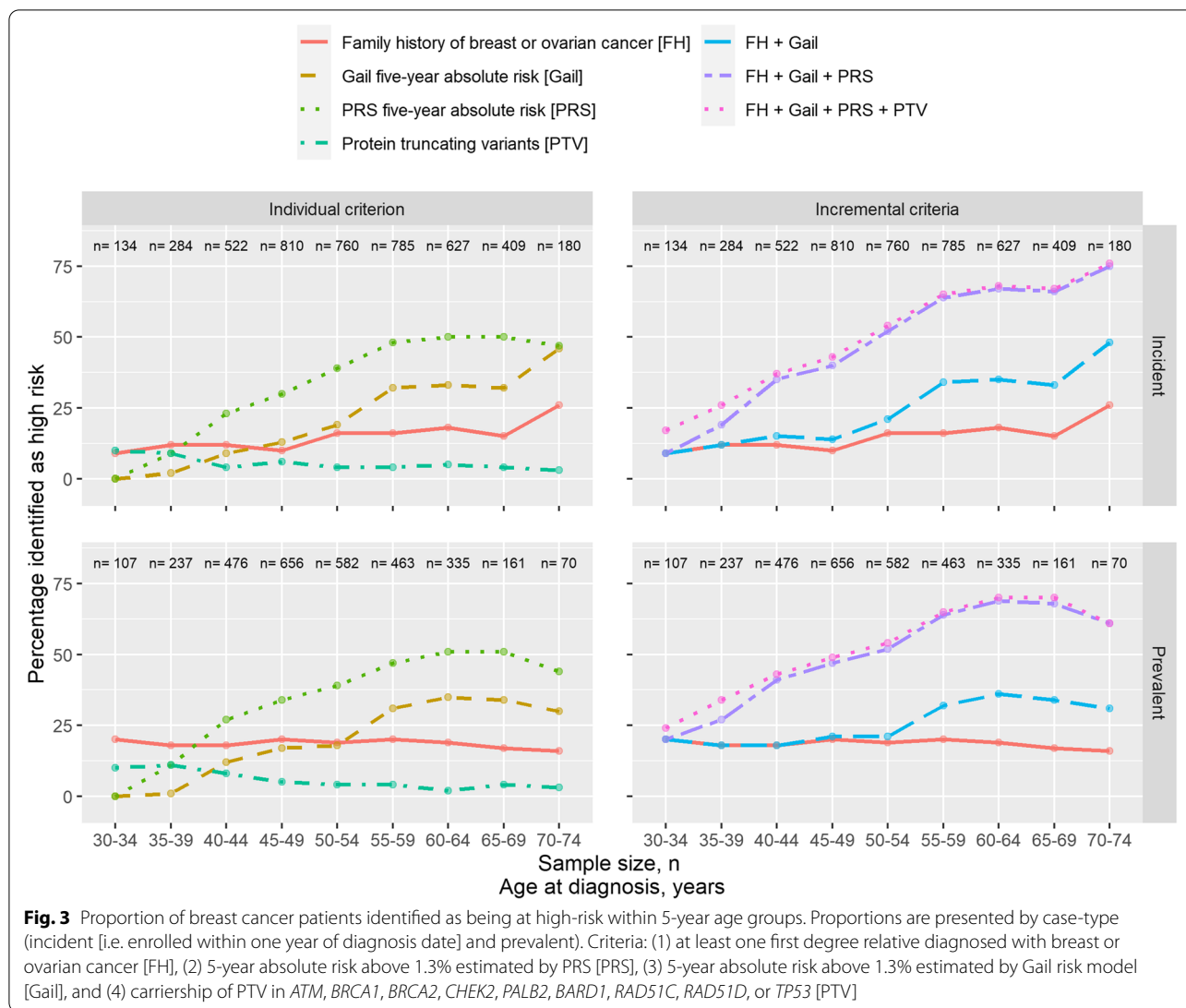


Fig. 3 Proportion of breast cancer patients identified as being at high-risk within 5-year age groups. Proportions are presented by case-type (incident [i.e. enrolled within one year of diagnosis date] and prevalent). Criteria: (1) at least one first degree relative diagnosed with breast or ovarian cancer [FH], (2) 5-year absolute risk above 1.3% estimated by PRS [PRS], (3) 5-year absolute risk above 1.3% estimated by Gail risk model [Gail], and (4) carriership of PTV in *ATM, BRCA1, BRCA2, CHEK2, PALB2, BARD1, RAD51C, RAD51D, or TP53* [PTV]

absolute risk by the Gail model → above 1.3% 5-year absolute risk by PRS → PTV carriership) (Fig. 3). The addition of the PRS criterion produced the largest change in proportion identified as at high risk (Fig. 3). Additional file 2 – Figs. S3 and S4 show the proportion of high-risk patients in each age group by study.

Validation in SCHS, a prospective cohort of healthy individuals

Summary characteristics and allele frequencies corresponding to the 313 variants included in the PRS of the 10,213 women from SCHS are presented in Additional file 2 – Tables S4 and S7, respectively. Four percent (n=418) developed breast cancer over a median follow-up of 20 years (IQR: 18 to 21), of which 19% (n=81) occurred within 5 years of recruitment. Among the 81,

38% (n=31) were above 1.3% 5-year absolute risk by PRS, while only 6% were above 1.3% 5-year absolute risk by the Gail model, suggesting a higher performance of the PRS than the Gail model for breast cancer risk stratification for middle-aged or older Chinese women.

Due to the small number of events within 5 years of recruitment, we studied all breast cancer patients ignoring time to event. The proportion of patients with above 1.3% 5-year absolute risk by PRS (in SCHS’s breast cancer patients) was higher than that observed in SGBCC and MyBrCa’s patients (SCHS = 41%, SGBCC + MyBrCa = 36%) (Additional file 2 – Fig. S5). However, the proportion of patients with above 1.3% 5-year absolute risk by the Gail model was lower (SCHS = 7%, SGBCC + MyBrCa = 21%).

Identifying women (with no personal history of breast cancer, $n=10,213$) at above 1.3% 5-year absolute risk by PRS resulted in a larger proportion of high-risk women ($n=2761$, 27%) as compared to using the Gail model (6%) (Additional file 2 – Fig. S5). However, 6% ($n=172$, of 2761) of these high-risk women identified by PRS developed breast cancer, which was similar to the 5% ($n=28$, of 587) by the Gail model, and 6% ($n=11$, of 180) by family history. Adding PRS to risk stratification identified 165 more breast cancer cases. Figure 4 shows the percentage of women identified as high risk by adding the Gail model and PRS to family history across different age groups. The Gail model identified up to 10% more high-risk individuals, while PRS identified an additional ~27% on top of the Gail model and family history.

Discussion

Currently, in many countries, population-based mammography screening is recommended based on age alone. However, not every woman is at the same level of risk of developing breast cancer. In practice, family history of the disease is widely used as a risk assessment tool. Breast cancer risk of women with a sister or a mother with breast cancer is reported to be approximately twice as high as those who do not have first-degree family members diagnosed with the disease [26]. In addition, family history information of high quality is reported to be highly correlated to the carriership of actionable genomic variants [21]. Prediction models using breast cancer risk factor information collected using questionnaires, such as the Gail model, are also widely used [27]. On the individual level, these risk estimates are encouraged to be included in conversations with clinicians to help make informed decisions about potential interventions, including chemoprevention with tamoxifen [27, 28].

While family history and conventional breast cancer risk factors may change over time and thus require updates and reassessments, an individual's genetic risk based on either established breast cancer predisposition genes or PRS may be determined at birth. However, the implementation of genetic tests in population-wide screening is highly debatable. Pathogenic variants in high-penetrance breast cancer genes are rare; hence, most women in the general population will not benefit and may develop a false sense of security [29]. Previously, the evidence that common genetic variants (used in the calculation of PRS) provide superior risk stratification over conventional breast cancer risk factors is lacking [30, 31]. There was also no consensus on which variants to include in the PRS calculation. However, recent international mega-consortia studies examining over a hundred thousand women show that the tail ends of PRS enable more precise risk differentiation [11–13].

With the latest developments in genetic risk prediction, it is timely to consider whether every woman in the general population should be genetically screened for high-risk genes and the use of PRS in a screening program. Our findings show that both genetic and conventional risk stratification tools have their own merits and are able to identify unique individuals at risk. Each risk assessment tool is a partial predictor at best. The inclusion of multiple predictive tools can pick up additional high-risk individuals who are missed out from using any one tool alone. In our study, family history and genetic risk perform better for women below age 50, as compared to the Gail model. This is noteworthy as the entry age for subsidized breast screening in many countries is 50 years. Genetic risk profiles will help younger women in making informed decisions on whether they should start screening at an earlier age. High-risk individuals may benefit from specific recommendations or interventions based on their personal breast cancer risk profiles.

In countries where breast screening uptake is low, breast cancer risk assessment tools function more than just predictive scores. The knowledge of breast cancer risk on an individual level may serve as a tool to motivate behavioral change. For example, a Finnish study studied the impact of genetic and non-genetic personal risk scores for cardiovascular diseases on health behavior in over 7000 participants. The results show that risk-reducing behavior is observed in participants across all risk strata, although more individuals at high risk made a health behavioral change (42.6% vs 33.5% of individuals not at high risk) [32]. The contributions of genetic and non-genetic risk profile feedback were reported to be independent of each other [32], further supporting the inclusion of both genetic and non-genetic risk factors for stratification in screening programs.

In terms of mammography screening, PRS has been reported to perform well at identifying the women who are most likely to benefit from this mode of detection [33]. The association between PRS and tumor characteristics in our study confirms this observation. Not all tumors grow at the same rate. Despite the advances in technology, routine mammography screening on average fails to detect ~10–30% of all breast cancers [5, 34]. Some of these missed tumors are interval cancers that are diagnosed between two screening episodes [35]. Women at high risk based on PRS will thus benefit from increased screening frequency and compliance to screening.

The main strength of our study is that this is one of the largest and most well-characterized breast cancer cohorts of Asian women. However, many of the breast cancer risk assessment tools and the PRS are established based on European populations and their utility in Asian breast cancer populations remains unclear. Approximately

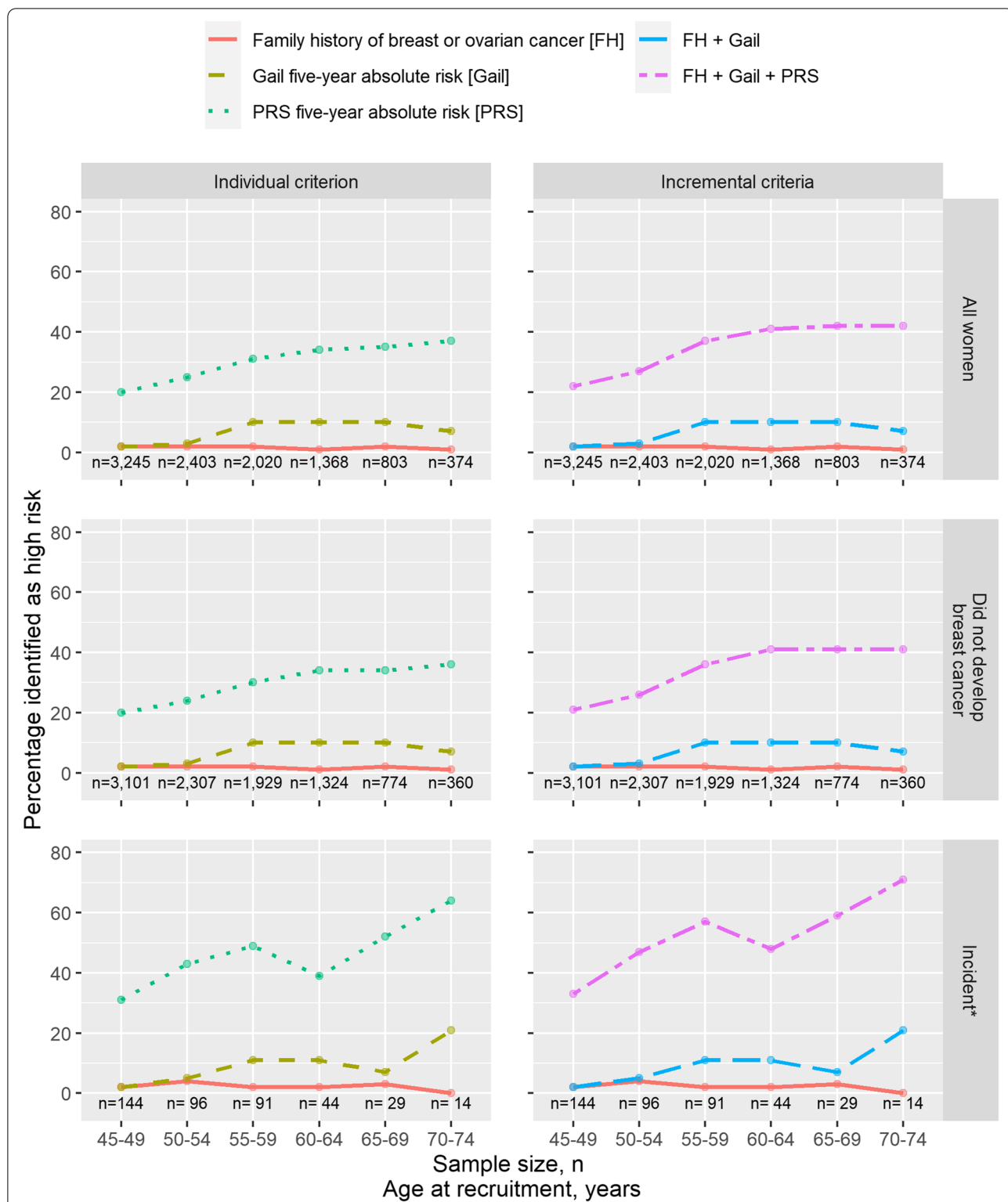


Fig. 4 Proportion of breast cancer patients from the validation dataset identified as being at high risk. The prospective cohort of healthy women—the Singapore Chinese Health Study [SCHS]—was used for validation. Criteria for high-risk: (1) at least one first degree relative diagnosed with breast or ovarian cancer [FH], (2) 5-year absolute risk above 1.3% estimated by the polygenic risk score [PRS], and (3) 5-year absolute risk above 1.3% estimated by Gail risk model [Gail]. PRS is standardized with mean and standard deviation of Chinese controls from the Singapore and Malaysia dataset. *Note: This plot uses age at recruitment; breast cancer may not occur within 5 years of recruitment

half of the breast cancer population were identified as high risk, suggesting that other factors not considered in the risk prediction models (e.g., mammographic density, physical activity, alcohol, smoking) studied may be responsible. We classified breast cancer patients into risk categories based on 5-year absolute risks at the age of breast cancer diagnosis; this may not be representative of women without breast cancer. While we are likely to overestimate the 5-year absolute risk, results from the prospective cohort (SCHS) support the use of genetic factors on top of family history and the Gail model. As this is a case-only cohort, the proportions from the risk classification analysis are not representative of the general population, where most women will be classified as low risk. Nonetheless, this will not affect the comparison of how different criteria identify women at high risk.

While this study's main focus was to highlight the lack of an overlap between high-risk women identified by genetic and non-genetic risk factors, it is worth noting that other works in the field have studied the potential improvements in risk prediction by combining different risk factors. For instance, Choudhury et al. explored the value of adding mammographic density and PRS to classical risk factors in a population of women of European ancestry [36]. In a more recent study, Yang et al. assessed the performance of breast cancer risk prediction models incorporating genetic and non-genetic risk factors in 20,444 breast cancer cases and 106,450 controls from the Asia Breast Cancer Consortium [37]. These developments are complementary to the findings of this study and will help pave the way for more patient-centric, data-driven healthcare systems in the future.

Conclusions

In summary, our assessment shows that family history and genetic and non-genetic risk stratification tools have the potential to complement one another to identify women at high risk in breast screening programs. The results add to the growing body of evidence to support a paradigm shift from an approach that is age-based to risk-based.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12916-022-02334-z>.

Additional file 1. Details on the study populations, DNA isolation and genotyping, polygenic risk score (PRS) calculation, and methods to obtain absolute risk by PRS and the Gail model.

Additional file 2: Table S1. Targeted regions of gene panel (hg19). **Table S2.** List of predicted protein-truncating variants. **Table S3.** Carriers of protein truncating variants. **Table S4.** Characteristics of cohorts. **Tables S5 and S6.** Agreement between the five criteria in identifying breast

cancer patients at high risk of breast cancer. **Table S7.** List of 313 variants in the polygenic risk score. **Figure S1.** Comparing the five-year absolute risk prediction. **Figures S2 and S5.** Venn diagram of breast cancer patients at high risk of breast cancer. **Figures S3 and S4.** Proportion of breast cancer patients identified as being at high risk.

Additional file 3. The incidence and mortality rates of Singapore.

Additional file 4: Table SM1 and SM2. List of 313 variants in the polygenic risk score. **Table SM3.** Mean and standard deviation of polygenic risk scores. **Table SM4.** Relative risk of the percentile corresponding to Figure SM4. **Figure SM1-SM3.** Flowcharts for individuals selected. **Figure SM4 and SM5.** Five-year absolute risk of developing breast cancer.

Acknowledgements

The study was made possible with the corporation of the participants and the assistance of research team members; from SGBCC – Jenny Liu, Siew Li Tan, Siok Hoon Yeo, Ting Ting Koh, Amanda Ong, Michelle Jia Qi Mo, Ying Jia Chew, Jin Yee Lee, Jing Jing Hong, Hui Min Lau, Ganga Devi D/O Chandrasegaran, and Nur Khaliesah Binte Mohamed Riza. We thank the Singapore Cancer Registry for assistance in the identification of cancer cases in the SCHS cohort. MyBrCa and MyMammo 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 Ernie Azwa Yusop, Hanani Che Halim, Faizah Harun, Farhana Fadzli, Leelavathy Krishnan, Shivaani Mariapun, Maheswari Jaganathan, Meow Keong Thong, Daphne S. C Lee, Sheau-Yee Lee, Sze-Yee Phuah, Kah-Nyin Lai, Shao-Yan Lau, Pui-Yoke Kwan, Pei-Sze Ng, Sook-Yee Yoon, Siti Norhidayu Hasan, Siu-Wan Wong, and Heamanthaa Padmanabhan) for their contributions and commitment to this study.

Authors' contributions

Study design: JL and PJH; writing group: JL, PJH, MH, SHT, and WKH; data management: PJH, MCT, and YSY; statistical analysis: PJH, JL, and WKH; provision and processing of DNA samples and phenotypic data: MH, VKMT, BKTT, SMT, EYT, GHL, CHY, NAMT, KR, TI, TH, W-PK, J-MY, AJK, CCK, and XS; breast cancer expertise: MH, VKMT, BKTT, SMT, EYT, GHL, FYW, EHL, JN, WYC, LCHL, WSY, CMS, SWT, CWQN, ZY, JAL, KR, TI, and TH; statistics and epidemiology expertise: MH, SHT, J-MY, MCT, AKB, and ACA. The authors read and approved the final version of the paper.

Funding

SGBCC was supported by the National Research Foundation Singapore [NRF-NRFF2017-02, awarded to J Li], BMRC Central Research Fund [Applied Translational Research, awarded to J Li], PRECISE CIP [awarded to MH and J Li], NUS start-up Grant [awarded to MH], National University Cancer Institute Singapore (NCIS) Centre Grant [NMRC/CG/NCIS/2010, NMRC/CG/012/2013 and CGAug16M005, awarded to MH], Breast Cancer Prevention Programme [BCPP, awarded to MH], Breast Cancer Screening Prevention Programme [BCSPP, awarded to MH], Asian Breast Cancer Research Fund [awarded to MH], and the NMRC Clinician Scientist Award (SI Category) [NMRC/CSA-SI/0015/2017, awarded to MH], NMRC Centre Grant [CGAug16M012, awarded to EYT]. The Malaysian Breast Cancer Genetic Study was established using funds from the Malaysian Ministry of Science, and the Malaysian Ministry of Higher Education High Impact Research Grant [grant no: UM.C/HIR/MOHE/06]. The Malaysian Mammographic Density Study was established using funds raised through the Sime Darby LPGA tournament and the High Impact Research Grant. Additional funding was received from Newton-Ungku Omar Fund [grant no: MR/P012930/1] and Wellcome Trust [grant no: v203477/Z/16/Z], Yayasan Sime Darby, PETRONAS, Estee Lauder Group of Companies, and other donors of Cancer Research Malaysia. The BRIDGES panel sequencing was supported by the European Union Horizon 2020 research and innovation program BRIDGES (grant number, 634935) and the Wellcome Trust (v203477/Z/16/Z). Genotyping of the OncoArray was funded by the NIH Grant U19 CA148065, and Cancer UK Grant (1287/A16563) and the PERSPECTIVE project supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research (grant GPH-129344) and, the Ministère de l'Économie, Science et Innovation du Québec through Genome Québec and the PSRSIIRI-701 grant, and the Quebec Breast Cancer Foundation.

The Singapore Chinese Health Study was supported by the National Medical Research Council, Singapore (NMRC/CIRG/1456/2016) and National Institutes of Health, USA (NCI R01 CA55069, R35 CA53890, R01 CA80205, and R01 CA144034).

Availability of data and materials

The data that support the findings of this study are available on request from Mikael Hartman (epfbamh@nus.edu.sg). The data are not publicly available due to privacy or ethical restrictions.

Declarations

Ethics approval and consent to participate

All studies were performed in accordance with the Declaration of Helsinki and all participants provided written informed consent. This study was approved by the institutional review board at the Agency for Science, Technology and Research (A-STAR) (ref: 2020-005). SGBCC was approved by the National Healthcare Group Domain Specific Review Board (reference number: 2017/00797) and the SingHealth Centralised Institutional Review Board (reference number: 2019/2246). MyBrCa was approved by the Independent Ethics Committee, Ramsay Sime Darby Health Care (reference number: 201109.4 and 201208.1), and the Medical Ethics Committee, University Malaya Medical Centre (reference number: 842.9).

Consent for publication

Not applicable.

Competing interests

J.N receives research funding from AstraZeneca. The other authors declare that they have no potential competing interests.

Author details

¹Genome Institute of Singapore, Human Genetics, Singapore, Singapore. ²Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore. ³Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore. ⁴Cancer Research Malaysia, 1 Jalan SS12/1A, 47500 Subang Jaya, Selangor, Malaysia. ⁵School of Mathematical Sciences, Faculty of Science and Engineering, University of Nottingham Malaysia, Jalan Broga, 43500 Semenyih, Selangor, Malaysia. ⁶Department of General Surgery, Sengkang General Hospital, Singapore, Singapore. ⁷Department of Breast Surgery, Singapore General Hospital, Singapore, Singapore. ⁸Division of Surgery and Surgical Oncology, National Cancer Centre Singapore, Singapore, Singapore. ⁹Department of General Surgery, Tan Tock Seng Hospital, Singapore 308433, Singapore. ¹⁰Lee Kong Chian School of Medicine, Nanyang Technology University, Singapore, Singapore. ¹¹Institute of Molecular and Cell Biology, Singapore, Singapore. ¹²KK Breast Department, KK Women's and Children's Hospital, Singapore 229899, Singapore. ¹³Division of Breast Surgery, Changi General Hospital, Singapore, Singapore. ¹⁴Subang Jaya Medical Centre, Subang Jaya, Selangor, Malaysia. ¹⁵Department of Surgery, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia. ¹⁶Universiti Malaya Cancer Research Institute, Kuala Lumpur, Malaysia. ¹⁷Division of Radiation Oncology, National Cancer Centre Singapore, Singapore, Singapore. ¹⁸Division of Medical Oncology, National Cancer Centre Singapore, Singapore, Singapore. ¹⁹Cancer Genetics Service, National Cancer Centre Singapore, Singapore, Singapore. ²⁰Department of Diagnostic Radiology, Singapore General Hospital, Singapore, Singapore. ²¹Department of Surgery, University Surgical Cluster, National University Hospital, Singapore, Singapore. ²²Healthy Longevity Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. ²³UPMC Hillman Cancer Center, Pittsburgh, PA, USA. ²⁴Department of Epidemiology, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA, USA. ²⁵Singapore Institute for Clinical Sciences, Agency for Science Technology and Research (A*STAR), Singapore 117609, Singapore. ²⁶Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK. ²⁷Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. ²⁸Department of Surgery, Faculty of Medicine, University of Malaya, Jalan Universiti, 50630 Kuala Lumpur, Malaysia.

Received: 15 December 2021 Accepted: 14 March 2022
Published online: 26 April 2022

References

- Wells J. Mammography and the politics of randomised controlled trials. *BMJ*. 1998;317(7167):1224–9.
- Ebell MH, Thai TN, Royalty KJ. Cancer screening recommendations: an international comparison of high income countries. *Public Health Rev*. 2018;39:7.
- Ernst VL. Mammography screening for women aged 40 through 49—a guidelines saga and a clarion call for informed decision making. *Am J Public Health*. 1997;87(7):1103–6.
- Lee CI, Chen LE, Elmore JG. Risk-based Breast cancer screening: implications of breast density. *Med Clin North Am*. 2017;101(4):725–41.
- Ray KM, Price ER, Joe BN. Breast density legislation: mandatory disclosure to patients, alternative screening, billing, reimbursement. *AJR Am J Roentgenol*. 2015;204(2):257–60.
- Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat Med*. 2004;23(7):1111–30.
- Gail MH. Twenty-five years of breast cancer risk models and their applications. *J Natl Cancer Inst*. 2015;107(5):djv042.
- Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med*. 2000;343(2):78–85.
- Moller S, Mucci LA, Harris JR, Scheike T, Holst K, Halekoh U, et al. The Heritability of breast cancer among women in the Nordic Twin Study of Cancer. *Cancer Epidemiol Biomarkers Prev*. 2016;25(1):145–50.
- Breast Cancer Association C, Dorling L, Carvalho S, Allen J, Gonzalez-Neira A, Luccarini C, et al. Breast cancer risk genes - association analysis in more than 113,000 women. *N Engl J Med*. 2021;384(5):428–39.
- Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am J Hum Genet*. 2019;104(1):21–34.
- Ho WK, Tan MM, Mavaddat N, Tai MC, Mariapun S, Li J, et al. European polygenic risk score for prediction of breast cancer shows similar performance in Asian women. *Nat Commun*. 2020;11(1):3833.
- Mavaddat N, Pharoah PD, Michailidou K, Tyrer J, Brook MN, Bolla MK, Wang Q, et al. Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst*. 2015;107(5):djv036. <https://doi.org/10.1093/jnci/djv036>.
- Ho PJ, Yeoh YS, Miao H, Lim SH, Tan EY, Tan BKT, et al. Cohort profile: The Singapore Breast Cancer Cohort (SGBCC), a multi-center breast cancer cohort for evaluation of phenotypic risk factors and genetic markers. *PLoS One*. 2021;16(4):e0250102.
- Tan MM, Ho WK, Yoon SY, Mariapun S, Hasan SN, Lee DS, et al. A case-control study of breast cancer risk factors in 7,663 women in Malaysia. *PLoS One*. 2018;13(9):e0203469.
- Tan KHX, Tan LWL, Sim X, Tai ES, Lee JJ, Chia KS, et al. Cohort profile: The Singapore Multi-Ethnic Cohort (MEC) study. *Int J Epidemiol*. 2018;47(3):699–699j.
- Dorajoo R, Chang X, Gurung RL, Li Z, Wang L, Wang R, et al. Loci for human leukocyte telomere length in the Singaporean Chinese population and trans-ethnic genetic studies. *Nat Commun*. 2019;10(1):2491.
- Michailidou K, Lindstrom S, Dennis J, Beesley J, Hui S, Kar S, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature*. 2017;551(7678):92–4.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–75.
- National Registry of Diseases Office N: Singapore Cancer Registry Annual Registry Report 2015. 2017.
- Wen W, Shu XO, Guo X, Cai Q, Long J, Bolla MK, 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.
- Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst*. 1989;81(24):1879–86.

23. Age-Specific Death Rates, Annual. <https://www.tablebuilder.singstat.gov.sg/publicfacing/viewMultiTable.action>. Accessed 8 Mar 2022.
24. Shieh Y, Eklund M, Madlensky L, Sawyer SD, Thompson CK, Stover Fiscalini A, et al. Breast cancer screening in the precision medicine era: risk-based screening in a population-based trial. *J Natl Cancer Inst*. 2017;109(5). <https://doi.org/10.1093/jnci/djw290>.
25. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20(1):37–46.
26. Colditz GA, Willett WC, Hunter DJ, Stampfer MJ, Manson JE, Hennekens CH, et al. Family history, age, and risk of breast cancer. Prospective data from the Nurses' Health Study. *JAMA*. 1993;270(3):338–43.
27. Rockhill B, Spiegelman D, Byrne C, Hunter DJ, Colditz GA. Validation of the Gail et al. model of breast cancer risk prediction and implications for chemoprevention. *J Natl Cancer Inst*. 2001;93(5):358–66.
28. Chay WY, Ong WS, Tan PH, Jie Leo NQ, Ho GH, Wong CS, et al. Validation of the Gail model for predicting individual breast cancer risk in a prospective nationwide study of 28,104 Singapore women. *Breast Cancer Res*. 2012;14(1):R19.
29. Wentzensen N, Berg CD. Population testing for high penetrance genes: are we there yet? *J Natl Cancer Inst*. 2018;110(7):687–9.
30. Narod SA. Personalised medicine and population health: breast and ovarian cancer. *Hum Genet*. 2018;137(10):769–78.
31. Chowdhury S, Dent T, Pashayan N, Hall A, Lyratzopoulos G, Hallowell N, et al. Incorporating genomics into breast and prostate cancer screening: assessing the implications. *Genet Med*. 2013;15(6):423–32.
32. Widén E, Junna N, Ruotsalainen S, Surakka I, Mars N, Ripatti P, et al. Communicating polygenic and non-genetic risk for atherosclerotic cardiovascular disease - An observational follow-up study. medRxiv. 2020.09.18.20197137. <https://doi.org/10.1101/2020.09.18.20197137>.
33. Li J, Ugalde-Morales E, Wen WX, Decker B, Eriksson M, Torstensson A, et al. Differential burden of rare and common variants on tumor characteristics, survival, and mode of detection in breast cancer. *Cancer Res*. 2018;78(21):6329–38.
34. Kamal RM, Abdel Razek NM, Hassan MA, Shaalan MA. Missed breast carcinoma; why and how to avoid? *J Egypt Natl Canc Inst*. 2007;19(3):178–94.
35. Bulliard JL, Sasieni P, Klabunde C, De Landtsheer JP, Yankaskas BC, Fracheboud J. Methodological issues in international comparison of interval breast cancers. *Int J Cancer*. 2006;119(5):1158–63.
36. Pal Choudhury P, Wilcox AN, Brook MN, Zhang Y, Ahearn T, Orr N, et al. Comparative Validation of breast cancer risk prediction models and projections for future risk stratification. *J Natl Cancer Inst*. 2020;112(3):278–85.
37. Yang Y, Tao R, Shu X, Cai Q, Wen W, Gu K, Gao YT, Zheng Y, Kweon SS, Shin MH, Choi JY, Lee ES, Kong SY, Park B, Park MH, Jia G, Li B, Kang D, Shu XO, Long J, Zheng W. Incorporating polygenic risk scores and nongenetic risk factors for breast cancer risk prediction among Asian women. *JAMA Netw Open*. 2022;5(3):e2149030.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

