The Discovery of Breast Cancer Risk Gene and Establishment of Prediction Model Based on Estrogen Metabolism Regulation

Feng Zhao  
Xuzhou Medical University

Zhixiang Hao  
Xuzhou Medical University

Yanan Zhong  
Xuzhou Medical University

Yinxue Xu  
Xuzhou Medical University

Meng Guo  
the Affiliated Hospital of Xuzhou Medical University

Bei Zhang  
Xuzhou Central Hospital

Xiaoxing Yin  
Xuzhou Medical University

Ying Li  
the Affiliated Hospital of Xuzhou Medical University

Xueyan Zhou (✉️ zxy851107@126.com)  
Xuzhou Medical University  https://orcid.org/0000-0002-2993-9842

Research article

Keywords: Breast cancer, Risk prediction, Estrogens, Estrogen metabolizing enzyme, Gene polymorphism, Polygenic Risk Score

DOI: https://doi.org/10.21203/rs.3.rs-50139/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background

In this study, we aim to uncover the relationship between estrogen levels and the genetic polymorphism of estrogen metabolism-related enzymes with breast cancer (BC) and establish a risk prediction model based on polygenic risk score.

Methods

Unrelated BC patients and healthy subjects were recruited for analysis of the estrogen levels and the single nucleotide polymorphisms (SNPs) of estrogen metabolism-related enzymes. The polygenic risk score (PRS) was used to explore the combined effect of multiple genes which was calculated using a Bayesian approach. The independent sample t test was used to evaluate the difference between PRS scores of BC and healthy subjects. Discriminatory accuracy of the models was compared using the area under the receiver operating characteristic curve (ROC).

Results

The estrogen homeostasis profile was disturbed in BC patients, with parent estrogens (E1, E2) and carcinogenic catechol estrogens (2/4-OHE1, 2-OHE2, 4-OHE2) significantly accumulated in the serum of BC patients. Then, we established PRS model to evaluate the role of multiple genes SNPs. The PRS model 1 (M1) was established from 6 GWAS-identified high risk genes SNPs. On the basis of M1, we added 7 estrogen metabolism enzyme genes SNPs to establish PRS model 2 (M2). The independent sample t test results show that there is no difference between BC and healthy subjects in M1 ($P = 0.17$), however, there is significant difference between BC and healthy subjects in M2 ($P = 4.9 \times 10^{-5}$). The ROC curve results also show that the accuracy of M2 (AUC = 62.18%) in breast cancer risk identification was better than M1 (AUC = 54.56%).

Conclusion

Estrogens and the related metabolic enzymes gene polymorphisms are closely related to BC. The model constructed by adding estrogen metabolic enzyme genes SNPs has a good ability in breast cancer risk prediction, and the accuracy is greatly improved comparing PRS model only includes GWAS-identified genes SNPs.

1. Background

Breast cancer is the most common cancer among women, accounting for 33% of all women's cancer cases, and 15% of all cancer deaths among women worldwide. According to the World Health
Organization (WHO), there will be more than 20 million new cases of breast cancer worldwide by 2025, which will seriously endanger women's lives and health [1]. With the continuous improvement of diagnosis and treatment methods, the survival rate of breast cancer patients has been greatly improved. Early prediction, early detection, and early treatment of high-risk groups are the key issues that need to be solved urgently in the clinic.

The occurrence and development of breast cancer are closely related to genetic and environmental factors. Gail in 1989 proposed the breast cancer risk prediction model, which included factors such as age at evaluation, age at menarche, age at first live birth, race, number of breast -, and family history of breast cancer [2, 3]. Some subsequent prediction models also involved BRCA1/2, estrogen replacement therapy, mammography screening times, and genetic polymorphism. Rare high-risk mutations particularly in the BRCA1 and BRCA2 genes explain less than 20% of the two-fold familial relative risk (FRR) and account for a small proportion of breast cancer cases in the general population. Low frequency variants conferring intermediate risk, such as those in CHEK2, ATM, and PALB2, explain 2–5% of the FRR [4]. Genome-wide association studies (GWAS) have led to the discovery of multiple common, low-risk variants (single nucleotide polymorphisms [SNPs]) associated with breast cancer risk [5]. Recently, it is found that genetic risk factors can account for 31% in the breast cancer risk evaluation [6], which indicates that breast cancer is a multifactorial disease, and genetic factor is the important etiological factor for the occurrence and development of breast cancer. At present, more and more researchers are inclined to develop a comprehensive genetic risk scoring method to evaluate the polygenic effects of single nucleotide polymorphisms (SNPs) basing on GWAS [7–9]. Some well-known studies such as Mavaddat et al. used GWAS-selected 77-SNPs to construct a PRS for BC. Comparing the highest 1% with middle quintile polygenic scores, they found that the risk increased by three times [38].

GWAS also have their own limitations. Firstly, a major limitation of genome-wide approaches is the need to adopt a high level of significance to account for the multiple tests. Secondly, GWAS explain only a modest fraction of the missing heritability [10]. As we all know, estrogen is an important risk factor for breast cancer. Long-term exposed, super-physiological concentrations of estrogen can bind to estrogen receptors, mediate the over-expression of various growth factors, and promote the growth and proliferation of cells; and various metabolites of estrogen can form adducts with DNA It induces genetic mutations and produces direct genotoxicity[11]. Thus, it is an important risk factor for breast cancer development that the estrogen and its toxic metabolites accumulate abnormally in the breast tissue. Estrogen homeostasis is regulated by estrogen-related metabolic enzymes. Endogenous estrogens are metabolized to be 2-, 4- and 16α-hydroxy estrogens, which are catalyzed by phase I metabolizing enzymes of cytochrome P450 CYP1A1, CYP1B1 and CYP3A4, respectively [12–14]. Hydroxy-estrogens are detoxified by conjugation reactions catalyzed by phase II metabolizing enzymes such as COMT, UGTs and SULTs. Thus, the expression level of the estrogen and its toxic metabolites can be considered to be a comprehensive reflection of the role of these estrogen metabolic enzymes to a certain extent. Polymorphisms in genes encoding these estrogen-related metabolic enzymes,which are reported to be closely related to differences in the enzyme activities and alter the levels of DNA-damaging species to influence the individual's susceptibility to breast cancer [13, 15, 16]. Genetic epidemiology studies
suggested that there is a correlation between polymorphisms in estrogen metabolism genes and breast cancer risk, however, these results are not consistent [16–18]. It is an important reason for the inconsistency of existing research results that studying the correlation between gene polymorphisms of estrogen metabolic enzymes and breast cancer in isolation. Currently, Breast cancer risk gene prediction models have not taken estrogen metabolic enzymes genes into consideration, therefore, the further optimization is needed from the perspective of estrogen metabolism overall level.

Based on the above analysis, our research aims to reveal the form of estrogen homeostasis disorders in breast cancer and explore the association between the metabolic enzyme genes polymorphisms and breast cancer occurrence from the overall level of estrogen metabolism. Furthermore, we developed a risk score composing GWAS-selected SNPs and the estrogens metabolic enzyme genes SNPs to optimize the breast cancer risk prediction model.

2. Methods

2.1 Chemicals

The standards and other chemical reagents came from our previously published study [19].

2.2 Clinical sample collection

The serum in the follicular and luteal phase of premenopausal 64 women (mean age: 45.5 ± 5.04 years) firstly diagnosed with BC and 49 matched healthy women (mean age: 43.7 ± 8.80 years) were collected to detect the level of estrogens. The blood of premenopausal 140 women (mean age: 43.3 ± 6.24 years) firstly diagnosed with BC and 140 matched healthy women (mean age: 40.2 ± 3.52 years) were collected to extract DNA and analyze SNP genotype. All samples and related data have been obtained from the Affiliated Hospital of Xuzhou Medical University, Xuzhou, China from June 2017 to May 2019. The patients with BC were enrolled from the Department of Nail Surgery, whereas the healthy subjects were enrolled from the Physical examination center. The collection of blood samples were operated before any therapy.

The enrollment criteria were as follows: No history of smoking; BMI ranged from 19 to 26; menarche age is between 12 and 16 years; all healthy volunteers and breast cancer patients did not receive chemotherapy, radiotherapy, or estrogen-related endocrine therapy during collecting blood samples. This protocol was approved by the Ethics Committee of the Affiliated Hospital of Xuzhou Medical University. Written informed consent was obtained from each subject before the study.

2.3 Quantification of estrogens using a LC-MS/MS method

The LC-MS/MS method referred to our previously published [19].

2.4 Genotyping analysis
DNA was extracted from peripheral whole blood with a Tiangen DNA extraction kit (Biotech, Beijing, China). The main metabolic enzymes CYP19A1, CYP1A1, CYP1B1, HSD17B1, COMT, UGTs, and SULTs are involved in the regulation of estrogen metabolism. In this study, according to the previous study and pharmacogenomic database, 1 gene loci that are more common or affect the function and activity of metabolic enzymes were screened from each metabolic enzyme. At the same time, we also using GWAS-identified breast-cancer related SNPs according some previous studies reported [20]. All selected SNPs were potentially functional variants, with minor allelic frequencies of more than 10% in MAF. The allelic discrimination of the following SNPs was performed by SNaPshot assay (Applied Biosystems Inc., Waltham, MA, USA): Estrogens metabolic enzymes genes SNPs including CYP19A1 (rs700519), CYP1A1 (rs1048943), CYP1B1 (rs1056827), COMT (rs4680), HSD17B1 (rs605059), SULT1A1 (rs1042028), UGT2B7 (rs7439366) and the GWAS-identified high breast risk genes SNPs including ZNF365 (rs10822013), FGFR2 (rs2981579), RAD51B (rs3784099), TOX3 (rs3803662), MAP3K1 (rs889312), HCN1 (rs981782). The allelic discrimination analysis was performed by Genesky Biotechnologies Inc., Shanghai, China (http://www.geneskybiotech.com). Detailed information about the SNPs basic information can be found in Table 1. To assure genotyping quality, detailed quality control (QC) procedures, including the duplicate identification of genotypes, a Hardy–Weinberg equilibrium (HWE) test, were carried out. All the 13 SNPs were successfully genotyped in 280 subjects with call rates of 100%. 
### Table 1
The basic information and HWE testing of each estrogens metabolizing enzymes gene polymorphisms

| Gene   | rs number  | Chromosome position       | Domain | Alleles | Amino acid change | Metabolism estrogens | Test for HWE (p) |
|--------|------------|---------------------------|--------|---------|-------------------|----------------------|-----------------|
| CYP19A1 | rs700519   | Chr15: 51507968 exon7    | G/A    | Arg264Cys | E1(E2)            |                      | 0.392           |
| CYP1A1  | rs1048943  | Chr15: 75012985 exon7    | T/C    | Ile462Val | 2-OHE1(E2)        |                      | 0.241           |
| CYP1B1  | rs1056827  | Chr2: 38302177 exon2     | C/A    | Ala119Ser | 4-OHE1(E2)        |                      | 0.602           |
| HSD17B1 | rs605059   | Chr17: 40706906 exon6    | G/A    | Gly313Ser | E2                |                      | 0.106           |
| COMT    | rs4680     | Chr22: 19951271 exon4    | G/A    | Val158Met | 2(4)-MeOE1(2)    |                      | 1.000           |
| SULT1A1 | rs1042028  | Chr16: 28617514 exon7    | C/T    | Arg213His | Sulfated metabolites |                | 0.144           |
| UGT2B7  | rs7439366  | Chr4: 69964338 exon2     | C/T    | Tyr268His | Glucuronide metabolites |        | 0.086           |
| ZNF365  | rs10822013 | Chr10: 64251977 intron4  | C/T    | /        | /                 |                      | 0.478           |
| FGFR2   | rs2981579  | Chr10: 123337335 intron2 | G/A    | /        | /                 |                      | 0.665           |
| TOX3    | rs3803662  | Chr16: 52586341 exon4    | G/A    | /        | /                 |                      | 0.360           |
| RAD51B  | rs3784099  | Chr14: 68749927 intron7  | G/A    | /        | /                 |                      | 0.456           |
| MAP3K1  | rs889312   | Chr5: 56031884 /         | A/C    | /        | /                 |                      | 0.776           |
| HCN1    | rs981782   | Chr5: intron6 A/C        | /      | /        | /                 |                      | 0.818           |

HWE, Hardy–Weinberg equilibrium.

### 2.5 Statistical analysis

SPSS 22.0 software was used to perform Statistical analysis. We used means ± SEM to express all estrogens' data, and the Student's t-test to test differences between the two groups. Multivariate analysis was performed using SIMCA 14.0 software.

HWE was examined among controls using a goodness-of-fit Chi-squared test. The odds ratio (OR) and 95% confidence interval (CI) were calculated using a logistic regression model to assess the association.
between the SNPs and the risk of breast cancer.

We established PRS to estimate the multi-gene contribution of estrogen-metabolic enzyme gene loci for breast cancer susceptibility, which was created using marginally significant SNPs associated with breast cancer risk based on the per-allele models. For strong linkage disequilibrium SNPs located on the same gene or chromosome, we choose the one variant with the highest OR value in the per-allele model as a candidate. The basic formula is as follows:

\[
\text{PRS} = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \beta_n x_n
\]

Where \( \beta_k \) is the per-allele OR for breast cancer associated with the minor allele for SNP \( k \), and \( x_k \) the number of alleles for the same SNP (0, 1, or 2).

3. Result

3.1 Disorders of estrogen expression in breast cancer patients

Using LC-MS/MS quantitative analysis, we measured the expression levels of 11 serum estrogens and metabolites in 64 cases of premenopausal BC (mean age: 45.5 ± 5.04 years) and 49 matched controls (mean age: 43.7 ± 8.80 years). We found that there was no significant difference in the age between the BC group and NC group. As shown in Fig. 1A, compared with NC group, the estrogen levels of BC group, especially E1, E2, 2-OHE2, 4-OHE2 (P < 0.01) and 2/4-OHE1 (P < 0.05) increased significantly. OPLS-DA was constructed as an unsupervised statistical method to identify potential estrogen homeostatic changes between the two groups. As can be observed in Fig. 1B, the metabolic profile of the NC group was clearly separated from the BC group, indicating there was a considerable metabolite difference between the BC group and NC group. Meanwhile, we also found that the potential biomarkers, with the VIP value higher than 1.0 in the OPLS-DA model were E1, E2, 2-OHE2, 4-OHE2 and 2/4-OHE1 in the serum of BC patients (Fig. 1C). Overall, these results supported the view that the disorder of estrogen homeostasis was closely related to the increased risks of BC.

3.2 Cohort description and Hardy–Weinberg equilibrium testing

We enrolled 140 patients firstly diagnosed with breast cancer and 140 corresponding healthy women in this study. The mean ages at diagnosis (for patients with cancer) were (43.3 ± 6.24) years and the healthy women in enrollment were (40.2 ± 3.52) years. The blood of these people was collected to extract DNA and analyze the SNP genotype. We found that there was no significant difference in the age between the BC group and NC group. The chi-square test was used to test the HWE value, \( P > 0.05 \) explained that the samples in enrollment were representative of the group. As can be seen in Table 1, all polymorphisms
were found to be in genetic equilibrium, which indicated that the observed genotype frequencies of the case and control groups were constant and representative.

### 3.3 Association of estrogen metabolizing enzyme genetic variants with breast cancer risk

Table 2 showed univariate analysis and ORs related to each metabolizing enzyme SNPs. The polymorphic genotypes of CYP1A1 rs1048943 (P = 0.007), CYP1B1 rs1056827 (P = 0.004), SULT1A1 rs1042028 (P = 0.029) showed significantly difference in distributions. Compared with the wild-type genotypes of CYP1A1 rs1048943 (TT) or SULT1A1 rs1042028 (CC), the heterozygous variant genotypes of CYP1A1 rs1048943 (TC) or SULT1A1 rs1042028 (CT) showed significantly higher risk in breast cancer, with an OR of 2.37 (95% confidence interval [CI] = 1.27–4.43) and 2.21 (95% CI = 1.20–4.05), respectively. Compared with the wild-type genotypes of CYP1B1 rs1056827 (CC), the homozygous variant genotypes (AA) showed significantly higher risk in breast cancer, yielding an OR of 6.90 (95% CI = 1.50–31.76). In addition, no associations with breast cancer risk were observed for the estrogen metabolic enzymes genes SNPs: CYP19A1 (rs700519), HSD17B1 (rs605059), COMT (rs4680), UGT2B7 (rs7439366) and the GWAS-selected SNPs including ZNF365 (rs10822013), FGFR2 (rs2981579), RAD51B (rs3784099), TOX3 (rs3803662), MAP3K1 (rs889312), HCN1 (rs981782).
Table 2
Genotype frequencies and ORs associated with each gene polymorphism in breast cancer cases and controls

| Gene and SNPs | Genotype | Control n (%) | Case n (%) | \( P \)-value\# | OR (95% CI) | \( P \)-value* |
|--------------|----------|---------------|------------|----------------|-------------|--------------|
| CYP19A1      | GG       | 97 (69.3%)    | 92 (65.7%) | 0.813          | 1           | —            |
| (rs700519)   | GA       | 37 (26.4%)    | 41 (29.3%) | 1.17 (0.69–1.98) | 0.564       |              |
|              | AA       | 6 (4.3%)      | 7 (5.0%)   | 1.23 (0.40–3.80) | 0.719       |              |
| CYP1A1       | TT       | 100 (71.4%)   | 80 (57.1%) | 0.007          | 1           | —            |
| (rs1048943)  | TC       | 31 (22.2%)    | 55 (39.3%) | 2.37 (1.27–4.43) | 0.003       |              |
|              | CC       | 9 (6.4%)      | 5 (3.6%)   | 1.10 (0.30–4.00) | 0.528       |              |
| CYP1B1       | CC       | 92 (65.7%)    | 80 (57.1%) | 0.004          | 1           | —            |
| (rs1056827)  | CA       | 48 (34.3%)    | 50 (35.7%) | 1.20 (0.73–1.97) | 0.802       |              |
|              | AA       | 0 (0.0%)      | 10 (7.2%)  | 6.90 (1.50–31.76) | 0.001      |              |
| HSD17B1      | GG       | 47 (33.6%)    | 46 (32.9%) | 0.713          | 1           | —            |
| (rs605059)   | GA       | 73 (52.1%)    | 69 (49.3%) | 0.97 (0.57–1.63) | 0.896       |              |
|              | AA       | 20 (14.3%)    | 25 (17.8%) | 1.28 (0.63–2.61) | 0.502       |              |
| COMT         | GG       | 91 (65.0%)    | 80 (57.1%) | 0.402          | 1           | —            |
| (rs4680)     | GA       | 42 (30.0%)    | 51 (36.4%) | 1.38 (0.83–2.29) | 0.212       |              |
|              | AA       | 7 (5.0%)      | 9 (6.4%)   | 1.46 (0.52–4.11) | 0.470       |              |
| SULT1A1      | CC       | 117 (83.6%)   | 98 (70.0%) | 0.029          | 1           | —            |
| (rs1042028)  | CT       | 20 (14.3%)    | 37 (26.4%) | 2.21 (1.20–4.05) | 0.010       |              |
|              | TT       | 3 (2.1%)      | 5 (3.6%)   | 1.99 (0.46–8.54) | 0.354       |              |
| UGT2B7       | CC       | 69 (49.30%)   | 64 (45.7%) | 0.824          | 1           | —            |

Values are presented as number (%) or OR (95% CI). OR, odds ratio; CI, confidence interval; SNP, single nuclear polymorphism.
| Gene and SNPs | Genotype | Control n (%) | Case n (%) | $P$-value* | OR (95% CI) | $P$-value* |
|--------------|----------|---------------|------------|-----------|-------------|-----------|
| (rs7439366)  | CT       | 60 (42.80%)   | 65 (46.4%) | 1.17      | (0.72–1.90) | 0.533     |
|              | TT       | 11 (7.90%)    | 11 (7.90%) |           | 1.08       | (0.44–2.66)| 0.870     |
| ZNF365       | CC       | 36 (25.71%)   | 43 (30.71%)| 0.640     | 1           |           |
| (rs10822013) | CT       | 75 (53.57%)   | 71 (50.71%)|           | 0.79       | (0.46–1.37)| 0.407     |
|              | TT       | 29 (20.71%)   | 26 (18.57%)|           | 0.75       | (0.38–1.50)| 0.415     |
| FGFR2        | GG       | 47 (33.57%)   | 40 (28.57%)| 0.418     | 1           |           |
| rs2981579    | GA       | 70 (50.00%)   | 69 (49.29%)|           | 1.16       | (0.68–1.98)| 0.592     |
|              | AA       | 23 (16.43%)   | 31 (22.14%)|           | 1.58       | (0.80–3.14)| 0.188     |
| RAD51B       | GG       | 111 (79.29%)  | 109 (77.86%)| 0.848     | 1           |           |
| rs3784099    | GA       | 25 (17.86%)   | 28 (20.00%)|           | 1.14       | (0.63–2.08)| 0.668     |
|              | AA       | 4 (2.86%)     | 3 (2.14%)  |           | 0.76       | (0.17–3.49)| 0.728     |
| TOX3         | GG       | 15 (10.71%)   | 18 (12.86%)| 0.664     | 1           |           |
| rs3803662    | GA       | 61 (43.57%)   | 54 (38.57%)|           | 0.83       | (0.51–1.38)| 0.475     |
|              | AA       | 64 (45.71%)   | 68 (48.57%)|           | 1.13       | (0.53–2.43)| 0.755     |
| MAP3K1       | CC       | 42 (30.00%)   | 35 (25.00%)| 0.460     | 1           |           |
| rs889312     | CA       | 67 (47.86%)   | 66 (47.14%)|           | 1.18       | (0.67–2.08)| 0.560     |
|              | AA       | 31 (22.14%)   | 39 (27.86%)|           | 1.51       | (0.79–2.89)| 0.215     |
| HCN1         | CC       | 16 (11.43%)   | 25 (17.86%)| 0.475     | 1           |           |
| rs981782     | CA       | 69 (49.29%)   | 63 (45.00%)|           | 0.92       | (0.55–1.53)| 0.737     |

Values are presented as number (%) or OR (95% CI). OR, odds ratio; CI, confidence interval; SNP, single nuclear polymorphism.
| Gene and SNPs | Genotype | Control n (%) | Case n (%) | P-value* | OR (95% CI) |
|--------------|----------|---------------|------------|----------|-------------|
| AA           |          | 55 (39.29%)   | 52 (37.14%)| 1.45 (0.68–3.08) | 0.336 |

Values are presented as number (%) or OR (95% CI). OR, odds ratio; CI, confidence interval; SNP, single nuclear polymorphism.

### 3.4 PRS breast cancer risk prediction model establishment and evaluation

Using the binary logistic regression method to calculate the OR of the per-allele model and the detailed results are shown in Table 3. We used the GWAS-identified genes high breast risk genes SNPs, namely ZNF365 (rs10822013), FGFR2 (rs2981579), RAD51B (rs3784099), TOX3 (rs3803662), MAP3K1 (rs889312), HCN1 (rs981782) to create PRS model 1 (M1) in the per-allele model. On the basis of M1, we also added the estrogens metabolic enzyme genes SNPs, namely CYP1A1 (rs1048943), CYP1B1 (rs1056827), SULT1A1 (rs1042028), CYP19A1 (rs700519), COMT (rs4680), HSD17B1 (rs605059), UGT2B7 (rs7439366) to create PRS model 2 (M2). The PRS scores were expressed as means ± SEM to find the difference between the two groups. Under the M1 and M2, the PRS data of the two groups obeyed the normal distribution; so we used an independent sample t-test to evaluate the difference between the two groups data. As shown in Table 4 and Fig. 2, the PRS scores in NC group was significantly lower than BC group in the M2 (P = 4.9*10^-5), however, there was no significant difference between NC and BC in the M1 (P = 0.17). Finally, the ROC was calculated to evaluate how the risk models discriminated between women with and without breast cancer (Fig. 3). The ROC estimated for the M2 was 62.18%, whereas that for the M1 was only 54.56%. Therefore, the accuracy of M2 in breast cancer risk identification was better than M1.
4. Discussion

Breast cancer (BC) is an estrogen-dependent tumor, and the occurrence of BC is closely related to the imbalance of estrogen homeostasis [21]. The accumulation of estrogen and its toxic metabolites in vivo is a significant risk factor for BC development. Different types of estrogens have different physiological and
pathological activities and can play an important role in the process of cancer development through different mechanisms. Parent estrogens are postulated to promote tumorigenesis directly through the stimulation of estrogen receptor (ER) [21]. The endogenous conversion of estrogen to genotoxic metabolites has been reported as an alternative, potentially ER-independent mechanism for estrogen-dependent breast tumorigenesis [22]. The catechol estrogens can form adducts with DNA, causing gene mutations, and produce direct genotoxicity [23]. Methoxyestrogens, including 2-methoxyestradiol, have been shown to inhibit carcinogenesis by suppressing cell proliferation and estrogen oxidation due to effects on microtubule stabilization [24].

In this study, the LC-MS/MS quantitative analysis method was used to determine the serum estrogens in the BC group and NC group. Comparing the levels of serum estrogens in the follicular phase and luteal phase of premenopausal breast cancer patients with healthy female volunteers, we found that the level of parent and hydroxylated estrogen in the BC group was significantly higher than that of NC, which indicated that estrogens metabolism disorder is closely related to the occurrence and development of breast cancer. Using OPLS-DA analysis, we have also noticed that E1, E2, 4-OHE2, 2-OHE2, and 2/4-OHE1 are BC-related disease markers. This result was consistent with the epidemiologic characteristics of patients with BC [25].

A large number of studies have confirmed that breast cancer existed heritability [26, 27]. However, high-risk genes such as BRCA1 and BRCA2 account for less than 15% of breast cancer cases [28, 29], which suggests that numerous breast cancer-related risk genes have not been discovered, and these gene polymorphisms influence susceptibility to breast cancer.

Estrogen is an important risk factor for breast cancer. However, no research has incorporated estrogens into the breast cancer risk prediction model. The possible main reason is that there is no clinically effective estrogen evaluation method, because the steady-state of estrogen is affected by various physiological and pathological factors such as menstrual cycle fluctuations. However, estrogen homeostasis is regulated by various metabolic enzymes. Therefore, we believe that the estrogen metabolic enzyme gene polymorphisms are closely related to estrogen homeostasis and the occurrence and development of breast cancer. In this study, univariate logistic regression analysis showed that CYP1A1, CYP1B1, and SULT1A1 gene polymorphisms are closely related to the occurrence of breast cancer.

CYP1A1 and CYP1B1 are the major phases I drug metabolism enzymes that catalyze hydroxylation of estrogens. The increasing polarity of estrogens may be related to the risk of breast cancer [30]. Our experiments also verified this view. In this study, we found that the variant allele of CYP1B1 rs1086836 was involved in reducing the risk of breast cancer, and the exact mechanism of the protection of this variant allele was not clear[31], we assumed that the heterozygote model of CYP1B1 rs1086836 (GC vs.GG: OR = 0.37, 95%CI: 0.21–0.67, P = 0.001) may result in decreased function of the CYP1B1 enzyme, reducing the production of 4-hydroxy estrogen and even catechol estrogen-3,4-quinone (CE-3,4-Q) to form adducts with DNA. At the same time, this study also proved that the variant alleles of CYP1A1 rs1048943 (TC vs.TT: OR = 2.37, 95%CI: 1.27–4.43, P = 0.003) and CYP1B1 rs1056827 (AA vs.CC: OR = 6.90, 95%Cl:
1.50-31.76, \( P = 0.001 \)) are closely related to the risk of breast cancer, which is consistent with the most research [32, 33]. The possible reason is that the mutations promote the activity of CYP1A1 and CYP1B1 enzymes to increase the production of hydroxylated estrogens or promote the individual’s susceptibility to estrogen.

SULTs catalyze the sulfate conjugation of a broad range of substrates and play an important role in the metabolism of endogenous and exogenous compounds including thyroid and steroid hormones, neurotransmitters, drugs and procarcinogens [34]. SULTs catalyzes the sulfated metabolism of estrogen (E1 and E2) and its metabolites (such as catechol estrogen) and eliminate the activity of estrogen, by forming the sulfate compounds: sulfated estrogens which can not combine with estrogen receptors (ERs). At the same time, it promotes the rapid excretion of sulfated estrogen from the cells, which can reduce the level of estrogen exposure in the circulation and target tissues. The SULT1A1 rs1042028 is the most widely studied gene polymorphism. Its allelic variation can reduce enzyme activity and thermal stability, resulting in increased estrogen accumulation and increased individual susceptibility to breast cancer [35]. In this study, the heterozygote model of rs1042028 had 2.21 times higher risk of breast cancer than the wild model. It is consistent with the results of multiple studies [36, 37].

Previous studies investigated associations between the PRS of multiple SNPs and breast cancer risk to study the cumulative effect of genes on the disease. Mavaddat et al. constructed a 77-SNP PRS for breast cancer and found a threefold increase in risk when comparing the polygenic scores of the highest 1% and the middle quintiles [38]. Harlid et al. investigated the combined effect of low-penetrant SNPs on breast cancer including ten SNPs and founded that the cumulative effect is strongly correlated with breast cancer [39]. However, most of this research on PRS comes from the Caucasian population sample database. Although Sueta and Chan and others have also conducted similar studies in Asian populations, the evidence is still limited [40, 41]. So far, there has been no relevant report on the establishment of a breast cancer PRS risk prediction model from the perspective of estrogen metabolizing enzymes. Based on this, a multi-gene PRS model including estrogen metabolic enzyme genes SNPs and GWAS-selected SNPs was constructed in this study to evaluate the comprehensive effects of multiple estrogen metabolic enzymes SNPs on breast cancer.

In this study, we evaluated possible relationships between the increased breast cancer risk estrogen metabolic enzyme genes SNPs and GWAS-identified genes SNPs in an Asian population. Among them, the GWAS-identified SNPs were unassociated with breast cancer risk in the per-allele model or dominant model in our study. This finding was inconsistent with previous study [20]. Further, we established PRS model 1 just including GWAS-identified SNPs and PRS model 2 which added estrogen metabolic enzyme genes SNPs on the basis of M1. By calculating the PRS score of each individual under the M1 and M2 PRS models, and performing a t-test analysis on the PRS score of the BC and NC group, we found that the \( P \)-value \( (4.9*10^{-5}) \) of the M2 PRS model was far less than M1 \( (0.17) \). Meanwhile, the ROC (62.18\%) of M2 models was better than the M1 (54.56\%). Therefore, the model constructed by adding estrogen metabolic enzyme genes SNPs has a good ability in breast cancer risk prediction, and the accuracy is greatly improved.
5. Conclusion

Estrogens and the related metabolic enzymes gene polymorphisms are closely related to BC. The model constructed by adding estrogen metabolic enzyme genes SNPs has a good ability in breast cancer risk prediction, and the accuracy is greatly improved comparing PRS model only includes GWAS-identified genes SNPs.

Abbreviations

BC: Breast cancer
BMI: Body mass index
CI: Confidence interval
COMT: Catechol-O-methyltransferase
CYP: Cytochrome P450
E1: Estrone
E2: 17β-estradiol
2-OHE2/1: 2-hydroxy estradiol/estrone
4-OHE2/1: 4-hydroxy estradiol/estrone
16α-OHE1: 16α-hydroxy estrone
2-MeOE2/1: 2-methoxy estradiol/estrone
4-MeOE2/1: 4-methoxy estradiol/estrone
ER: Estrogen receptor
ESI: Electrospray ionization source
FRR: Familial relative risk
GWAS: Genome-wide association studies
HWE: Hardy–Weinberg equilibrium
LC-MS/MS: liquid chromatography-tandem mass spectrometry
M1: PRS model 1
Declarations

Ethics approval and consent to participate

All procedures performed in studies involving human participants were following the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This study is registered on the Clinical Test Public Management Platform (Registration number: ChiCTR1800014658). Informed consent was obtained from all individual participants included in the study.

Consent for publication

Not applicable.

Availability of data and materials

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Competing interest

The authors declare that they have no competing interests.

Funding
This study was supported by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China [No. 18KJA350002]; the Natural Science Foundation of Jiangsu Province [No. BK20181470]; the Science and Technology Foundation of Xuzhou [No. KC18044]; the Natural Science Foundation of China [No. 81403001]; the Six talent peaks project in Jiangsu Province [YY-045]; the Qing Lan Project in Jiangsu Province; the Provincial commission of health and family planning in Jiangsu Province [No. H2017079]; the Natural Science Foundation general project of Jiangsu Province [No. BK20171173]; the Science and Technology planning project of Jiangsu Province [No. BE2019636]. The roles are to provide essential funding to pay personal efforts.

Authors contributions

All authors read and approved the final manuscript. XYZ, YL and FZ contributed to conception and design. FZ, XYZ, and ZXH performed the methodology. FZ, ZXH and YNZ collected experimental data. FZ, ZXH analyzed and explained the data. FZ and XYZ wrote, reviewed, and revised the manuscript. YXX, BZ, MG, and XXY contributed to administrative, technical, and material support. XYZ supervised the research. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Rivera-Franco MM, Leon-Rodriguez E. Delays in Breast Cancer Detection and Treatment in Developing Countries[J]. Breast cancer: basic clinical research. 2018;12:1178223417752677.
2. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually[J]. J Natl Cancer Inst. 1989;81(24):1879–86.
3. Crispo A, D’Aiuto G, De Marco M, et al. Gail model risk factors: impact of adding an extended family history for breast cancer[J]. The breast journal. 2008;14(3):221–7.
4. Bonache S, Gutierrez-Enriquez S, Tenés A, Masas M, Balmaña J, Diez O. Mutation analysis of the BCCIP gene for breast cancer susceptibility in breast/ovarian cancer families. Gynecol Oncol. 2013;131(2):460–3.
5. Chan M, Ji SM, Liaw CS, et al. Association of common genetic variants with breast cancer risk and clinicopathological characteristics in a Chinese population. Breast Cancer Res Treat. 2012;136(1):209–20.
6. Möller S, Mucci LA, Harris JR, 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.
7. Mavaddat N, Pharoah PD, Michailidou K, et al. Prediction of breast cancer risk based on profiling with common genetic variants[J]. Journal of the National Cancer Institute,2015,107(5).
8. Warren Andersen S, Trentham-Dietz A, Gangnon RE, et al. The associations between a polygenic score, reproductive and menstrual risk factors and breast cancer risk[J]. Breast cancer research and
9. Reeves GK, Travis RC, Green J, et al. Incidence of breast cancer and its subtypes in relation to individual and multiple low-penetrance genetic susceptibility loci. JAMA, 2010, 304(4): 426–434.
10. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. Nat Rev Genet. 2019;20(8):467–84.
11. Warner M, Gustafsson JA. On estrogen, cholesterol metabolism, and breast cancer. N Engl J Med. 2014;370(6):572–3.
12. Zhang Y, Gaikwad NW, Olson K, et al. Cytochrome P450 isoforms catalyze formation of catechol estrogen quinones that react with DNA. Metabolism. 2007;56:887–94.
13. Kiruthiga PV, Kannan MR, Saraswathi C, et al. CYP1A1 gene polymorphisms: lack of association with breast cancer susceptibility in the southern region (Madurai) of India. Asian Pac J Cancer Prev. 2011;12:2133–8.
14. Crooke PS, Ritchie MD, Hachey DL, et al. Estrogens, enzyme variants and breast cancer: a risk model. Cancer Epidemiol Biomarkers Prev. 2006;15:1620–9.
15. Ghisari M, Eiberg H, Long M, et al. Polymorphisms in phase I and phase II genes and breast cancer risk and relations to persistent organic pollutant exposure: a case-control study in Inuit women. Environ Health. 2014;13:19.
16. Qiu J, Du Z, Liu J, et al. Association between polymorphisms in estrogen metabolism genes and breast cancer development in Chinese women: A prospective case-control study. Medicine 2018; 97(47):e13337.0.
17. Sangrajrang S, Sato Y, Sakamoto H, et al. Genetic polymorphisms of estrogen metabolizing enzyme and breast cancer risk in Thai women. International journal of cancer. 2009;125(4):837–43.
18. Ghisari M, Long M, Røge DM, et al. Polymorphism in xenobiotic and estrogen metabolizing genes, exposure to perfluorinated compounds and subsequent breast cancer risk: A nested case-control study in the Danish National Birth Cohort. Environmental research 2017, 154:325–333.
19. Zhao F, Wang X, Wang Y, et al. The function of uterine UDP-glucuronosyltransferase 1A8 (UGT1A8) and UDP-glucuronosyltransferase 2B7 (UGT2B7) is involved in endometrial cancer based on estrogen metabolism regulation. Hormones (Athens). 2020;19(3):403–12.
20. Hsieh YC, Tu SH, Su CT, et al. A polygenic risk score for breast cancer risk in a Taiwanese population. Breast Cancer Res Treat. 2017;163(1):131–8.
21. Eliassen AH, Spiegelman D, Xu X, Keefer LK, Veenstra TD, Barbieri RL, et al. Urinary estrogens and estrogen metabolites and subsequent risk of breast cancer among premenopausal women. Can Res. 2012;72:696–706.
22. Newbold RR, Liehr JG. Induction of uterine adenocarcinoma in CD-1 mice by catechol estrogens. Can Res. 2000;60:235–7.
23. Warner M, Gustafsson JA. On estrogen, cholesterol metabolism, and breast cancer. N Engl J Med. 2014;370(6):572–3.
24. Nehal J, Laldmni, Mohamadi A, et al. 2-Methoxyestradiol, a Promising Anticancer Agent. Pharmacotherapy. 2003;23(2):165–72.

25. Sampson JN, Falk RT, Schairer C, et al. Association of Estrogen Metabolism with Breast Cancer Risk in Different Cohorts of Postmenopausal Women. Cancer Res. 2017;77:918–25.

26. Blazer KR, Slavin T, Weitzel JN. Increased reach of genetic cancer risk assessment as a tool for precision management of hereditary breast cancer. JAMA Oncol. 2016;2:723–4.

27. Doherty J, Bonadies DC, Matloff ET. Testing for hereditary breast cancer: panel or targeted testing? experience from a clinical cancer genetics practice. J Genet Counsel. 2015;24:683–7.

28. Bogdanova N, Helbig S, Dork T. Hereditary breast cancer: ever more pieces to the polygenic puzzle. Hered Cancer Clin Pract. 2013;11:12.

29. El Saghir NS, Zgheib NK, Assi HA, et al. BRCA1 and BRCA2 mutations in ethnic Lebanese Arab women with high hereditary risk breast cancer. Oncologist. 2015;20:357–64.

30. Zhang Y, Gaikwad NW, Olson K, et al. Cytochrome P450 isoforms catalyze formation of catechol estrogen quinones that react with DNA. Metabolism 2007;56:887–94. with high hereditary risk breast cancer. Oncologist 2015;20:357–64.

31. Gajjar K, Martin-Hirsch PL, Martin FL. CYP1B1 and hormone-induced cancer. Cancer Lett. 2012;324:13–30.

32. Martínez-Ramírez OC, Pérez-Morales R, Castro C, et al. Polymorphisms of catechol estrogens metabolism pathway genes and breast cancer risk in Mexican women. Breast. 2013;22:335–43.

33. Reding KW, Weiss NS, Chen C, et al. Genetic polymorphisms in the catechol estrogen metabolism pathway and breast cancer risk. Cancer Epidemiol Biomarkers Prev. 2009;18:1461–7.

34. Xiao J, Zheng Y, Zhou Y, Zhang P, Wang J, Shen F, et al. Sulfotransferase SULT1A1 Arg213His polymorphism with cancer risk: a meta-analysis of 53 case-control studies. PLoS One. 2014;9(9):e106774.

35. Nagar S, Walther S, Blanchard RL. Sulfotransferase (SULT) 1A1 polymorphic variants *1, *2, and *3 are associated with altered enzymatic activity, cellular phenotype, and protein degradation. Mol Pharmacol. 2006;69:2084–92.

36. Lee H, Wang Q, Yang F, Tao P, Li H, Huang Y, et al. SULT1A1 Arg213His polymorphism, smoked meat, and breast cancer risk: a case-control study and meta-analysis. DNA Cell Biol. 2012;31(5):688–99.

37. Forat-Yazdi M, Jafari M, Kargar S, Abolbaghaei SM, Nasiri R, et al. Association between SULT1A1 Arg213His (rs9282861) Polymorphism and Risk of Breast Cancer: a Systematic Review and Meta-Analysis. J Res Health Sci. 2017 Oct 14;17(4):e00396.

38. Mavaddat N, Pharoah PD, Michailidou K, et al. Prediction of breast cancer risk based on profiling with common genetic variants. Journal of the National Cancer Institute, 2015, 107(5).

39. Harlid S, Ivarsson MI, Butt S, et al. Combined effect of low-penetrant SNPs on breast cancer risk. British journal of cancer, 2012, 106(2):389–396.

40. Sueta A, Ito H, Kawase T, et al. Genetic risk predictor for breast cancer using a combination of low-penetrance polymorphisms in a Japanese population. Breast cancer research and
41. Chan M, Ji SM, Liaw CS, et al. Association of common genetic variants with breast cancer risk and clinicopathological characteristics in a Chinese population [J]. Breast cancer research and treatment, 2012, 136(1): 209–220.