

# Association of Estrogen-Related Polygenetic Risk Scores with Breast Cancer and Interactions with Alcohol Intake, Early Menarche, and Nulligravida

Sang Shin Song, Suna Kang, Sunmin Park\*

## Abstract

**Backgrounds:** Early menstruation, late menopause, no pregnancy, and genetic factors are known risk factors of the disease, but their effects may differ in Asian and Caucasian women. The purpose of this study was to identify genetic variants of genes related to estrogen signaling in a large city hospital-based cohort and to determine their interactions with lifestyles. **Methods:** This is a case-control study. Three hundred ninety participants diagnosed with breast cancer were compared with 36,290 controls (no cancer) to explore the genetic variants to influence breast cancer risk. Based on GWAS results, the selected genetic variants were subjected to their interactions by generalized multifactor dimensionality reduction (GMDR) analysis. **Results:** Early menstruation (OR=1.55), early menopause (OR=1.70), and no experience of pregnancy (OR=2.86) had a positive association with breast cancer risk ( $P < 0.05$ ). The selected polygenetic risk score (PRS) models included four SNPs and seven SNPs: The four-SNP PRS model included CDH13\_rs12600325, SMYD3\_rs3753686, FGF12\_rs2134635, and ESRRB\_rs10873289, and in the seven-SNP PRS model, ESR1\_rs2046210, estrogen-related receptor gamma (ESRRG)\_rs17043393, and EGFR\_rs6958497 were added into the four-SNP PRS model. Early menstruation, early menopause, and no pregnancy experience interacted with four-SNP PRS. For the participants who had early menstruation and early menopause, high-PRS had an association with a much higher breast cancer risk than the low-PRS in the four-SNP model. However, metabolic parameters, nutrient intakes, and different dietary patterns did not interact with PRS for breast cancer risk. However, alcohol intake interacted with PRS for breast cancer risk (OR=2.33 and 8.07 for mild and moderate alcohol consumption, respectively;  $P = 0.0004$ ). **Conclusion:** Consideration of age at menarche and menopause, pregnancy experience, and alcohol intake may be required to reduce breast cancer risk in women with a high-PRS of genes related to the estrogen signaling pathway.

**Keywords:** Gene-gene interaction- breast cancer- oestrogen receptor- polygenetic risk scores- alcohol

*Asian Pac J Cancer Prev*, 23 (1), 13-24

## Introduction

Breast cancer is the most common cancer and has a lifetime prevalence of about 12% in women in the USA (Harbeck and Gnant, 2017). Endogenous and exogenous factors influence the etiology of breast cancer. Breast epithelial cells grow and function under female hormone control (Harbeck et al., 2019). Estrogen is associated with the mitotic activity of breast epithelial cells, and thus, excessive estrogen exposure increases their excessive proliferation and increased the mutation rates to increase breast cancer risk (Harbeck et al., 2019). The risk factors of breast cancer known to interact with estrogen exposure are age, age at menarche and menopause, pregnancy experience, breastfeeding period, hormone replacement therapy, and obesity (Harkness et al., 2020). However, the association between estrogen exposure and breast cancer development remains unclear.

Genetic factors influence breast cancer development. For example, women with a family history of breast cancer are at greater risk of developing the disease (Chen and Parmigiani, 2007). Genetic factors include rare coding variants in susceptible genes such as breast cancer type 1 susceptibility protein (BRCA)-1 and *BRCA2* and common non-coding variants. *BRCA1* and *BRCA2* mutations have been demonstrated as major genetic factors in Caucasians. However, a recent study showed that *BRCA* mutations are associated with substantial differences in breast cancer risk between Caucasians and Chinese (Bhaskaran et al., 2019). Genetic variants related to estrogen signaling need to be examined to influence breast cancer risk because estrogen exposure is a known risk factor of breast cancer in Asians. The genes involved in estrogen production, degradation, and estrogen receptor signaling have also been studied (Sun et al., 2015), but their associations with genetic variants in breast cancer

Department of Food and Nutrition, Obesity/Diabetes Research Center, Hoseo University, Asan, Korea. \*For Correspondence: [smpark@hoseo.edu](mailto:smpark@hoseo.edu)

need to be examined in Asians. Furthermore, poly-genetic variants have shown to exhibit accumulative effects on breast cancer risk (Mavaddat et al., 2019).

Estrogen secretion is affected by lifestyles and nutrient intakes, which influence breast cancer risk directly or indirectly through obesity. Furthermore, interactions between genetic variants and obesity, hyperglycemia, and lifestyles are known to influence breast cancer risk (Engin, 2017). Obesity is a primary risk factor of breast cancer and is associated with increased estrogen exposure by estrogen overproduction and early menstruation due to excessive aromatization activity in adipose tissues. Obese women also overproduce pro-inflammatory cytokines, and exhibit increased insulin resistance, up-regulation of insulin-like growth factor pathways, and excessive oxidative stress, which all impact breast cancer risk (Engin, 2017). Obesity is associated with excessive energy intakes, high-fat diets, and lack of exercise that are independent risk factors of breast cancer. Furthermore, advanced-glycated endproduct (AGE) level increases caused by hyperglycemia are positively associated with estrogen receptor-positive breast cancer patients, and reductions in AGE levels modulate estrogen receptor- $\alpha$  signaling (Walter et al., 2019). The results suggest that less AGE production mitigates the risk of estrogen-positive breast cancer development, and it has therapeutic potential. AGE generation is mainly associated with blood glucose concentration and is modulated by lifestyle factors. Therefore, obesity and hyperglycemia, directly and indirectly, influence breast cancer risk, and nutrient intake and dietary patterns may interact with genetic factors to influence breast cancer risk.

Here, we hypothesized that polygenetic variants related to the estrogen signaling pathway influence breast cancer risk, and that they interact with metabolic parameters and nutrient intake. This hypothesis was assessed in 36,680 women aged >40 years who participated in a large hospital-based city cohort.

## Materials and Methods

### Participants

During the ten years from 2004 to 2013, a total of 37,116 women aged >40 years participated in the Korean Genome and Epidemiology Study (KoGES), a hospital-based city cohort study organized by the Korean Center for Disease and Control. The present study was approved by the institutional review boards of the Korean National Institute of Health (KBP-2015-055) and Hoseo University (1041231-150811-HR-034-01). Written informed consent was obtained from all participants.

### Criteria of breast cancer

This is a case-control study. Participants were asked for the diagnosis of breast cancer, and if they answered yes, they were allocated to a breast-cancer group as the control group (n=390). As the participants had a history of any cancers other than breast cancer, they were excluded from the study (n=436). The participants in the control group (n=36,290) did not have a history of any cancers.

### Anthropometric and biochemical measurements

The participants provided age, education, income, smoking history, alcohol consumption, and physical activity during a health interview (Liu et al., 2020). Menarche, the age of initial menstruation, was provided. Menopause was defined as no menstruation for the last 12 months, and age at menopause was provided. If the participant had ovariectomy and hysterectomy, the age at the surgery was considered as menopause age. As the participants provided the periods to have hormone replacement therapy, the periods were extended into the menstruation age, although they had menopause. The pregnancy experience was also provided. Education level was categorized into three groups: less than high school, high school, and college or more. Household income (USD/month) was categorized into four groups: very low (<\$1,000), low (\$1,000–2,000), intermediate (\$2,000–4,000), and high (>\$4,000) (Park et al., 2015). Smoking status was divided into three categories: current smoker, past smoker, and never-smoker (Park et al., 2015). Alcohol consumption status was classified into three categories according to average daily alcohol consumption: nondrinker (0 g), mild drinker (0–20 g), and moderate drinker (> 20 g) (Table 2) (Park et al., 2015). The World Health Organization has recommended < 40g for men <20 g for women per day (Daya et al., 2020). The women's recommendation amount was used for the cutoff for daily alcohol intake since the participants did not drink alcohol a lot in the present study.

Body-weight, height, and waist circumference were measured using a standardized procedure (Kim and Han, 2017). Body mass index (BMI) was calculated by dividing weight in kilograms by the height in meters squared. Blood was collected after an overnight fast, and plasma and serum samples were subjected to laboratory testing (Kim and Han, 2017). Fasting serum glucose and blood hemoglobin A1c (HbA1c; glycated hemoglobin) concentrations were determined using a Hitachi 7600 Automatic Analyzer (Hitachi, Tokyo, Japan). Blood pressure was measured on the right arms in a sitting position at heart level using a sphygmomanometer.

### Assessments of food and nutrient intakes using a semi-quantitative food frequency questionnaire (SQFFQ) and dietary pattern analysis

Dietary intakes were estimated using an SQFFQ developed and validated for KoGES (Ahn et al., 2007). This questionnaire requested information regarding the consumption of food items, and the participants completed it. The consumption of 106 food items was calculated by multiplying the frequencies of specified serving sizes by the portion size over the previous year. The intakes of 23 nutrients on the SQFFQ were estimated using CAN Pro 3.0 (Computer-Aided Nutritional Analysis Program), a nutrient database developed by the Korean Nutrition Society (Ahn et al., 2007).

The 106 food items in the SQFFQ were categorized into 29 predefined food groups (Supplemental Table 1). These food groups were included as independent variables during factor analysis conducted to find dietary patterns using the FACTOR procedure. We determined the number

of factors to retain using eigenvalues of  $>1.5$  and 4 dietary factors describing the distinct dietary patterns of participants. The orthogonal rotation procedure (varimax) was applied during the principal component (PCA) analysis (Park and Kang, 2020). Factor-loading values  $\geq 0.40$  were considered to make major contributions to distinct dietary patterns (Supplemental Table 2).

#### *Genotyping and quality control*

Genomic DNA of the participants was extracted from whole blood, and their genotypes were assessed on an Affymetrix Korean Chip (Affymetrix, Santa Clara, CA), and results were made available to scientists from the Center for Genome Science of the Korea National Institute of Health. This chip has been used to study Korean genetic variants and includes prevalent disease-related single nucleotide polymorphisms (SNPs) (Lee and Kim, 2016). Genotyping accuracies were determined by Bayesian Robust Linear Modeling using the Mahalanobis Distance Genotyping Algorithm (Rabbee and Speed, 2006). Genotype analysis from the DNA samples had a genotyping accuracy of  $\geq 98\%$ , a missing genotype call rate of  $<4\%$ , heterozygosity of  $<30\%$ , Hardy-Weinberg equilibrium (HWE) of  $P>0.05$ , and showed no gender bias (Park and Kang, 2020).

#### *Identification of the best models for gene-gene interactions by generalized multifactor dimensionality reduction (GMDR)*

The flow chart used to produce polygenetic risk scores that influence breast cancer risk is presented in Figure 1. Participants were categorized as having (breast-cancer group,  $n=390$ ) or not having any cancer (control group,  $n=36,290$ ). A genome-wide association study (GWAS) was performed to find genetic variants associated with breast cancer risk by the breast-cancer and control study using a  $P$ -value  $< 0.0001$ . The 493 genetic variants were selected from the GWAS results. The names of selected genetic variants were determined using scandb.org. SNPs of the genes that interacted with estrogen signaling were selected using genemania.org. GMDR analysis included the SNPs of the genes involved in estrogen signaling. Linkage disequilibrium (LD) analyses were performed on the selected genetic variants in the same chromosomes using Haploview 4.2 in PLINK. The genetic variants exhibiting strong LD ( $D'>0.4$ ) were excluded in the GMDR analysis because they provided similar information concerning breast cancer risk. The final ten potential genetic variants for the best model generated by GMDR analysis in the same chromosome did not have a strong correlation among the SNPs, determined by LD analysis ( $D'<0.4$ ). The model for gene-gene interactions influencing breast cancer risk was selected by trained balanced accuracy (TRBA), testing balanced accuracy (TEBA), and cross-validation consistency (CVC) from the GMDR analysis. The genetic variants to be selected from the genetic variant-genetic variant interaction were used to calculate the polygenetic risk score (PRS) models.

#### *Statistical analyses*

Statistical analysis was performed using GPLINK

version 2.0 (<http://pngu.mgh.harvard.edu/~purcell/plink>) and SAS version 9.3 (SAS Institute, Cary, NC, USA). Using a GMDR program, the best gene-gene interaction model was selected using  $P$ -values of  $<0.05$  in sign rank tests of TRBA and TEBA after adjusting for age, gender and living area or age, gender, living area, education, income level, and body mass index (Uma Jyothi and Reddy, 2015). Ten-fold cross-validation was also used to check CVC since the sample size was greater than 1000 (Uma Jyothi and Reddy, 2015). In the best PRS models determined by GMDR analysis, the non-risk and risk alleles of each SNP were counted as 0 and 1, respectively (Hong et al., 2018). For example, when the G allele was associated with an increased risk of breast cancer, then TT, GT, GG were given 0, 1, and 2, respectively. The PRS was calculated by summing risk allele scores in each SNP included in the PRS models. PRS obtained using the 4-SNP PRS model were divided into three categories (0-3, 4-5, and  $>5$ ) by tertile, and these are referred to as low-, medium- and high-PRS, respectively. The model containing seven SNPs (the seven-SNP PRS model) was also classified into three categories (0-6, 7-8, and  $>8$ ) by tertile, and they are also referred to as low-, medium-, and high-PRS, respectively. A high-PRS indicated a higher number of risk alleles in the best gene-gene-interaction model. Adjusted odds ratios (ORs) and 95% confidence intervals of the high-PRS were calculated for breast cancer risk, as the reference of the low-PRS in either the four- or seven-SNP PRS models after adjusting covariates to eliminate the effect of covariates. Covariates in the group 1 included age, education, income, residence area, survey year, and body mass index, whereas covariates in the group 2 contained the covariates in the covariate group 1 plus menopause, menopause age, energy intake, menstrual initiation, and experience of pregnancy.

Descriptive statistics of categorical variables, such as gender and lifestyle, were calculated by frequency distributions according to PRS tertiles. Frequency distributions of classification variables were statistically analyzed using the chi-squared test. Means and standard deviations were analyzed for continuous variables according to PRS categories or breast-cancer and control groups, and the significances of differences between PRS categories or breast-cancer/control groups were determined by one-way analysis of variance (ANOVA) after adjustment for two different covariate groups. Multiple comparisons were performed using the Tukey test. The values of each lifestyle parameter were dichotomized into the 'high' or 'low' groups to explore the interactions between PRS and lifestyles. Two-way ANOVA with main effects and an interaction term was conducted to examine interactions between PRS and lifestyles and estrogen exposure parameters after adjusting for covariates.  $P$ -values of  $\leq 0.05$  were considered statistically significant.

## **Results**

#### *General characteristics of the participants according to the presence of breast cancer*

Table 1 describes the demographic and clinical

Table 1. Socioeconomic and Metabolic Characteristics of the Participants According to Breast Cancer

	No breast cancer (n=36,290)	Breast cancer (n=390)	Adjusted OR (95% CI)
Age <sup>1</sup> (years)	52.9±5.5	54.2±5.4***	0.928 (0.711-1.212)
Early menstration <sup>2</sup> (Yes, %)	3061 (13.4)	72 (21.6)***	1.550 (1.077-2.229)*
Late menopausal women <sup>3</sup> (Yes, %)	12,747 (55.8)	131 (39.2)***	0.529 (0.389-0.720)***
Pregnancy experience <sup>4</sup> (No, %)	1,206 (3.29)	27 (6.94)***	2.862 (1.813-4.519)***
BMI <sup>5</sup> (kg/m <sup>2</sup> )	23.6±2.8	23.4±2.9	0.810 (0.891-1.040)
Waist circumference <sup>6</sup>	78.1±8.1	78.0±8.1	0.965 (0.728-1.279)
Plasma total cholesterol <sup>7</sup> (mg/dL)	200±35.8	192±35.6***	0.666 (0.455-0.974)*
Plasma HDL <sup>8</sup> (mg/dL)	56.2±13.1	54.9±13.1	1.102 (0.882-1.376)
Plasma triglyceride <sup>9</sup> (mg/dL)	113±72.9	124±77.5*	1.212 (0.958-1.535)
Type 2 diabetes <sup>10</sup>	7135 (19.4)	83 (21.3)	1.495 (1.023-2.183)*
Blood pressure <sup>11</sup>	7656 (20.9)	73 (18.7)	0.816 (0.618-1.078)
White blood cell counts <sup>12</sup> (10 <sup>9</sup> /L)	5.49±1.45	5.15±1.40***	0.629 (0.471-0.840)**
Plasma hs-CRP <sup>13</sup> (ng/mL)	0.12±0.31	0.16±0.85	1.100 (0.404-2.994)
Education <sup>14</sup> (Number, %)			1
<High school	5,948(20.8)	56(18.6)	1.183 (0.620-2.257)
High school, college	6,520(22.8)	65(21.6)	1.118 (0.435-2.873)
College more	16,092(56.3)	180(59.8)	
Income <sup>15</sup> (Number, %) >2000			1
<\$1000/y	3,977 (11.5)	43 (11.8)	0.938 (0.734-1.199)
\$1000-2000	7,420 (24.4)	86 (23.6)	
\$2000-4000	14,950 (43.2)	151 (41.4)	
>\$4000	8,227 (23.8)	85 (23.3)	

The values represent means ± standard deviations or number of the subjects (percentage of each group); The cutoff points of the reference were as following: <sup>1</sup>< 60 years old for age, <sup>2</sup><14 years old for initial menstration age, <sup>3</sup><50 years old for menopause age, <sup>4</sup>no pregnancy experience, <sup>5</sup>< 25 kg/m<sup>2</sup> BMI, <sup>6</sup>< 90 cm for men and 85 cm for women waist circumferences, <sup>7</sup><230 mg/dL plasma total cholesterol concentrations, <sup>8</sup>>40 mg/dL for men and 50 mg/dL for women plasma HDL cholesterol, <sup>9</sup><150 mg/dL plasma triglyceride concentrations, <sup>10</sup><126 ml/dL fasting serum glucose plus diabetic drug intake, <sup>11</sup><140 mmHg SBP, 90 mmHg DBP plus hypertension medication, <sup>12</sup><5.6X10<sup>9</sup>/L white blood cell counts, <sup>13</sup><0.5 mg/dL serum high sensitive-C-reactive protein (hs-CRP) concentrations, <sup>14</sup>high school graduation and <sup>15</sup><\$2000/month income; Adjusted odds ratio (ORs) after adjusting for covariates including initial menstration age, menopause, pregnancy experience, body mass index, energy intake, education, income, residence are, and survey year in logistic regression models. \* Significant differences by breast cancer at P<0.05, \*\* at P<0.01, \*\*\* P<0.001.

characteristics of the study population, which included 390 women in the breast-cancer group and 36,290 in no cancer (control) group. Age was significantly higher in the breast-cancer group than in the control group. Participants that experienced early menstration, early menopause, or had no experience of pregnancy had higher incidences of breast cancer (Table 1). Participants that had an early initiation of menstration or no pregnancy experience had 1.6- and 2.8-fold greater risks of breast cancer, respectively, and those that experienced late menopause had a 0.53-fold lower the risk. These results strongly suggest that female hormones influence breast cancer risk. BMI and waist circumferences were not found to be associated with breast cancer. Blood pressure and lipid profiles, except plasma total cholesterol concentrations, were not significantly different in the breast-cancer and control groups. Plasma total cholesterol in the breast-cancer group was significantly lower in the breast-cancer group than in the control group (Table 1). However, hyperglycemia was associated with a 1.5-fold increase in breast cancer risk. White blood cell counts, but not serum high-sensitive C-reactive protein (hs-CRP) concentrations, were negatively associated with breast cancer risk. Education and income levels were not significantly different in the

breast-cancer and control groups.

#### *Nutrient intake and dietary patterns of the participants*

Table 2 provides a summary of nutrient intakes and dietary patterns in breast-cancer and non-breast cancer (control) groups. Intakes of fat (P<0.001), alcohol (P<0.05), and coffee (P<0.05) were significantly greater in the breast-cancer group than in the control group. Principle component analysis of dietary patterns showed that the breast-cancer group had a more prudent diet (P<0.01) and less rice-based food intake (P<0.05) than the control group (Table 2). Consumption intakes of a traditional balanced diet or a noodle/meat diet were not significantly different in the breast-cancer and control groups (Table 2), and neither were energy, carbohydrate, protein, or cholesterol intakes, exercise, or smoking (Table 2).

#### *Selection of the model of gene-gene interactions related to estrogen receptor signaling to influence breast cancer risk*

The following ten SNPs involved in estrogen signaling metabolism were selected to influence breast cancer risk; rs17043393 of *ESRRG*, rs3753686 of *Set* and *Mynd* domain-containing protein 3 (*SMYD3*) on chromosome

Table 2. Nutrient Intake and Dietary Patterns of the Participants According to Breast Cancer Presence

	No breast cancer (n=36,290)	Breast cancer (n=390)	Adjusted OR (95% CI)
Energy intake <sup>1</sup> (kcal)	1693±531	1635±516	0.831 (0.670-1.030)
CHO percent intake <sup>2</sup>	70.7±20.8	72.8±20.0***	1.340 (0.920-1.952)
Protein percent intake <sup>3</sup>	13.2±5.8	13.1±5.6**	0.721 (0.503-1.033)
Fat percent intake <sup>4</sup>	13.4±8.7	13.0±8.0**	0.457 (0.332-0.630)***
Cholesterol intake <sup>5</sup>	167±125	157±106	0.799 (0.601-1.061)
Exercise (Number, %)			
No	17,536 (47.8)	166 (42.7)*	1
Yes	19,093 (52.1)	223 (57.3)	1.203 (0.971-1.490)
Smoking (Number, %)			
No	35,458 (96.8)	385 (99.0)	1
Former smoking	439 (1.2)	1 (0.26)	0.301 (0.075-1.216)
Smoking	726 (2.0)	3 (0.77)	0.233 (0.033-1.664)
Alcohol intake (Number, %)			
No	25,941 (70.6)	332 (85.1)***	1
Mild drink (0-20g)	979 (2.7)	5 (1.3)	0.390 (0.124-1.226)
Moderate drink (≥20g)	9,806(26.7)	53 (13.6)	0.378 (0.254-0.563)*
Coffee intake <sup>6</sup> (Number %)			1
Low (<3 g/day)	14,865 (40.5)	208 (53.3)***	0.729 (0.555-0.955)*
Medium (3-16 g/day)	21,560 (58.7)	181 (46.4)	
High (≥16 g/day)	301 (0.82)	1 (0.26)	
Traditional balanced diet <sup>7</sup>	10604 (28.9)	95 (24.4)	0.865 (0.664-1.128)
Prudent diet <sup>7</sup>	14065 (38.3)	179 (45.9)	1.408 (1.117-1.775)**
Noodle/meat diet <sup>7</sup>	10121 (27.6)	90 (23.1)*	0.887 (0.673-1.167)
Rice-based diet <sup>7</sup>	12039 (32.8)	112 (28.7)	0.746 (0.585-0.951)*

The values represent means ± standard deviations or number of the subjects (percentage of each group). The cutoff points of the reference were as following: <sup>1</sup>< estimated energy intake, <sup>2</sup>< 65 energy % carbohydrate (CHO) intake, <sup>3</sup><13 energy % protein intake <sup>4</sup><15 energy % fat intake, and <sup>5</sup><250 mg/day cholesterol intake, <sup>6</sup>< 3 g/day coffee drinking, and <sup>7</sup><70<sup>th</sup> percentile intake of each dietary pattern. Adjusted odds ratio (ORs) after adjusting for covariates including initial menstruation age, menopause, pregnancy experience, body mass index, energy intake, education, income, residence area, and survey year in logistic regression models. \* Significant differences by cataract at P<0.05, \*\* at P<0.01, \*\*\* P<0.001.

1, rs62255841 of thyroid hormone receptor beta (*THRB*), rs2134635 of fibroblast growth factor 12 (*FGF12*) on chromosome 3, rs2046210 of estrogen receptor-1 (*ESR1*) on chromosome 6, rs6958497 of epidermal growth factor receptor (*EGFR*) on chromosome 7, rs57120191 of microcephalin (*MCPHI*) on chromosome 8, rs10873289 of estrogen-related receptor beta (*ESRRB*) on chromosome 14, rs12600325 of H-cadherin (*CDH13*) on chromosome 16, and rs6125040 of nuclear receptor coactivator 3 (*NCOA3*) on chromosome 20; after adjusting for age, gender, residence area, survey year, body mass index, daily energy intake, education, and income (Table 3). Each genetic variant was significantly associated with estrogen signaling metabolism in breast cancer (ORs = 0.65 -2.65, P-value = 3.60XE-06 to 2.77XE-02). Genotype frequency distributions met HWE (P>0.05), and their MAF values were > 0.01 (Table 3).

Interactions between genetic variants related to estrogen receptor signaling were assessed using GMDR. Table 4 shows ten models generated from the ten SNPs. Of these models, the best models were selected using TRBA, TEBA, and CVC values with and without adjusting for the covariates detailed in Table 4. Models including four, seven, and nine genetic variants showed significant

(P<0.05) associations with genetic variant-genetic variant interactions in breast cancer (Table 4). However, a model included four SNPs that had the lowest P-value and a CVC of 10/10. This model included four SNPs (four-SNP PRS model) that were selected from the best model of genetic variant-genetic variant interaction by GMDR analysis. The four-SNP PRS model contained *CDH13\_rs12600325*, *SMYD3\_rs3753686*, *FGF12\_rs2134635*, and *ESRRB\_rs10873289* (Table 4). TRBA, TEBA, and CVC of this model were 0.6035, 0.5488, and 10/10, respectively, after adjustment for age, residence area, and BMI. Furthermore, the model with seven SNPs also met the best model criteria of GMDR analysis, and it could be also applied for estimating breast cancer risk (Table 4). The seven-SNP PRS model included *CDH13\_rs12600325*, *SMYD3\_rs3753686*, *FGF12\_rs2134635*, *ESRRB\_rs10873289*, *ESR1\_rs2046210*, *ESRRG\_rs17043393*, *EGFR\_rs6958497* and their TRBA, TEBA and CVC values were 0.7105, 0.5518, and 10/10, respectively, after adjusting for age, residence area, and BMI (Table 4).

*Association between PRS and breast cancer after adjusting for covariates*

As shown in Figure 1, the adjusted ORs of breast

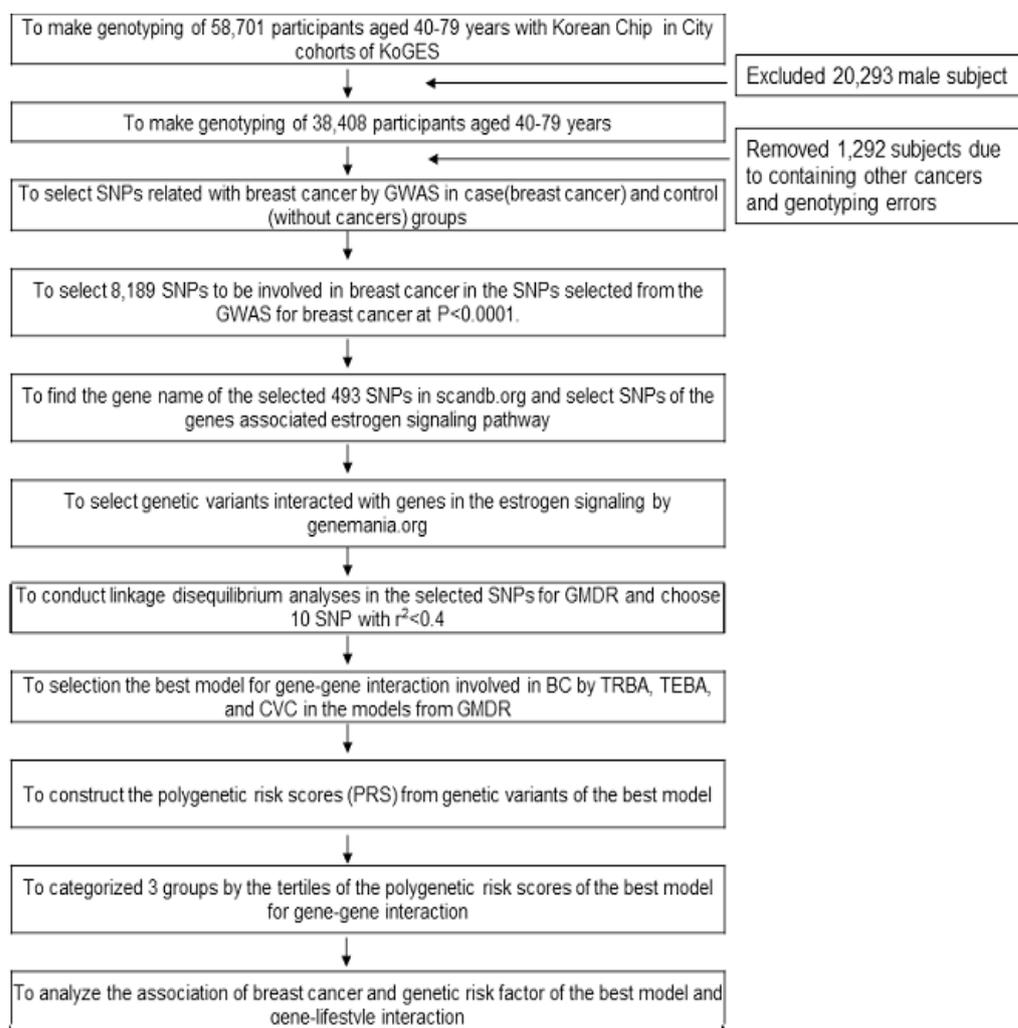


Figure 1. Flow Chart of Haplotypes that Influence Breast Cancer Risk

cancer in the high-PRS group of the four-SNP model were 2.55 and 2.89 ( $P < 0.001$ ), respectively, as compared with subjects in the low-PRS group after adjusting different covariates (covariate groups 1 and 2). This result indicated that the participants in the high-PRS group were at a 2.55- and a 2.89-fold higher risk of breast cancer, respectively than subjects in the low-PRS group. Individuals with a high-PRS group of the seven-SNP model had a higher risk of breast cancer by 2.01- and 2.23-times ( $P < 0.001$ ), respectively, than subjects in the low-PRS group after adjusting covariates model 1 and 2 (Figure 1). However, there was no significant association between the low-PRS and high-PRS groups after adjusting covariates in BMI, waist circumference, type 2 diabetes, hypertension, serum concentrations of total cholesterol, LDL-C, HDL-C, triglyceride, and hs-CRP, and white blood cell counts (Supplemental Table 2).

#### *Interaction between PRS and menstruation and lifestyles in breast cancer*

Age at menarche ( $P = 0.049$ ) and menopause ( $P = 0.004$ ) ages showed significant interactions with the PRS groups of the 4-SNP model for breast cancer risk, and pregnancy experience ( $P = 0.042$ ) also interacted with the PRS groups for breast cancer risk (Table 5). Early menopause and

early initiation of menstruation had a positive association with a high-PRS for breast cancer risk. The participants in early menarche and early menopause with a high-PRS had 2.98 and 3.22 times had higher risks of breast cancer, respectively than those with a low-PRS (Figures 2A, 2B). In having no history of pregnancy, the participants with a high-PRS had a similar risk of breast cancer to those with a low-PRS. However, in having a history of pregnancy, the participants with a high-PRS had a 2.73-fold higher risk of breast cancer than those with a low-PRS (Figure 2C). However, no significant interactions were found between PRS and intakes of energy, protein, carbohydrate, or fat (Table 5) and dietary patterns (data not shown) for the risk of breast cancer. On the other hand, the interaction between alcohol intake and PRS modulated breast cancer risk ( $P = 0.004$ ). In a mild alcohol intake group, the subjects with a high-PRS had a 2.33-fold higher risk of breast cancer than those with a low-PRS. In the moderate alcohol intake group, the participants with a high-PRS had an 8.07-fold higher risk of breast cancer than those with a low-PRS. However, the prevalence of breast cancer was much higher in the participants with mild alcohol intake than in those with moderate alcohol intake, regardless of the PRS (Figure 2D).

Table 4. Generalized Multifactor Dimensionality Reduction (GMDR) Results of Multi-Locus Interaction With Genes Related to Estrogen Signaling Metabolism in Breast Cancer

Model	No Adjusted					Adjusted for age, residence area, bmi				
	TRBA	TEBA	P value	CVC	TRBA	TEBA	P value	CVC		
CDH13_rs12600325	0.5407	0.4997	6 (0.3770)	4/10	0.541	0.4997	6 (0.3770)	4/10		
Model 1 plus SMYD3_rs3753686	0.558	0.5209	6 (0.3770)	5/10	0.558	0.5262	7 (0.1719)	5/10		
Model 2 plus FGF12_rs2134635	0.5791	0.5358	8 (0.0547)	4/10	0.579	0.5349	8 (0.0547)	5/10		
Model 3 plus ESRRB_rs10873289	0.6036	0.5495	10 (0.0010)	10/10	0.604	0.5488	10 (0.0010)	10/10		
Model 4 plus ESR1_rs2046210	0.6354	0.5189	6 (0.3770)	6/10	0.635	0.524	7 (0.1719)	5/10		
Model 5 plus ESRRG_rs17043393	0.6812	0.5323	8 (0.0547)	10/10	0.681	0.5304	8 (0.0547)	10/10		
Model 6 plus EGFR_rs6958497	0.7104	0.5518	9 (0.0107)	10/10	0.711	0.5518	9 (0.0107)	10/10		
Model 7 plus THRB_rs62255841	0.7245	0.5325	8 (0.0547)	9/10	0.725	0.5339	8 (0.0547)	9/10		
Model 8 plus NCOA3_rs6125040	0.7351	0.5463	9 (0.0107)	9/10	0.735	0.5468	9 (0.0107)	9/10		
Model 9 plus MCPH1_rs57120191	0.7444	0.5258	8 (0.0547)	10/10	0.745	0.5267	8 (0.0547)	10/10		

TRBA, trained balanced accuracy; TEBA, test balance accuracy; CVC, cross-validation consistency; sign test, result and P value for the significance of GMDR model by sign test with and without adjusting for covariates designated in the table; BMI, body mass index.

Table 3. The Characteristics of the Ten Genetic Variants of Genes Related to Estrogen Signaling Metabolism in Breast Cancer Used for the Generalized Multifactor Dimensionality Reduction Analysis

Chr <sup>1</sup>	SNP <sup>2</sup>	Position	Mi <sup>3</sup>	Ma <sup>4</sup>	OR and 95% CI	P value adjusted	MAF	HWE	Gene	Functional Consequence	Left gene	right gene
1	rs17043393	216871412	G	A	0.65 (0.52-0.83)	3.76XE-04	0.149	0.435	ESRRG	intron	USH2A	GPATCH2
1	rs3753686	246474033	A	G	1.26 (1.09-1.45)	1.91XE-03	0.366	0.171	SMYD3	intron	KIF26B	LOC100132379
3	rs62255841	24200751	A	G	2.65 (1.75-3.99)	3.60XE-06	0.012	0.731	THRB	intron	NR1D2	LOC644990
3	rs2134635	192275595	G	A	0.74 (0.63-0.86)	1.25XE-04	0.386	0.098	FGF12	intron	LOC100132663	C3orf59
6	rs2046210	151948366	A	G	1.19 (1.02-1.38)	2.50XE-02	0.316	0.474	ESR1	intergenic	C6orf97	ESR1
7	rs6958497	55161746	C	T	0.65 (0.45-0.95)	2.77XE-02	0.055	0.528	EGFR	intron	LOC643168	LOC100133256
8	rs57120191	6347162	C	T	2.46 (1.59-3.82)	6.04XE-05	0.012	0.477	MCPH1	intron	LOC392180	ANGPT2
14	rs10873289	76865501	T	A	0.77 (0.64-0.93)	5.39XE-03	0.234	0.809	ESRRB	intron	C14orf118	LOC100130399
16	rs12600325	83723511	T	C	0.7 (0.59-0.83)	3.51XE-05	0.298	0.159	CDH13	intron	MPHOSPH6	HSBP1
20	rs6125040	46162833	T	A	1.82 (1.12-2.96)	1.64XE-02	0.012	0.124	NCOA3	intron	RPL35AP	SULF2

<sup>1</sup>Chromosome; <sup>2</sup>Single nucleotide polymorphism; <sup>3</sup>Minor allele; <sup>4</sup>Major allele; <sup>5</sup>Odds ratio; <sup>6</sup>P-value for OR and 95% confidence intervals (CI) after adjusting for age, gender, residence area, survey year, body mass index, daily energy intake, education and income; <sup>7</sup>Minor allele frequency; <sup>8</sup>Hardy-Weinberg equilibrium

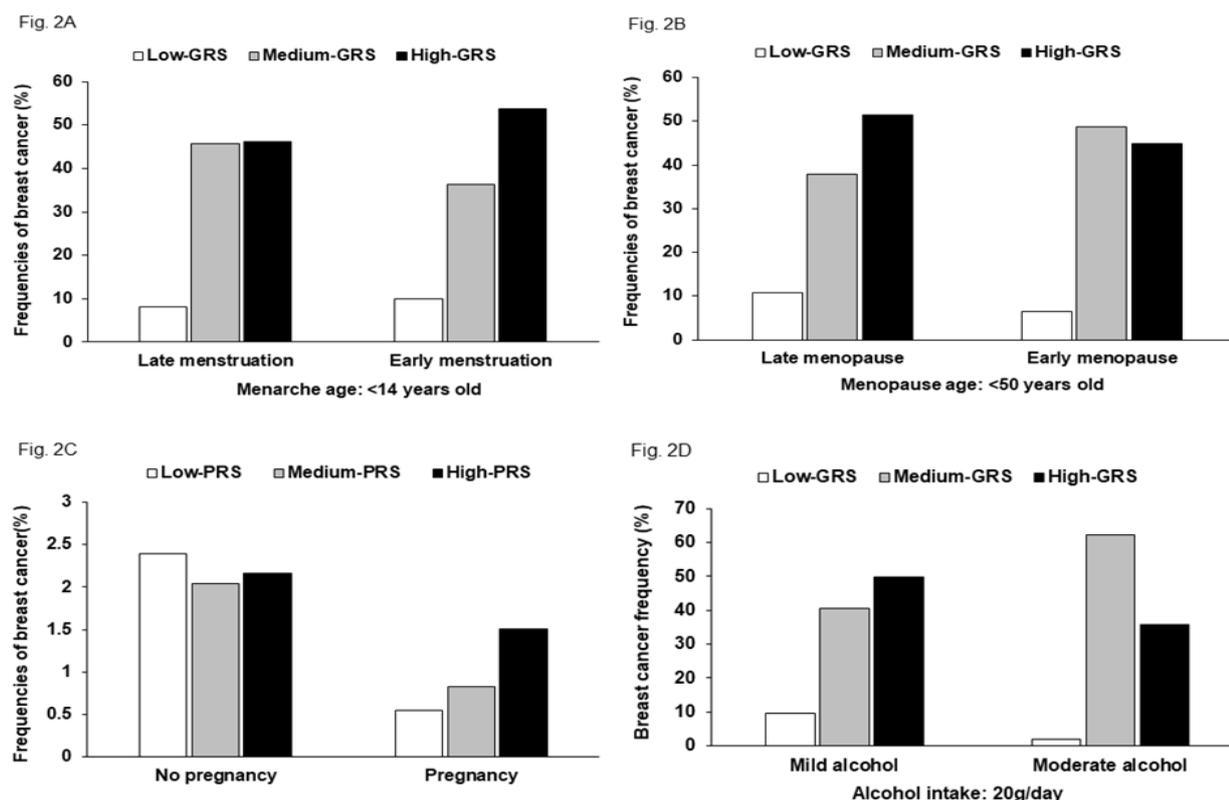


Figure 2. Frequencies of breast cancer in participants categorized with low, medium, or high 4- or 7- SNP PRS groups. A, Participants categorized by initial menstruation age (a cutoff value: 14 years old); B, Participants categorized by age at menopause (a cutoff value: 50 years old); C, Participants categorized by pregnancy experience (a cutoff: no pregnancy); D, Participants categorized by alcohol intake (a cutoff value: 20 g/day); \*\* Significantly different among low and high-PRS groups at  $P < 0.01$  and \*\*\* at  $P < 0.001$

## Discussion

We explored the genetic variants of genes related to estrogen receptor signaling and their influence on breast cancer. GMDR analysis of 493 genetic variants detected in the breast-cancer and the control groups identified ten gene variants related to estrogen signaling. Based on considerations of these ten variants, we identified a model best suited for evaluating gene-gene interactions related to the estrogen receptor. These models included SNP-SNP interactions of four or seven SNPs. The four-SNP model included *CDH13*\_rs12600325, *SMYD3*\_rs3753686, *FGF12*\_rs2134635, and *ESRRB*\_rs10873289 while the seven-SNP model contained the genetic variants in four-SNP model plus *ESR1*\_rs2046210, *ESRRG*\_rs17043393, and *EGFR*\_rs6958497.

Cadherin (*CDH13*) is involved in cell proliferation, growth, and survival and acts as a tumor suppressor gene. Its expression is reduced in human tumors, and it is methylated in breast and lung cancer cells (Yang et al., 2016), and thus, mutations in *CDH13* may induce tumor formation. *SMYD3* is required for estrogen-regulated gene transcription in the estrogen signaling pathway, and its overexpression is associated with the risk of breast cancer (Huang and Xu, 2017).

*FGF12* is a gene with broad mitogenic and cell survival activities and is involved in tumor growth and invasion, and over-methylation of *FGF12* is observed in breast tissue. The genes included in the four-SNP PRS

model are associated with the proliferation of estrogen-sensitive cells via indirectly promoting the estrogen receptor signaling pathway, and conformational changes in these genes by SNPs or methylation could induce breast cancer cell growth (Lian et al., 2012).

Estrogen is involved in various physiological and disease processes, including reproduction, bone remodeling, and breast cancer control. Its effects are transmitted through estrogen receptor- $\alpha$  and - $\beta$ , which are encoded by *ESR1* and *ESR2* and estrogen-related receptors (*ESRR- $\alpha$* , - $\beta$ , and - $\gamma$ ) (Madhu Krishna et al., 2018). These receptors are expressed in estrogen-sensitive organs such as the ovary, breast, and brain. In the present study, *ESRRB*\_rs10873289, *ESR1*\_rs2046210, and *ESRRG*\_rs17043393 were included in the seven-SNP PRS model. However, in the four-SNP PRS model, the only estrogen receptor included was *ESRRB*\_rs10873289. *ESRRB* plays an inhibitory role in cancer cell proliferation in estrogen-dependent cells (Tanida et al., 2015), and *ESRRG* is a nuclear receptor also involved in breast cancer development. In fact, in a study conducted in Thailand, *ESRRG* rs1857407 and rs945453 were found to be associated with breast cancer risk (Sangrajrang et al., 2009). Furthermore, *ESRRG* modulates cell proliferation and estrogen signaling in breast cancer (Ijichi et al., 2011). Recently, *ESR1* rs2046210 was reported to be associated with susceptibility to breast cancer and suggested to be a genetic biomarker in Asians and Caucasians (Campa et al., 2011; Jin et al., 2019), and to play a crucial role in

Table 5. Adjusted Odds Ratios for the Risk of Breast Cancer by Polygenetic Risk Scores of the Best Model (PRS) for Gene-Gene Interaction after Covariate Adjustments According to the Patterns of Lifestyles

	Low- PRS (n=14,420)	Medium-PRS (n=21,641)	High-PRS (n=4,201)	Gene-nutrient interaction P value
Early menstration <sup>1</sup>	1	1.198 (0.521-2.758)	2.976 (1.330-6.658)***	0.049
Late mentration		1.486 (0.942-2.345)	2.402 (1.521-3.795)***	
Early menopause <sup>2</sup>	1	2.096 (1.139-3.956)	3.222 (1.746-5.947)***	0.004
Late menopause		1.049 (0.614-1.792)	2.221 (1.316-3.747)***	
No pregnancy	1	0.846 (0.253-2.835)	1.063 (0.300-3.767)	0.042
Pregnancy		1.492 (0.972-2.288)	2.726 (1.784-4.167)***	
Low energy <sup>3</sup>	1	1.162 (0.728~1.855)	1.955 (1.225-3.118)***	0.561
High energy		2.256 (1.030-4.944)	4.395 (2.018-9.568)***	
Low CHO <sup>4</sup>	1	1.089 (0.626-1.895)	1.839 (1.059-3.196)**	0.608
High CHO		1.830 (1.022-3.277)	3.400 (1.906-6.066)***	
Low protein <sup>5</sup>	1	1.089 (0.687~1.725)	1.919 (1.215~3.030)***	0.415
High protein		2.700 (1.166~6.253)	4.920 (2.131~11.356)***	
Low fat <sup>6</sup>	1	1.264 (0.819~1.952)	2.337 (1.519~3.596)***	0.125
High fat		2.379 (0.846~6.690)	3.864 (1.376~10.854)**	
Low cholesterol <sup>7</sup>	1	1.165 (0.765~1.776)	2.052 (1.349~3.119)***	0.309
High cholesterol		4.797 (1.155~19.92)	9.015 (2.181~37.26)***	
No and mildd alcohol <sup>8</sup>	1	1.186 (0.786~1.791)	2.332 (1.554~3.501)***	0.004
Moderate alcohol		7.998 (1.086~58.88)	8.074 (1.068~60.13)**	

Values represent odd ratios and 95% confidence intervals; PRS with 4 SNPs was divided into 3 categories (0-3, 4-5, and >5) by tertiles as the low, medium and high groups of the best model of GMDR; The cutoff point were as following: <sup>1</sup> <14 years old, <sup>2</sup><50 years old, <sup>3</sup><estimated energy intake, <sup>4</sup>< 65% carbohydrate (CHO) intake, <sup>5</sup>< 13% protein intake, <sup>6</sup>< 15% fat intake, <sup>7</sup>< 250 mg/d cholesterol intake, and <sup>8</sup><20 g/day alcohol intake. Multiple logistic regression models include the corresponding main effects, interaction terms of SNPs and main effects (energy and nutrient intake), and potential confounders such as age, residence area, survey year, body mass index (BMI), education, income, smoking, alcohol, energy, physical activity, fat percent intake, and carbohydrate percent intake; Reference was the low-PRS; \*Significantly different from low-PRS in logistic regression analysis at \* P<0.05, \*\* P<0.01, \*\*\* P<0.001.

the etiology of breast cancer in Chinese, Japanese, and Vietnamese women (Li et al., 2016; Mizoo et al., 2013; Thanh et al., 2018). In the present study, ESR1 rs2046210 was one of the genetic variants in the seven-SNP PRS model. Epidermal growth factor receptor (EGFR) is a receptor tyrosine kinase (RTK), and the first RTK linked with cancer, and the overexpression and mutation of EGFR are observed in many cancers, including breast cancer (Wang, 2017). EGFR overexpression is shown in 15-30% of breast cancers, and EGFR expression is up-regulated by 3~6 fold in breast cancer cells. Estrogen-depleted breast cancer cells are more susceptible to epidermal growth factor mitotic effects, which results in cell death due to the simultaneous blocking of estrogen and EGFR signaling pathways (Jeong et al., 2019).

Polygenetic genetic variants affect breast cancer risk in an additive manner (Mavaddat et al., 2019). In the present study, the adjusted ORs of breast cancer in subjects with high-PRS as determined using the four- and seven-SNP PRS models were 2.89 (P<0.001) and 2.23 (P<0.001), respectively, after adjusting for covariates compared to subjects with a low-PRS. In previous studies, women with polygenetic *BRCAl/2* variants were found to be at elevated risk of breast and ovarian cancers (Barnes et al., 2020), and in a triple-negative breast cancer study, a four-fold risk difference was found between the highest and lowest PRS quintiles (Purrington et al., 2014). These findings suggest that the PRS of the genes related to

estrogen signaling influences breast cancer risk and that this risk is associated with estrogen exposure.

Menstruation-related events are important risk factors of breast cancer; early menarche and late menopause are known to increase the risk of breast cancer, as shown by a meta-analysis conducted on European women (Cancer, 2012) and a review of reproductive factors in breast cancer (Kapil et al., 2014). Since early menarche has reported being associated with childhood obesity and food intake, it can be prevented by modifying lifestyles (Brix et al., 2020; Kim and Lim, 2020). However, breast cancer risk has been reported to increase after menopause in Southeast Asia (Nindrea et al., 2017). It is not clear whether this is a result of comparing early menopause with late menopause. Although breast cancer risk increase doubles every ten years until menopause and then slows after menopause, the incidence of breast cancer remains high after menopause (Momenimovahed and Salehiniya, 2019). Consistent with this previous study (Momenimovahed and Salehiniya, 2019), the current study showed that early menarche increased breast cancer risk. This result might be involved in a longer estrogen exposure to the breast from an immature stage. However, in the present study, early menopause increased breast cancer risk, indicating that the estrogen exposure at an immature stage of the breast may play a critical role in inducing breast cancer more than the extended periods of estrogen exposure. When women may develop breast cancer during menstruation, the

chemotherapy results in menopause since chemotherapy has a 53~89% risk of developing early menopause (Zhou et al., 2015). As previously reported, reproductive period duration may be an indication of estrogen exposure and is associated with the risk of breast and endometrial cancers (Li et al., 2011).

Pregnancy experience and parity are also associated with breast cancer risk. Pregnancy experience and higher parity are associated with a lower risk of breast cancer. During the first pregnancy, breast tissue is either less susceptible to cancer or carcinogenic stimuli to the breast are reduced. Women produce three major estrogens: estrone, estradiol, and estriol. During pregnancy, total estrogen secretion greatly increases, but estriol levels increase significantly more than those of estrone or estradiol. Estriol may be less likely to enhance carcinogenic potential than other estrogens, and thus, younger women with low estriol to estrone plus estradiol ratio have a higher risk of breast cancer, which may reduce breast cancer risk during early pregnancy (Kapil et al., 2014). Many studies have revealed that higher parity is associated with a significantly lower risk of breast cancer (Terry et al., 2018). Furthermore, the present study shows that the risk of breast cancer among women with no experience of pregnancy was not significantly different between those with high- or low-PRS groups, which indicates that no experience of pregnancy offsets genetic impacts on breast cancer development. However, subjects with high-PRS and pregnancy experience had a 2.73 times higher risk of breast cancer than those with low-PRS.

A systematic review of the literature showed that moderate alcohol consumption increased the risks of breast cancer and other cancers (de Menezes et al., 2013). Alcoholic drinks are regarded as carcinogenic by the International Agency for Research on Cancer. Alcohol is metabolized in breast tissue to acetaldehyde, which is associated with the production of reactive oxygen species and DNA damage (Singletary and Gapstur, 2001). Furthermore, alcohol can increase circulating estrogen levels, which is a risk factor for breast cancer. Many recent studies have reported a positive association between alcohol intake and estrogen-positive breast cancer (Romieu et al., 2015). A pooled analysis reported a stronger positive association with alcohol at an occasion exceeding 15g per day (Jung et al., 2016). However, cancer risk associated with alcohol intake may be modulated by genetic factors (Park and Kang, 2020), for example, genetic polymorphisms of alcohol dehydrogenase (ADH) and CYP2E1 have been shown to affect breast cancer risk (Lu et al., 2017). A multitude of genetic factors is also probably associated with altering breast sensitivity to carcinogenic stimuli. The present study also showed a significant difference between high- and low-PRS groups ( $P=0.004$ ). Subjects with high-PRS and mild alcohol intake ( $<20\text{g/day}$ ) were found to have 2.33 times higher risk of breast cancer than those with low-PRS. Furthermore, subjects with high-PRS and moderate alcohol intake ( $\geq 20\text{g/day}$ ) had an 8.07 times higher risk of breast cancer than those with low-PRS. However, energy, carbohydrate, protein, and fat intakes, and four different dietary patterns (data not shown) did not show the interaction with genetic

variants to modulate breast cancer risk.

Since Asians have less association with BRCA1/2 than Caucasians (Bhaskaran et al., 2019) and estrogen exposure is a predominant risk factor for breast cancer in Asians (Sun et al., 2015), the SNPs associated with the estrogen signaling pathway play a critical role in breast cancer risk. The strength of the present study was to use the pooling four or seven genetic variants related to the estrogen signaling pathway to explain the breast cancer risk in Asians. The high-PRS in the four- and seven-SNP model elevated breast cancer risk. Furthermore, the PRS had an interaction with early menarche, early menopause, pregnancy experience, and alcohol intake. The results could be applied to the young persons who have the high-PRS to prevent breast cancer in later life. The limitations of the present study are as follows: 1) Since this study was conducted with the case-control, the results could not be interpreted as the cause-and-effect relation. 2) Subjects were recruited from an urban hospital-based cohort, and thus, our results cannot be extended to the Korean population. 3) Since lifestyles and nutrient intakes were self-reported, the results might exhibit some bias. However, the semi-quantitative food questionnaire used to assess nutrient intakes contained 106 commonly consumed food items and was validated using three day-food records over four seasons by KoGES (Ahn et al., 2007).

In conclusion, the four-SNP PRS model that included *CDH13\_rs12600325*, *SMYD3\_rs3753686*, *FGF12\_rs2134635*, *ESRRB\_rs10873289*, which are all related to estrogen signaling, were found to be positively associated with the risk of breast cancer and to interact with menstruation patterns and pregnancy experience to modulate breast cancer risk. PRS also interacted with alcohol intake to amplify the impact of alcohol on breast cancer risk. The present study cautions that women with a high-PRS should avoid alcohol intake and that lifestyles be changed to delay menarche and reduced the risk of later breast cancer development. The persons who have early menarche need to check the genetic variants in the clinical setting, and their lifestyles can be modified as a prevention measure.

## Author Contribution Statement

authors confirm contribution to the paper as follows: study conception and design: S.S.S, S.P; data collection: S.S.S.; analysis and interpretation of results: S.K., S.P., S.S.S; draft manuscript preparation: S.S.S. All authors reviewed the results and approved the final version of the manuscript.

## Acknowledgments

### *Informed consent*

Written informed consent was obtained from all participants.

### *Compliance with Ethical Standards*

All procedures conducted in the study were under the ethical standards of the institutional review boards of the

Korean National Institute of Health (KBP-2015-055) and Hoseo University (1041231-150811-HR-034-01). The study was processed under the ethical standards of the Helsinki declaration.

#### Funding

This research was supported by the fund from the National Research Foundation of Korea (2019R1A2C1007203).

#### Disclosure of potential conflicts of interest

No authors declare that there are no conflicts of interest.

## References

- Ahn Y, Kwon E, Shim JE, et al (2007). Validation and reproducibility of food frequency questionnaire for Korean genome epidemiologic study. *Eur J Clin Nutr*, **61**, 1435-41.
- Barnes DR, Rookus MA, McGuffog L, et al (2020). Polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of BRCA1 and BRCA2 pathogenic variants. *Genet Med*, **22**, 1653-66.
- Bhaskaran SP, Chandratre K, Gupta H, et al (2019). Germline variation in BRCA1/2 is highly ethnic-specific: Evidence from over 30,000 Chinese hereditary breast and ovarian cancer patients. *Int J Cancer*, **145**, 962-73.
- Brix N, Ernst A, Lauridsen LLB, et al (2020). Childhood overweight and obesity and timing of puberty in boys and girls: cohort and sibling-matched analyses. *Int J Epidemiol*, **49**, 834-44.
- Campa D, Kaaks R, Le Marchand L, et al (2011). Interactions between genetic variants and breast cancer risk factors in the breast and prostate cancer cohort consortium. *J Natl Cancer Inst*, **103**, 1252-63.
- Cancer CGoHFIB (2012). Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *Lancet Oncol*, **13**, 1141-51.
- Chen S, Parmigiani G (2007). Meta-analysis of BRCA1 and BRCA2 penetrance. *J Clin Oncol*, **25**, 1329-33.
- Daya NR, Rebholz CM, Appel LJ, et al (2020). Alcohol consumption and risk of hospitalizations and mortality in the atherosclerosis risk in communities study. *Alcohol Clin Exp Res*, **44**, 1646-57.
- de Menezes RF, Bergmann A, Thuler LC (2013). Alcohol consumption and risk of cancer: a systematic literature review. *Asian Pac J Cancer Prev*, **14**, 4965-72.
- Engin A (2017). Obesity-associated Breast Cancer: Analysis of risk factors. *Adv Exp Med Biol*, **960**, 571-606.
- Harbeck N, Gnant M (2017) Breast cancer. *Lancet*, 389, 1134-50.
- Harbeck N, Penault-Llorca F, Cortes J, et al (2019). Breast cancer. *Nat Rev Dis Primers*, **5**, 66.
- Harkness EF, Astley SM, Evans DG (2020). Risk-based breast cancer screening strategies in women. *Best Pract Res Clin Obstet Gynaecol*, **65**, 3-17.
- Hong KW, Kim SH, Zhang X, et al (2018). Interactions among the variants of insulin-related genes and nutrients increase the risk of type 2 diabetes. *Nutr Res*, **51**, 82-92.
- Huang L, Xu AM (2017) SET and MYND domain-containing protein 3 in cancer. *Am J Transl Res*, **9**, 1-14.
- Ijichi N, Shigekawa T, Ikeda K, et al (2011). Estrogen-related receptor  $\gamma$  modulates cell proliferation and estrogen signaling in breast cancer. *J Steroid Biochem Mol Biol*, **123**, 1-7.
- Jeong Y, Bae SY, You D, et al (2019). EGFR is a therapeutic target in hormone receptor-positive breast cancer. *Cell Physiol Biochem*, **53**, 805-19.
- Jin TF, Zhang WT, Zhou ZF (2019). The 6q25.1 rs2046210 polymorphism is associated with an elevated susceptibility to breast cancer: A meta-analysis of 261,703 subjects. *Mol Genet Genomic Med*, **7**, e553.
- Jung S, Wang M, Anderson K, et al (2016). Alcohol consumption and breast cancer risk by estrogen receptor status: in a pooled analysis of 20 studies. *Int J Epidemiol*, **45**, 916-28.
- Kapil U, Bhadoria AS, Sareen N, et al (2014). Reproductive factors and risk of breast cancer: A Review. *Indian J Cancer*, **51**, 571-6.
- Kim JH, Lim JS (2020). Early menarche and its consequence in Korean female: reducing fructose intake could be one solution. *Clin Exp Pediatr*, DOI: 10.3345/cep.2019.00353.
- Kim Y, Han BG (2017). Cohort Profile: The Korean Genome and Epidemiology Study (KoGES) Consortium. *Int J Epidemiol*, **46**, 1350.
- Lee SK, Kim MK (2016). Relationship of sodium intake with obesity among Korean children and adolescents: Korea National Health and Nutrition Examination Survey. *Br J Nutr*, **115**, 834-41.
- Li G, Xiang YB, Courtney R, et al (2011). Association of single nucleotide polymorphism at 6q25.1,rs2046210, with endometrial cancer risk among Chinese women. *Chin J Cancer*, **30**, 138-43.
- Li X, Yao GY, Li FX, et al (2016). Validation of associations between ESR1 variants and breast cancer risk in Chinese cohorts. *Genet Mol Res*, **15**.
- Lian ZQ, Wang Q, Li WP, et al (2012). Screening of significantly hypermethylated genes in breast cancer using microarray-based methylated-CpG island recovery assay and identification of their expression levels. *Int J Oncol*, **41**, 629-38.
- Liu M, Jin HS, Park S (2020). Protein and fat intake interacts with the haplotype of PTPN11\_rs11066325, RPH3A\_rs886477, and OAS3\_rs2072134 to modulate serum HDL concentrations in middle-aged people. *Clin Nutr*, **39**, 942-9.
- Lu Y, Zhu X, Zhang C, et al (2017). Role of CYP2E1 polymorphisms in breast cancer: a systematic review and meta-analysis. *Cancer Cell Int*, **17**, 11.
- Madhu Krishna B, Chaudhary S, Mishra DR, et al (2018). Estrogen receptor  $\alpha$  dependent regulation of estrogen-related receptor  $\beta$  and its role in the cell cycle in breast cancer. *BMC Cancer*, **18**, 607.
- Mavaddat N, Michailidou K, Dennis J, et al (2019). Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am J Hum Genet*, **104**, 21-34.
- Mizoo T, Taira N, Nishiyama K, et al (2013). Effects of lifestyle and single nucleotide polymorphisms on breast cancer risk: a case-control study in Japanese women. *BMC Cancer*, **13**, 565.
- Momenimovahed Z, Salehiniya H (2019). Epidemiological characteristics of and risk factors for breast cancer in the world. *Breast Cancer*, **11**, 151-64.
- Nindrea RD, Aryandono T, Lazuardi L (2017). Breast cancer risk from modifiable and non-modifiable risk factors among women in Southeast Asia: A Meta-Analysis. *Asian Pac J Cancer Prev*, **18**, 3201-6.
- Park S, Ahn J, Lee BK (2015). Self-rated subjective health status is strongly associated with sociodemographic factors, Lifestyle, Nutrient Intakes, and Biochemical Indices, but Not Smoking Status: KNHANES 2007-2012. *J Korean Med Sci*, **30**, 1279-87.
- Park S, Kang S (2020). A Western-style diet interacts with genetic variants of the LDL receptor to hyper-LDL cholesterolemia in Korean adults. *Public Health Nutr*, **2020**, 1-11.
- Purrington KS, Slager S, Eccles D, et al (2014). Genome-

- wide association study identifies 25 known breast cancer susceptibility loci as risk factors for triple-negative breast cancer. *Carcinogenesis*, **35**, 1012-9.
- Rabbee N, Speed TP (2006). A genotype calling algorithm for Affymetrix SNP arrays. *Bioinformatics*, **22**, 7-12.
- Romieu I, Scoccianti C, Chajès V, et al (2015). Alcohol intake and breast cancer in the European prospective investigation into cancer and nutrition. *Int J Cancer*, **137**, 1921-30.
- Sangrajrang S, Sato Y, Sakamoto H, et al (2009) Genetic polymorphisms of estrogen metabolizing enzyme and breast cancer risk in Thai women. *Int J Cancer*, **125**, 837-43.
- Singletonary KW, Gapstur SM (2001). Alcohol and breast cancer: a review of epidemiologic and experimental evidence and potential mechanisms. *JAMA*, **286**, 2143-51.
- Sun M-Y, Du H-Y, Zhu A-N, et al (2015) Genetic polymorphisms in estrogen-related genes and the risk of breast cancer among Han Chinese women. *Int J Mol Sci*, **16**, 4121-35.
- Tanida T, Matsuda KI, Yamada S, et al (2015). Estrogen-related receptor  $\beta$  reduces the subnuclear mobility of estrogen receptor  $\alpha$  and suppresses estrogen-dependent cellular function. *J Biol Chem*, **290**, 12332-45.
- Terry MB, Liao Y, Kast K, et al (2018). The influence of number and timing of pregnancies on breast cancer risk for women with BRCA1 or BRCA2 mutations. *JNCI Cancer Spectr*, **2**, pky078.
- Thanh NTN, Lan NTT, Phat PT, et al (2018). Two polymorphisms, rs2046210 and rs3803662, are associated with breast cancer risk in a Vietnamese case-control cohort. *Genes Genet Syst*, **93**, 101-9.
- Uma Jyothi K, Reddy BM (2015). Gene-gene and gene-environment interactions in the etiology of type 2 diabetes mellitus in the population of Hyderabad, India. *Meta Gene*, **5**, 9-20.
- Walter KR, Ford ME, Gregoski MJ, et al (2019). Advanced glycation end products are elevated in estrogen receptor-positive breast cancer patients, alter the response to therapy, and can be targeted by lifestyle intervention. *Breast Cancer Res Treat*, **173**, 559-71.
- Wang Z (2017). ErbB Receptors and Cancer. *Methods Mol Biol*, **1652**, 3-35.
- Yang J, Niu H, Huang Y, et al (2016). A systematic analysis of the relationship of CDH13 promoter methylation and breast cancer risk and prognosis. *PLoS One*, **11**, e0149185.
- Zhou Q, Yin W, Du Y, et al (2015). Prognostic impact of chemotherapy-induced amenorrhea on premenopausal breast cancer: a meta-analysis of the literature. *Menopause*, **22**, 1091-7.



This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.