

## Article

# Combination of a 15-SNP Polygenic Risk Score and Classical Risk Factors for the Prediction of Breast Cancer Risk in Cypriot Women

Kristia Yiangou <sup>1,2</sup>, Kyriacos Kyriacou <sup>1,2</sup>, Eleni Kakouri <sup>3</sup>, Yiola Marcou <sup>3</sup>, Mihalis I. Panayiotidis <sup>1,2</sup> , Maria A. Loizidou <sup>1,2,†</sup>, Andreas Hadjisavvas <sup>1,2,\*,†</sup>  and Kyriaki Michailidou <sup>2,4,\*,†</sup> 

<sup>1</sup> Department of Cancer Genetics, Therapeutics and Ultrastructural Pathology, The Cyprus Institute of Neurology and Genetics, Nicosia 2371, Cyprus; kristiay@cing.ac.cy (K.Y.); kyriacos@cing.ac.cy (K.K.); mihalisp@cing.ac.cy (M.I.P.); loizidou@cing.ac.cy (M.A.L.)

<sup>2</sup> The Cyprus School of Molecular Medicine, The Cyprus Institute of Neurology and Genetics, Nicosia 2371, Cyprus

<sup>3</sup> Department of Medical Oncology, Bank of Cyprus Oncology Center, Nicosia 2012, Cyprus; eleni.kakouri@bococ.org.cy (E.K.); yiola.marcou@bococ.org.cy (Y.M.)

<sup>4</sup> Biostatistics Unit, The Cyprus Institute of Neurology and Genetics, Nicosia 2371, Cyprus

\* Correspondence: ahsavvas@cing.ac.cy (A.H.); kyriakimi@cing.ac.cy (K.M.)

† These authors contributed equally to this study.



**Citation:** Yiangou, K.; Kyriacou, K.; Kakouri, E.; Marcou, Y.; Panayiotidis, M.I.; Loizidou, M.A.; Hadjisavvas, A.; Michailidou, K. Combination of a 15-SNP Polygenic Risk Score and Classical Risk Factors for the Prediction of Breast Cancer Risk in Cypriot Women. *Cancers* **2021**, *13*, 4568. <https://doi.org/10.3390/cancers13184568>

Academic Editor: Ettore Bidoli

Received: 2 August 2021

Accepted: 8 September 2021

Published: 11 September 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Simple Summary:** Breast cancer is the most commonly diagnosed type of cancer in women worldwide. Stratification of women based on their individual breast cancer risk could guide targeted preventative strategies and population screening. Integrated models that combine the effects of a polygenic risk score (PRS) with classical breast cancer risk factors could provide an individualized breast-cancer risk estimation. Although various studies have extensively evaluated the performance of such integrated models in populations of European ancestry, no previous studies have included individuals of Greek-Cypriot origin. To this end, we have assessed the predictive performance of a 15-SNP PRS (PRS<sub>15</sub>), in combination with classical breast-cancer risk factors, in women of Greek-Cypriot origin. This proof-of-concept study suggests that models combining genetic data with classical risk factors may be used in the future for the prediction of breast-cancer risk and, therefore, supports their potential clinical utility for targeted preventative strategies in Cypriot women.

**Abstract:** The PRS combines multiplicatively the effects of common low-risk single nucleotide polymorphisms (SNPs) and has the potential to be used for the estimation of an individual's risk for a trait or disease. PRS has been successfully implemented for the prediction of breast cancer risk. The combination of PRS with classical breast cancer risk factors provides a more comprehensive risk estimation and could, thus, improve risk stratification and personalized preventative strategies. In this study, we assessed the predictive performance of the combined effect of PRS<sub>15</sub> with classical breast-cancer risk factors in Cypriot women using 1109 cases and 1177 controls from the MASTOS study. The PRS<sub>15</sub> was significantly associated with an increased breast cancer risk in Cypriot women OR (95% CI) 1.66 (1.25–2.19). The integrated risk model obtained an AUC (95% CI) 0.70 (0.67–0.72) and had the ability to stratify women according to their disease status at the extreme deciles. These results provide evidence that the combination of PRS with classical risk factors may be used in the future for the stratification of Cypriot women based on their disease risk, and support its potential clinical utility for targeted preventative actions and population screening.

**Keywords:** breast cancer; polygenic risk score; classical risk factors; risk prediction; Cypriot women

## 1. Introduction

Breast cancer is the most commonly diagnosed type of cancer in women around the world [1]. Disease susceptibility varies between individuals, and is influenced by multiple

genetic and non-genetic risk factors such as age, height, BMI, reproductive and menstrual history, use of hormone replacement therapy and lifestyle risk factors [2–6], all of which confer individually a moderate effect on breast cancer risk. Family history is one of the most established risk factors associated with breast cancer predisposition. So far, multiple breast cancer susceptibility variants have been discovered. These include pathogenic variants in high-risk and moderate-risk genes, which are rare in the population, and account for about 20–25% of the excess familial relative risk (FRR) of breast cancer. Large-scale genome-wide association studies (GWAS) identified more than 200 SNPs common in the population, each individually conferring a small effect on the disease risk, but collectively account for about ~18.3% of the excess FRR of the disease [7–10].

Individualized risk estimation could be used for the stratification of women into different categories according to their breast cancer risk, which could potentially guide targeted risk management strategies, and improve population screening efficiency [11,12]. A polygenic risk score (PRS) combines multiplicatively the effects of common susceptibility variants and could be used for the stratification of women according to their personal breast cancer risk [13–17]. Recently, Mavaddat et al. 2019 have constructed a PRS including 313 SNPs for the prediction of overall, ER-positive and ER-negative, breast cancer risk in women of European ancestry [14]. Compared with females in the 40–60% quintile of the PRS<sub>313</sub> risk distribution, females in the 1st and 99th percentiles had 0.27-times and 4-times increased overall breast cancer risk, respectively [14]. Combination of PRS with classical risk factors can substantially improve the prediction of breast cancer risk and could detect individuals at different levels of the disease risk [18–24]. Furthermore, the incorporation of PRS into breast cancer risk prediction models such as BOADICEA, Tyrer–Cuzick, Gail and Rosner–Colditz can improve their discrimination power [25–29], and provide a more comprehensive individualized risk estimation [30].

Although large studies have assessed the performance of such combined models in populations of European descent, no previous study included individuals of Greek-Cypriot origin. In Cyprus, an island in the Mediterranean region, more than 500 new breast cancer cases are diagnosed in females annually, accounting for around 35% of all female cancer cases [31]. Currently, little is known about how common variants influence breast cancer susceptibility in Cypriot women or about the utility of their combined effect (PRS) for the prediction of breast cancer risk. A previous study, by our group, has evaluated 11 SNPs identified through the initial GWAS for association with breast cancer risk in Cypriot women and concluded that four of them were nominally significantly associated with breast cancer risk [32]. Furthermore, a previous study evaluated the strength of associations between classical risk factors and breast cancer risk in Cypriot women [33].

The aim of this study was to combine a PRS<sub>15</sub> with classical breast cancer risk factors and assess its predictive power in Cypriot women using samples from the MASTOS study [33].

## 2. Materials and Methods

### 2.1. Study Population

Study participants included 2286 females derived from the MASTOS study [33]. MASTOS is a population-based case-control study that includes 1109 female breast cancer cases with mean age (SD) at interview 55.99 (9.15), and 1177 female healthy controls with mean age (SD) 55.65 (7.04) of self-reported Greek-Cypriot origin. Cases were females who were diagnosed with breast cancer between January 1999 and December 2006. Healthy controls were participating in the National program for breast cancer screening, with negative mammography results, and no previous personal history of breast cancer. Demographic and phenotypic data of all the participants were collected using a specially designed questionnaire, through a standardized interview. All study participants were recruited from January 2004 to December 2006. Detailed information on the purpose, design of the study, data collection and study population is described elsewhere [33]. Three samples were excluded from the analysis due to the high missing rate of phenotypic data. Therefore, the

total number of individuals included in the analysis was 2283, consisting of 1174 controls and 1109 cases.

The Cyprus National Bioethics Committee approved the study protocol (EEBK/EII/2005/08), and all study subjects gave written consent. The study was conducted in compliance with the Helsinki Declaration.

## 2.2. SNP Selection and Genotyping

Fifteen SNPs that were previously identified via GWAS (prior to 2013) [34–40] were selected and genotyped in all MASTOS study participants, using the Taqman SNP genotyping assays from Applied Biosystems Inc. (ABI), according to the manufacturer’s instructions and as described in detail elsewhere [32]. Detailed information about the 15 SNPs included in this study is summarized in Table 1.

**Table 1.** Information about the 15 SNPs selected and genotyped in all MASTOS study participants.

| CHR <sup>1</sup> | SNP               | Position <sup>2</sup> | Alleles <sup>3</sup> | MAF <sup>4</sup> | iCOGS OR <sup>5</sup>             | iCOGS p-Value <sup>6</sup>               | MASTOS MAF <sup>7</sup> | MASTOS OR <sup>8</sup>            | MASTOS p-Value |
|------------------|-------------------|-----------------------|----------------------|------------------|-----------------------------------|--|-------------------------|-----------------------------------|----------------|
| 1                | rs11249433        | 121280613             | A/G                  | 0.4              | 1.09<br>(1.07–1.12)               | $4.43 \times 10^{-20}$                   | 0.46                    | 1.00<br>(0.89–1.12)               | 0.98           |
| 2                | <b>rs13387042</b> | <b>217905832</b>      | <b>A/G</b>           | <b>0.49</b>      | <b>0.88</b><br><b>(0.86–0.9)</b>  | <b><math>8.91 \times 10^{-41}</math></b> | <b>0.45</b>             | <b>0.85</b><br><b>(0.75–0.95)</b> | <b>0.005</b>   |
| 3                | rs4973768         | 27416013              | C/T                  | 0.47             | 1.1<br>(1.08–1.12)                | $4.65 \times 10^{-22}$                   | 0.45                    | 0.89<br>(0.78–1.00)               | 0.055          |
| 5                | <b>rs889312</b>   | <b>56031884</b>       | <b>A/C</b>           | <b>0.28</b>      | <b>1.12</b><br><b>(1.1–1.15)</b>  | <b><math>2.87 \times 10^{-27}</math></b> | <b>0.29</b>             | <b>1.18</b><br><b>(1.04–1.34)</b> | <b>0.01</b>    |
| 6                | <b>rs2046210</b>  | <b>151948366</b>      | <b>G/A</b>           | <b>0.34</b>      | <b>1.08</b><br><b>(1.06–1.1)</b>  | <b><math>2.13 \times 10^{-14}</math></b> | <b>0.41</b>             | <b>1.13</b><br><b>(1.00–1.27)</b> | <b>0.047</b>   |
| 8                | <b>rs13281615</b> | <b>128355618</b>      | <b>A/G</b>           | <b>0.4</b>       | <b>1.1</b><br><b>(1.08–1.12)</b>  | <b><math>3.26 \times 10^{-22}</math></b> | <b>0.48</b>             | <b>1.07</b><br><b>(0.95–1.20)</b> | <b>0.26</b>    |
| 9                | <b>rs1011970</b>  | <b>22062134</b>       | <b>G/T</b>           | <b>0.17</b>      | <b>1.06</b><br><b>(1.03–1.08)</b> | <b><math>2.68 \times 10^{-5}</math></b>  | <b>0.19</b>             | <b>1.15</b><br><b>(0.99–1.33)</b> | <b>0.07</b>    |
| 10               | <b>rs2981582</b>  | <b>123352317</b>      | <b>G/A</b>           | <b>0.38</b>      | <b>1.26</b><br><b>(1.24–1.28)</b> | <b><math>1.6 \times 10^{-120}</math></b> | <b>0.44</b>             | <b>1.16</b><br><b>(1.03–1.31)</b> | <b>0.01</b>    |
| 10               | <b>rs10995190</b> | <b>64278682</b>       | <b>G/A</b>           | <b>0.16</b>      | <b>0.86</b><br><b>(0.83–0.88)</b> | <b><math>1.6 \times 10^{-29}</math></b>  | <b>0.14</b>             | <b>0.97</b><br><b>(0.82–1.15)</b> | <b>0.7</b>     |
| 10               | <b>rs704010</b>   | <b>80841148</b>       | <b>C/T</b>           | <b>0.38</b>      | <b>1.08</b><br><b>(1.06–1.1)</b>  | <b><math>2.94 \times 10^{-15}</math></b> | <b>0.37</b>             | <b>1.01</b><br><b>(0.90–1.14)</b> | <b>0.83</b>    |
| 11               | rs3817198         | 1909006               | T/C                  | 0.31             | 1.07<br>(1.05–1.09)               | $1.09 \times 10^{-10}$                   | 0.31                    | 0.97<br>(0.85–1.09)               | 0.59           |
| 11               | <b>rs614367</b>   | <b>69328764</b>       | <b>C/T</b>           | <b>0.15</b>      | <b>1.21</b><br><b>(1.18–1.24)</b> | <b><math>1.5 \times 10^{-45}</math></b>  | <b>0.11</b>             | <b>1.09</b><br><b>(0.91–1.31)</b> | <b>0.36</b>    |
| 16               | <b>rs3803662</b>  | <b>52586341</b>       | <b>G/A</b>           | <b>0.26</b>      | <b>1.24</b><br><b>(1.21–1.26)</b> | <b><math>2.71 \times 10^{-86}</math></b> | <b>0.33</b>             | <b>1.01</b><br><b>(0.89–1.14)</b> | <b>0.86</b>    |
| 17               | <b>rs6504950</b>  | <b>53056471</b>       | <b>G/A</b>           | <b>0.28</b>      | <b>0.94</b><br><b>(0.92–0.96)</b> | <b><math>8.15 \times 10^{-9}</math></b>  | <b>0.26</b>             | <b>0.94</b><br><b>(0.82–1.07)</b> | <b>0.34</b>    |
| 21               | rs2823093         | 16520832              | G/A                  | 0.27             | 0.93<br>(0.91–0.95)               | $2.39 \times 10^{-12}$                   | 0.73                    | 1.07<br>(0.94–1.23)               | 0.28           |

<sup>1</sup> Chromosome. <sup>2</sup> Build 37 position. <sup>3</sup> Reference/Effect allele. <sup>4</sup> Mean frequency of the effect allele in the controls taken from the iCOGS study [7,8]. <sup>5</sup> Per allele odds ratio (95% Confidence Intervals) for the effect allele taken from the iCOGS study (Associations for overall breast cancer). <sup>6</sup> p-value taken from the iCOGS study. <sup>7</sup> Frequency of the effect allele in controls in the MASTOS study. <sup>8</sup> Per allele odds ratio (95% Confidence Intervals) for the effect allele in the MASTOS study. Eleven SNPs in the same direction as previously reported in the iCOGS study are shown in bold.

## 2.3. Statistical Analysis

Allele frequencies, odds ratios (ORs) and 95% confidence intervals (CIs) of the 15 SNPs were calculated in the MASTOS study using logistic regression analysis in R (version 3.6.3) [41]. A 15-SNP PRS model (PRS<sub>15</sub>) was subsequently constructed, and a score was created for each woman using the following Equation (1), as previously described in Pharoah et al. 2002 [42]:

$$\text{PRS} = \beta_1 x_1 + \dots + \beta_k x_k + \beta_{15} x_{15} \quad (1)$$

Briefly, in this Equation (1),  $\beta_k$  is the log OR of the minor allele for SNP<sub>k</sub> obtained from the iCOGS study [7,8], and  $x_k$  is the number of minor allele copies that are carried by each individual for SNP<sub>k</sub> and can take values 0, 1 or 2 (minor allele was defined based on

the published minor allele frequency (MAF)). Information about the published ORs and allele frequencies of the 15 SNPs included in the PRS is summarized in Table 1.

Logistic regression analysis was performed to evaluate for associations between PRS<sub>15</sub> and breast cancer risk, and by quartiles of the PRS<sub>15</sub> risk distribution, standardized by the controls, and using the 2nd quartile, 25–50%, as the reference. All calculations were carried out in R (version 3.6.3).

Associations between each risk factor and breast cancer risk were calculated using univariable logistic regression analysis (Table S1). Pair-wise Spearman correlations were calculated in the control group to assess for interactions between the PRS<sub>15</sub> and 10 classical breast cancer risk factors including: menopausal status (yes/no), age at menarche (years), parity (yes/no), age at first full-term pregnancy (FFTP) (per 5 years), breastfeeding among parous women (yes/no), height (cm), BMI (kg/m<sup>2</sup>), use of hormone-replacement therapy (HRT) (yes/no), smoking status (yes/no) and family history (in a first degree relative) (yes/no).

Subsequently, a multivariable model consisting of the PRS<sub>15</sub> and the risk factors was constructed. Multivariable logistic regression analysis was performed to determine the association between the integrated risk model and breast cancer risk. Then, MASTOS dataset was divided into deciles according to the predicted risk probability of the integrated risk model, to evaluate its ability to stratify women based on their disease status. Logistic regression analysis was performed to generate OR (95% CI) of each decile, by using the 5th decile as the reference. For the final integrated risk model, only individuals with complete observations were used for the analysis. Thus, the total number of individuals included in the final analysis was 1780, consisting of 900 controls and 880 cases. Sensitivity and specificity analysis were performed to evaluate the performance of the model.

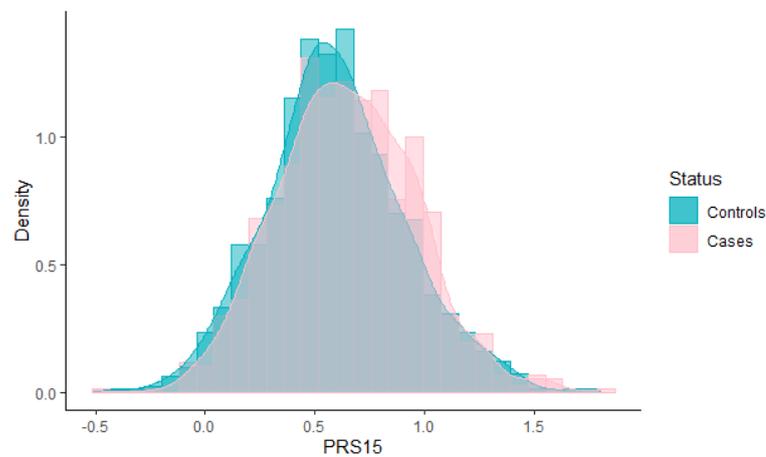
The global goodness-of-fit of each model was evaluated using the Hosmer–Lemeshow test. The area under the receiver operating characteristic curve (AUC) and 95% DeLong CI were calculated in order to evaluate the discrimination power of the models, using the pROC package in R [43]. All tests were two-sided, using a *p*-value threshold of 0.05, and were carried out in R (version 3.6.3).

### 3. Results

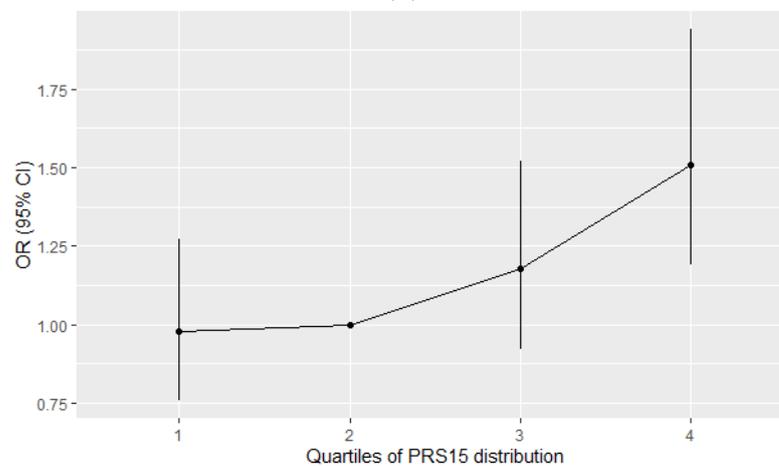
#### 3.1. Evaluation of the PRS<sub>15</sub> and Its Association with Breast Cancer Risk in Greek-Cypriot Women

Single SNP analysis showed that 4 of the 15 SNPs were associated with breast cancer risk at a nominal significant *p*-value < 0.05 (Table 1), and in total 11 of the 15 SNPs had the point estimates of the ORs in the same direction, as previously reported in the iCOGS study (indicated in bold in Table 1) [7,8]. Subsequently, a PRS consisting of these 15 SNPs was constructed.

PRS<sub>15</sub> distribution plots are shown for both controls (blue) and cases (pink) in Figure 1. The average PRS<sub>15</sub> was higher in cases [mean (SD) = 0.645 (0.312)] compared to controls [mean (SD) = 0.595 (0.316)] (Figure 1a). The PRS<sub>15</sub> was, statistically, significantly associated with increased breast cancer risk in Cypriot women with OR (95% CI) 1.66 (1.25–2.19) and *p*-value = 0.0004. The AUC (95% CI) of PRS<sub>15</sub> was 0.55 (0.52–0.57) and was well-calibrated (Hosmer–Lemeshow test  $\chi^2 = 11.77$ , *p*-value = 0.162). As illustrated in Figure 1b, compared with females in the 2nd quartile (25–50%) of the PRS<sub>15</sub> risk distribution, the estimated OR (95% CI) for those in the first quartile was 0.98 (0.76–1.27), *p*-value = 0.88, whereas for those in the fourth quartile was 1.51 (1.19–1.94), *p*-value = 0.0009 (Figure 1b; Table S2A). The estimated OR of PRS<sub>15</sub> did not change substantially when adjusted by age or family history (Table S2B).



(a)

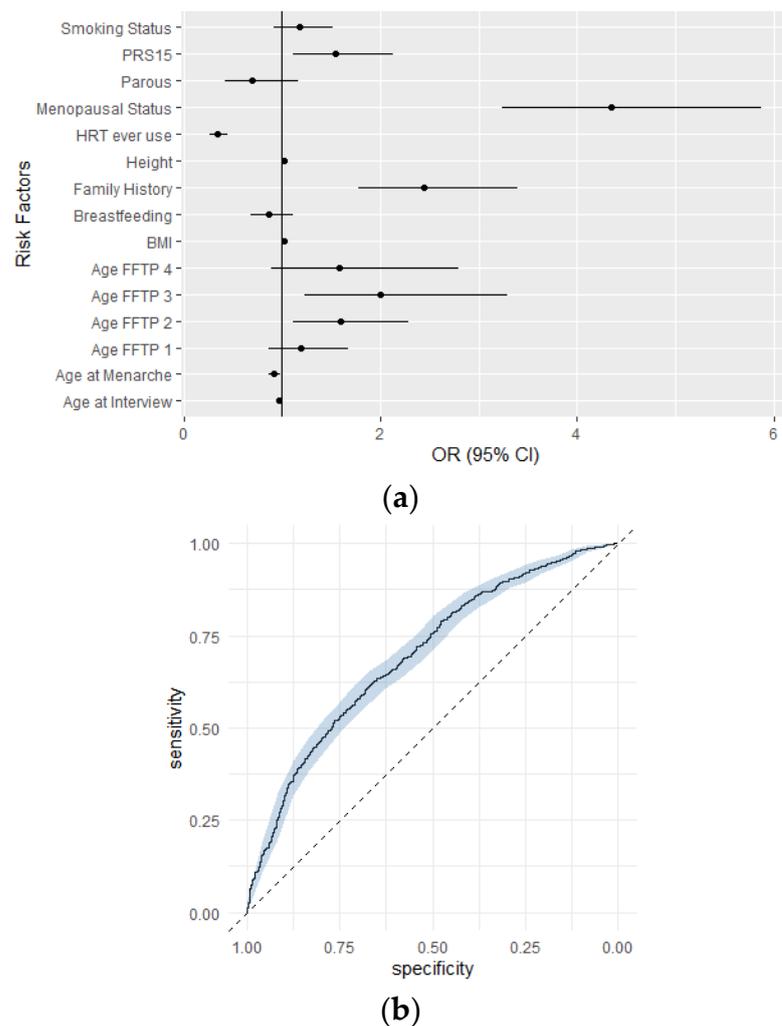


(b)

**Figure 1.** Distribution of the PRS<sub>15</sub> and its association, by quartiles, with breast cancer risk in the MASTOS study. (a) Distribution of the PRS<sub>15</sub> in controls (blue) and in cases (pink) in Cypriot women. The average PRS<sub>15</sub> was higher in cases compared with the controls (b) OR (95% CI) for breast cancer risk by quartiles of PRS<sub>15</sub> risk distribution, using as reference the 2nd quartile (25–50%). Women in the fourth quartile had a significant increased breast cancer risk compared with those in the 2nd quartile.

### 3.2. Association between the Integrated Risk Model Consisting of PRS<sub>15</sub> and Classical Risk Factors with Breast Cancer Risk in Greek-Cypriot Women

We further proceeded with the evaluation of a multivariable model including breast cancer risk factors, that were previously used in larger studies, and in combination with PRS<sub>15</sub>. Table S1 summarizes the distribution of classical risk factors in the MASTOS study. There was no evidence of interactions between PRS<sub>15</sub> and any of the classical breast cancer risk factors (Table S3). Thus, an integrated risk model that included all risk factors and PRS<sub>15</sub> was constructed. The integrated risk model achieved a risk discrimination score of AUC (95% CI) 0.70 (0.67–0.72) and was well calibrated (Hosmer-Lemeshow  $\chi^2 = 8.73$ ,  $p$ -value = 0.37) (Figure 2, Table S4). Sensitivity of the model was 0.644 and specificity was 0.624.

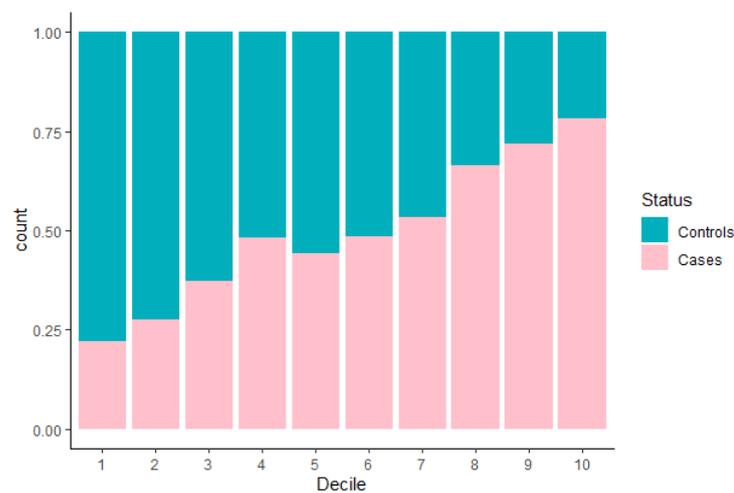


**Figure 2.** Association of the integrated risk model with breast cancer risk in the MASTOS study: (a) Associations between risk factors included in the final integrated risk model with breast cancer risk in the MASTOS study. Estimated ORs (95% CI) of each risk factor for breast cancer risk are illustrated; Age FFTP category 5 (Nulliparous) was included in the category 0 of Age FFTP (reference). (b) ROC curve for the integrated risk model, (AUC 0.70, 95% CI 0.67–0.72).

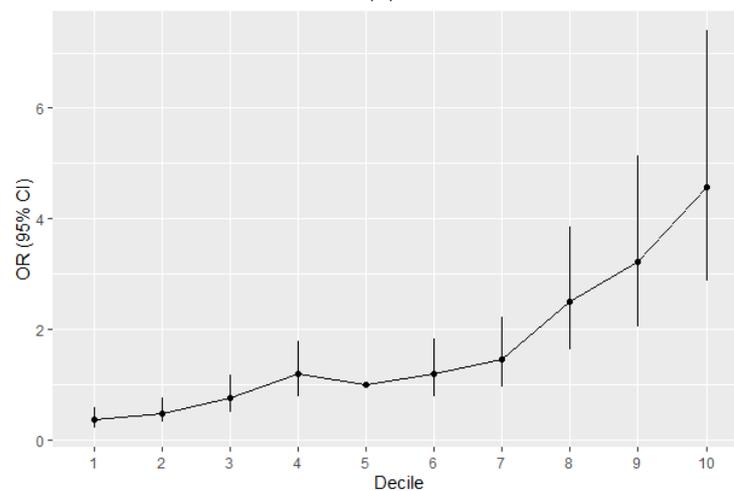
To assess the ability of the integrated risk model to discriminate individuals based on their disease status, MASTOS dataset was divided into deciles based on the predicted risk probability of the integrated risk model. As illustrated in Table 2, compared with the 5th decile, the estimated OR (95% CI) in the first decile was 0.36 (0.22–0.57), with  $p$ -value =  $1.55 \times 10^{-5}$ , and included 15.4% of controls, and 4.4% of cases. Similarly, the estimated OR (95% CI) in the 2nd quartile was 0.48 (0.31–0.75), with  $p$ -value = 0.001 and included 15% of controls and 5.8% of cases. In contrast, the estimated OR (95% CI) in the 9th decile was 3.22 (2.04–5.13) with  $p$ -value =  $6.46 \times 10^{-7}$  and included 5% of controls and 13% of cases, and in the last decile the estimated OR (95% CI) was 4.58 (2.88–7.4), with  $p$ -value =  $2.44 \times 10^{-10}$  and included 4.2% of controls and 15.6% of cases (Table 2, Figure 3).

**Table 2.** Total number of controls and cases of the MASTOS study included in each decile, when the dataset was divided based on the predicted risk probability of the integrated risk model. The estimated ORs (95% CI) of each decile for breast cancer risk were generated from logistic regression using the 5th decile as the reference.

| Decile. | Controls (%) | Cases (%)  | OR (95% CI)      | <i>p</i> -Value        |
|---------|--------------|------------|------------------|------------------------|
| 1       | 139 (15.4)   | 39 (4.4)   | 0.36 (0.22–0.57) | $1.55 \times 10^{-5}$  |
| 2       | 135 (15)     | 51 (5.8)   | 0.48 (0.31–0.75) | 0.001                  |
| 3       | 108 (12)     | 64 (7.3)   | 0.75 (0.49–1.16) | 0.2                    |
| 4       | 98 (10.9)    | 91 (10.3)  | 1.18 (0.78–1.79) | 0.44                   |
| 5       | 94 (10.4)    | 74 (8.4)   | 1                | -                      |
| 6       | 92 (10.2)    | 86 (9.8)   | 1.19 (0.78–1.82) | 0.43                   |
| 7       | 89 (9.9)     | 102 (11.6) | 1.46 (0.96–2.21) | 0.08                   |
| 8       | 62 (6.9)     | 122 (13.9) | 2.5 (1.63–3.86)  | $3.17 \times 10^{-5}$  |
| 9       | 45 (5)       | 114 (13)   | 3.22 (2.04–5.13) | $6.46 \times 10^{-7}$  |
| 10      | 38 (4.2)     | 137 (15.6) | 4.58 (2.88–7.4)  | $2.44 \times 10^{-10}$ |



(a)



(b)

**Figure 3.** Association between the integrated risk model and breast cancer risk in the MASTOS study (a) Distribution of controls (blue) and cases (pink) of the MASTOS study in each decile, when the dataset was divided based on the predicted risk probability of the integrated risk model; (b) ORs (95% CI) by decile for breast cancer risk, using the 5th decile as the reference.

#### 4. Discussion

Prevention and early detection of breast cancer are key objectives in the clinical management of the disease. The incorporation of PRS into risk prediction models that include classical breast cancer risk factors can potentially provide a more comprehensive personalized breast cancer risk estimation and, thus, has potential clinical applications in guiding targeted population screening and personalized preventative strategies. Although large studies have evaluated the performance of PRS and risk prediction models in populations of European ancestry, specific assessment among smaller European populations has not been extensively performed. Recent studies highlight the need of country-specific calibration of such risk prediction models for a more precise population-specific personalized risk estimation and classification [19,27,44].

The main aim of this study was to assess the predictive performance of a PRS consisting of 15 previously identified breast cancer susceptibility variants in combination with other classical breast cancer risk factors, and to evaluate its ability to discriminate Greek-Cypriot women based on their breast cancer risk. Hence, we demonstrated that the PRS<sub>15</sub> was associated with an increased breast cancer risk in Greek-Cypriot women, and the integrated risk model had the ability to stratify Greek-Cypriot women based on their disease status at the extreme deciles. The results of this validation study support the potential clinical utility of a combined risk estimation model that will include PRS and classical risk factors for providing individualized breast cancer risk information and guiding targeted screening and preventative actions in our population.

To date, more than 200 common, low-risk SNPs have been discovered through GWAS to be associated with breast cancer predisposition in women of European ancestry, and a PRS<sub>313</sub> has been constructed for the prediction of breast cancer risk in European women [9,14]. In our study, we included 15 SNPs identified by GWAS, before 2013, which have higher effect sizes compared to most of the recently discovered SNPs and constructed a PRS. Of the 15 SNPs included in the PRS only a small number (4/15) were individually associated with breast cancer risk at nominal significant *p*-value in our study population, with the effect size of a larger number of the SNPs (11/15) being in the same direction, as previously described in the iCOGS study [7,8]. Based on this observation, the discriminatory ability of the combined effect of these variants was investigated. PRS<sub>15</sub> was significantly associated with an increased breast cancer risk in Greek-Cypriot women, with OR (95% CI) of 1.66 (1.25–2.19), which falls within the range of ORs of the published PRSs constructed and evaluated in women of European descent, according to the Polygenic Score (PGS) Catalog [45]. The AUC (95% CI) of PRS<sub>15</sub> was 0.55 (0.52–0.57), which was slightly lower compared with previous studies in European populations where the AUC ranged between 0.58–0.65 [30]. Incorporation of additional susceptibility SNPs in the PRS could potentially improve its discrimination power. Despite the modest predictive accuracy of PRS<sub>15</sub>, women in the highest quartile of the PRS<sub>15</sub> risk distribution had a statistically significant 1.5-times greater breast cancer risk compared to the average.

Similar to other studies, we did not identify any significant interactions between the PRS and the other classical breast cancer risk factors (Table S3, Figure S1) [20,24,46]; thus, a multivariable model consisting of the PRS<sub>15</sub> and all the risk factors was constructed. Combination of PRS<sub>15</sub> with classical risk factors achieved a discrimination score of AUC (95% CI) 0.70 (0.67–0.72), and had the ability to stratify women based on their disease status at the extreme deciles which are the most important for risk-stratified preventative strategies (first decile included 15.4% of controls, and 4.4% of cases, whereas the last decile included 4.2% of controls and 15.6% of cases). Our results, provide evidence that such integrated risk models have the capacity to stratify Greek-Cypriot women based on their personal breast cancer risk.

These results are consistent with previously published studies, demonstrating that integrated risk models, including a PRS, and classical breast cancer risk factors have the ability to stratify women according to their disease risk [19,20,24,26]. Recently, Triviño et al. 2020 have evaluated the predictive performance of an integrated risk model consisting of a

PRS<sub>92</sub> and 5 phenotypic risk factors in a cohort of Spanish women. Similarly to our results, the integrated risk model had the ability to stratify women according to their disease risk at the extreme deciles, and had a slightly higher predictive power compared to our study, AUC (95% CI) 0.80 (0.77–0.83) [22]. Additionally, van Veen et al. (2018) have used a PRS incorporating the effects of 18 SNPs, 9 of which were identical to the SNPs included in our analysis while another 5 were surrogate variants within the same gene/locus. The study concluded that the incorporation of PRS<sub>18</sub> within the Tyrer–Cuzick model in combination with mammographic density, could substantially improve risk prediction accuracy, with AUC (95% CI) 0.67 (0.62–0.71) [26].

The main limitation of our study is the small number of SNPs included in the PRS. The inclusion of the PRS into a model with only including the classical risk factors did not make a substantial increase in the predictive performance of the model. In the future, larger studies, and incorporation of additional SNPs in the PRS, as well as additional risk factors in the integrated risk model, such as mammographic density, are needed for defining the best tool to be used in our population.

In this study, we evaluated the predictive performance of the combined effect of a PRS with classical risk factors for the prediction of breast cancer risk in Greek-Cypriot women. Despite the limited number of SNPs included in the risk prediction model, our study highlights that it is worth assessing further the clinical utility of PRS for Greek-Cypriot women.

## 5. Conclusions

In this study, we evaluated the predictive performance of a polygenic risk score consisting of 15 previously identified SNPs, in combination with classical breast cancer risk factors in women of Greek-Cypriot origin. Although the PRS and risk prediction models are extensively evaluated in individuals of European ancestry, no previous study included women from our population. This study demonstrates that polygenic information in combination with classical risk factors has the power to stratify Greek-Cypriot women based on their disease risk. These results suggest the potential clinical utility for the use of a combined model for the prediction of breast cancer risk in Cyprus.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/cancers13184568/s1>, Figure S1: Heatmap of Spearman Pairwise Correlation between all the risk variables included in the analysis, in the control group of the MASTOS study population, Table S1: Phenotypic characteristics in the MASTOS study, Table S2A: Odds ratio (95% CI) by quartiles of PRS<sub>15</sub> risk distribution generated using the 2nd quartile (25–50%) as the reference, Table S2B: PRS<sub>15</sub> adjusted by family history and age, Table S3A: Table of Spearman Pairwise Correlation  $r$  between risk variables in the MASTOS study, Table S3B: Table of Spearman Pairwise Correlation  $p$ -value between risk variables in the MASTOS study, Table S4: Associations between the risk factors of the final integrated risk model with breast cancer risk in Cypriot women.

**Author Contributions:** Conceptualization, K.M., A.H. and M.A.L.; methodology, K.M., A.H., M.A.L., K.K., E.K., Y.M. and K.Y.; formal analysis, K.Y. and K.M.; resources, A.H. and M.I.P.; writing—original draft preparation, K.M. and K.Y.; writing—review and editing, M.A.L., A.H., K.K., M.I.P., E.K., Y.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Cyprus Research Promotion Foundation grants 0104/13, 0104/17 and Telethon.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Cyprus National Bioethics Committee (EEBK/EII/2005/08).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data presented in this study are available on request from ahsavas@cing.ac.cy. The data are not publicly available due to ethical and privacy restrictions.

**Acknowledgments:** We would like to thank all the individuals who participated in this study.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)] [[PubMed](#)]
2. Green, J.; Cairns, B.J.; Casabonne, D.; Wright, F.L.; Reeves, G.; Beral, V. Height and cancer incidence in the Million Women Study: Prospective cohort, and meta-analysis of prospective studies of height and total cancer risk. *Lancet Oncol.* **2011**, *12*, 785–794. [[CrossRef](#)]
3. Hamajima, N.; Hirose, K.; Tajima, K.; Rohan, T.; Calle, E.E.; Heath, C.W.; Coates, R.J.; Liff, J.M.; Talamini, R.; Chantarakul, N.; et al. Alcohol, tobacco and breast cancer—Collaborative reanalysis of individual data from 53 epidemiological studies, including 58 515 women with breast cancer and 95 067 women without the disease. *Br. J. Cancer* **2002**, *87*, 1234–1245. [[CrossRef](#)]
4. Hamajima, N.; Hirose, K.; Tajima, K.; Rohan, T.; Friedenreich, C.M.; Calle, E.E.; Gapstur, S.M.; Patel, A.V.; Coates, R.J.; Liff, J.M.; et al. Menarche, menopause, and breast cancer risk: Individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *Lancet Oncol.* **2012**, *13*, 1141–1151. [[CrossRef](#)]
5. Morimoto, L.M.; White, E.; Chen, Z.; Chlebowski, R.T.; Hays, J.; Kuller, L.; Lopez, A.M.; Manson, J.A.; Margolis, K.L.; Muti, P.C.; et al. Obesity, body size, and risk of postmenopausal breast cancer: The women’s health initiative (United States). *Cancer Causes Control* **2002**, *13*, 741–751. [[CrossRef](#)] [[PubMed](#)]
6. Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormone replacement therapy: Collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Collaborative Group on Hormonal Factors in Breast Cancer. *Lancet* **1997**, *350*, 1047–1059. [[CrossRef](#)]
7. Michailidou, K.; Beesley, J.; Lindstrom, S.; Canisius, S.; Dennis, J.; Lush, M.J.; Maranian, M.J.; Bolla, M.K.; Wang, Q.; Shah, 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. [[CrossRef](#)] [[PubMed](#)]
8. Michailidou, K.; Hall, P.; Gonzalez-Neira, A.; Ghoussaini, M.; Dennis, J.; Milne, R.L.; Schmidt, M.K.; Chang-Claude, J.; Bojesen, S.E.; Bolla, M.K.; et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* **2013**, *45*, 353–361. [[CrossRef](#)] [[PubMed](#)]
9. Michailidou, K.; Lindström, S.; Dennis, J.; Beesley, J.; Hui, S.; Kar, S.; Lemaçon, A.; Soucy, P.; Glubb, D.; Rostamianfar, A.; et al. Association analysis identifies 65 new breast cancer risk loci. *Nature* **2017**, *551*, 92–94. [[CrossRef](#)] [[PubMed](#)]
10. Zhang, H.; Ahearn, T.U.; Lecarpentier, J.; Barnes, D.; Beesley, J.; Qi, G.; Jiang, X.; O’Mara, T.A.; Zhao, N.; Bolla, M.K.; et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nat. Genet.* **2020**, *52*, 572–581. [[CrossRef](#)]
11. Pashayan, N.; Morris, S.; Gilbert, F.J.; Pharoah, P.D.P. Cost-effectiveness and Benefit-to-Harm Ratio of Risk-Stratified Screening for Breast Cancer A Life-Table Model. *JAMA Oncol.* **2018**, *4*, 1504–1510. [[CrossRef](#)]
12. Burton, H.; Chowdhury, S.; Dent, T.; Hall, A.; Pashayan, N.; Pharoah, P. Public health implications from COGS and potential for risk stratification and screening. *Nat. Genet.* **2013**, *45*, 349–351. [[CrossRef](#)]
13. Khera, A.V.; Chaffin, M.; Aragam, K.G.; Haas, M.E.; Roselli, C.; Choi, S.H.; Natarajan, P.; Lander, E.S.; Lubitz, S.A.; Ellinor, P.T.; et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **2018**, *50*, 1219–1224. [[CrossRef](#)]
14. Mavaddat, N.; Michailidou, K.; Dennis, J.; Lush, M.; Fachal, L.; Lee, A.; Tyrer, J.P.; Chen, T.H.; Wang, Q.; Bolla, M.K.; et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am. J. Hum. Genet.* **2019**, *104*, 21–34. [[CrossRef](#)] [[PubMed](#)]
15. Mavaddat, N.; P Pharoah, P.D.; Michailidou, K.; Tyrer, J.; Brook, M.N.; Bolla, M.K.; Wang, Q.; Dennis, J.; Dunning, A.M.; Shah, M.; et al. Prediction of Breast Cancer Risk Based on Profiling With Common Genetic Variants. *JNCI J. Natl. Cancer Inst.* **2015**, *107*, 36. [[CrossRef](#)] [[PubMed](#)]
16. Torkamani, A.; Wineinger, N.E.; Topol, E.J. The personal and clinical utility of polygenic risk scores. *Nat. Rev. Genet.* **2018**, *19*, 581–590. [[CrossRef](#)]
17. Zhang, Y.D.; Hurson, A.N.; Zhang, H.; Choudhury, P.P.; Easton, D.F.; Milne, R.L.; Simard, J.; Hall, P.; Michailidou, K.; Dennis, J.; et al. Assessment of polygenic architecture and risk prediction based on common variants across fourteen cancers. *Nat. Commun.* **2020**, *11*, 3353. [[CrossRef](#)] [[PubMed](#)]
18. Chatterjee, N.; Shi, J.; García-Closas, M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat. Rev. Genet.* **2016**, *17*, 392–406. [[CrossRef](#)] [[PubMed](#)]
19. Hurson, A.N.; Pal Choudhury, P.; Gao, C.; Hüsing, A.; Eriksson, M.; Shi, M.; Jones, M.E.; Evans, D.G.R.; Milne, R.L.; Gaudet, M.M.; et al. Prospective evaluation of a breast-cancer risk model integrating classical risk factors and polygenic risk in 15 cohorts from six countries. *Int. J. Epidemiol.* **2021**. [[CrossRef](#)]
20. Maas, P.; Barrdahl, M.; Joshi, A.D.; Auer, P.L.; Gaudet, M.M.; Milne, R.L.; Schumacher, F.R.; Anderson, W.F.; Check, D.; Chattopadhyay, S.; et al. Breast Cancer Risk from Modifiable and Nonmodifiable Risk Factors among White Women in the United States. *JAMA Oncol.* **2016**, *2*, 1295–1302. [[CrossRef](#)] [[PubMed](#)]
21. Wand, H.; Lambert, S.A.; Tamburro, C.; Iacocca, M.A.; O’Sullivan, J.W.; Sillari, C.; Kullo, I.J.; Rowley, R.; Dron, J.S.; Brockman, D.; et al. Improving reporting standards for polygenic scores in risk prediction studies. *Nature* **2021**, *591*, 211–219. [[CrossRef](#)]

22. Triviño, J.C.; Ceba, A.; Rubio-Solsona, E.; Serra, D.; Sanchez-Guiu, I.; Ribas, G.; Rosa, R.; Cabo, M.; Bernad, L.; Pita, G.; et al. Combination of phenotype and polygenic risk score in breast cancer risk evaluation in the Spanish population: A case–control study. *BMC Cancer* **2020**, *20*, 1079. [[CrossRef](#)]
23. Garcia-Closas, M.; Gunsoy, N.B.; Chatterjee, N. Combined associations of genetic and environmental risk factors: Implications for prevention of breast cancer. *J. Natl. Cancer Inst.* **2014**, *106*, dju305. [[CrossRef](#)]
24. Kapoor, P.M.; Mavaddat, N.; Choudhury, P.P.; Wilcox, A.N.; Lindström, S.; Behrens, S.; Michailidou, K.; Dennis, J.; Bolla, M.K.; Wang, Q.; et al. Combined Associations of a Polygenic Risk Score and Classical Risk Factors With Breast Cancer Risk. *J. Natl. Cancer Inst.* **2021**, *113*, 329–337. [[CrossRef](#)]
25. Lee, A.; Mavaddat, N.; Wilcox, A.N.; Cunningham, A.P.; Carver, T.; Hartley, S.; Babb de Villiers, C.; Izquierdo, A.; Simard, J.; Schmidt, M.K.; et al. BOADICEA: A comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genet. Med.* **2019**, *21*, 1708–1718. [[CrossRef](#)] [[PubMed](#)]
26. Van Veen, E.M.; Brentnall, A.R.; Byers, H.; Harkness, E.F.; Astley, S.M.; Sampson, S.; Howell, A.; Newman, W.G.; Cuzick, J.; Evans, G.R. Use of single-nucleotide polymorphisms and mammographic density plus classic risk factors for breast cancer risk prediction. *JAMA Oncol.* **2018**, *4*, 476–482. [[CrossRef](#)] [[PubMed](#)]
27. Choudhury, P.P.; Wilcox, A.N.; Brook, M.N.; Zhang, Y.; Ahearn, T.; Orr, N.; Coulson, P.; Schoemaker, M.J.; Jones, M.E.; Gail, M.H.; et al. Comparative validation of breast cancer risk prediction models and projections for future risk stratification. *J. Natl. Cancer Inst.* **2021**, *112*, 278–285. [[CrossRef](#)]
28. Lakeman, I.M.M.; Rodriguez-Gironde, M.; Lee, A.; Ruiters, R.; Stricker, B.H.; Wijnant, S.R.A.; Kavousi, M.; Antoniou, A.C.; Schmidt, M.K.; Uitterlinden, A.G.; et al. Validation of the BOADICEA model and a 313-variant polygenic risk score for breast cancer risk prediction in a Dutch prospective cohort. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **2020**, *22*, 1803–1811. [[CrossRef](#)] [[PubMed](#)]
29. Zhang, X.; Rice, M.; Tworoger, S.S.; Rosner, B.A.; Eliassen, A.H.; Tamimi, R.M.; Joshi, A.D.; Lindstrom, S.; Qian, J.; Colditz, G.A.; et al. Addition of a polygenic risk score, mammographic density, and endogenous hormones to existing breast cancer risk prediction models: A nested case-control study. *PLoS Med.* **2018**, *15*, e1002644. [[CrossRef](#)] [[PubMed](#)]
30. Yanes, T.; Young, M.A.; Meiser, B.; James, P.A. Clinical applications of polygenic breast cancer risk: A critical review and perspectives of an emerging field. *Breast Cancer Res.* **2020**, *22*, 21. [[CrossRef](#)]
31. Pavlou, P.; Demetriou, A. Cyprus Cancer Registry (CyCR). In *Progress Report August 2016*; Ministry of Health: Nicosia, Cyprus, 2016; p. 86.
32. Loizidou, M.A.; Hadjisavvas, A.; Ioannidis, J.P.A.; Kyriacou, K. Replication of genome-wide discovered breast cancer risk loci in the Cypriot population. *Breast Cancer Res. Treat.* **2011**, *128*, 267–272. [[CrossRef](#)] [[PubMed](#)]
33. Hadjisavvas, A.; Loizidou, M.A.; Middleton, N.; Michael, T.; Papachristoforou, R.; Kakouri, E.; Daniel, M.; Papadopoulos, P.; Malas, S.; Marcou, Y.; et al. An investigation of breast cancer risk factors in Cyprus: A case control study. *BMC Cancer* **2010**, *10*, 447. [[CrossRef](#)]
34. Ahmed, S.; Thomas, G.; Ghousaini, M.; Healey, C.S.; Humphreys, M.K.; Platte, R.; Morrison, J.; Maranian, M.; Pooley, K.A.; Luben, R.; et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat. Genet.* **2009**, *41*, 585–590. [[CrossRef](#)] [[PubMed](#)]
35. Easton, D.F.; Pooley, K.A.; Dunning, A.M.; Pharoah, P.D.P.; Thompson, D.; Ballinger, D.G.; Struwing, J.P.; Morrison, J.; Field, H.; Luben, R.; et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **2007**, *447*, 1087–1093. [[CrossRef](#)]
36. Ghousaini, M.; Fletcher, O.; Michailidou, K.; Turnbull, C.; Schmidt, M.K.; Dicks, E.; Dennis, J.; Wang, Q.; Humphreys, M.K.; Luccarini, C.; et al. Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nat. Genet.* **2012**, *44*, 312–318. [[CrossRef](#)]
37. Stacey, S.N.; Manolescu, A.; Sulem, P.; Rafnar, T.; Gudmundsson, J.; Gudjonsson, S.A.; Masson, G.; Jakobsdottir, M.; Thorlacius, S.; Helgason, A.; et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat. Genet.* **2007**, *39*, 865–869. [[CrossRef](#)]
38. Thomas, G.; Jacobs, K.B.; Kraft, P.; Yeager, M.; Wacholder, S.; Cox, D.G.; Hankinson, S.E.; Hutchinson, A.; Wang, Z.; Yu, K.; 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. [[CrossRef](#)] [[PubMed](#)]
39. Turnbull, C.; Ahmed, S.; Morrison, J.; Pernet, D.; Renwick, A.; Maranian, M.; Seal, S.; Ghousaini, M.; Hines, S.; Healey, C.S.; et al. Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat. Genet.* **2010**, *42*, 504–507. [[CrossRef](#)] [[PubMed](#)]
40. Zheng, W.; Long, J.; Gao, Y.T.; Li, C.; Zheng, Y.; Xiang, Y.B.; Wen, W.; Levy, S.; Deming, S.L.; Haines, J.L.; et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat. Genet.* **2009**, *41*, 324–328. [[CrossRef](#)]
41. Team, R.C. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
42. Pharoah, P.D.; Antoniou, A.; Bobrow, M.; Zimmern, R.L.; Easton, D.F.; Ponder, B.A. Polygenic susceptibility to breast cancer and implications for prevention. *Nat. Genet.* **2002**, *31*, 33–36. [[CrossRef](#)] [[PubMed](#)]
43. Robin, X.; Turck, N.; Hainard, A.; Tiberti, N.; Lisacek, F.; Sanchez, J.C.; Müller, M. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **2011**, *12*, 77. [[CrossRef](#)] [[PubMed](#)]
44. Garcia-Closas, M.; Chatterjee, N. Assessment of breast cancer risk: Which tools to use? *Lancet Oncol.* **2019**, *20*, 463–464. [[CrossRef](#)]

- 
45. Lambert, S.A.; Gil, L.; Jupp, S.; Ritchie, S.C.; Xu, Y.; Buniello, A.; McMahon, A.; Abraham, G.; Chapman, M.; Parkinson, H.; et al. The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. *Nat. Genet.* **2021**, *53*, 420–425. [[CrossRef](#)] [[PubMed](#)]
  46. Rudolph, A.; Song, M.; Brook, M.N.; Milne, R.L.; Mavaddat, N.; Michailidou, K.; Bolla, M.K.; Wang, Q.; Dennis, J.; Wilcox, A.N.; et al. Joint associations of a polygenic risk score and environmental risk factors for breast cancer in the Breast Cancer Association Consortium. *Int. J. Epidemiol.* **2018**, *47*, 526–536. [[CrossRef](#)] [[PubMed](#)]