Am J Cancer Res 2015;5(8):2467-2475
www.ajcr.us /ISSN:2156-6976/ajcr0011173

Original Article

Association of five single nucleotide polymorphisms at 6q25.1 with breast cancer risk in northwestern China

Long Zhou[1,2], Na He[3,4], Tian Feng[2], Tingting Geng[2], Tianbo Jin[1,2,3,4], Chao Chen[1,2]

[1]School of Life Sciences, Northwest University, Xi'an, Shaanxi 710069, China; [2]National Engineering Research Center for Miniaturized Detection Systems, Xi'an 710069, China; [3]Key Laboratory of High Altitude Environment and Genes Related to Diseases of Tibet Autonomous Region, School of Medicine, Tibet University for Nationalities, Xianyang 712082, China; [4]Key Laboratory for Basic Life Science Research of Tibet Autonomous Region, School of Medicine, Xizang Minzu University for Nationalities, Xianyang, Shaanxi 712082, China

Received June 8, 2015; Accepted June 17, 2015; Epub July 15, 2015; Published August 1, 2015

Abstract: CCDC170 and *ESR1*, located at 6q25.1, were associated with breast cancer (BC) risk by genome-wide association studies. Our goal was to validate the association between *CCDC170-ESR1* polymorphisms and BC risk in the population of northwestern China. A case-control study of 551 patients with BC and 577 control individuals was conducted from January 2011 to November 2014. We analyzed five BC-associated single nucleotide polymorphisms (SNPs) identified in *CCDC170-ESR1* by previous studies. Logistic regression models were used to derive odds ratios (ORs) and 95% confidence intervals after adjusting for body mass index and age. The minor alleles of rs3757318, rs3734805, and rs2046210 were associated with increased BC risk (OR = 1.30, p = 0.005; OR = 1.28, p = 0.006; OR = 1.20, p = 0.033, respectively) in an allelic model analysis. Those three SNPs had a coincident significant association with increased BC risk in genetic models and stratification analyses. A new haplotype, "CT", was associated with a 1.31-fold increased risk of BC (OR = 1.31, p = 0.006). The "C" allele of rs9383951 was associated with a reduced risk of BC (OR = 0.69, p = 0.048) in estrogen receptor-positive individuals under the log-additive model. Our data provide new evidence of the association between *CCDC170-ESR1* and BC susceptibility in the population of northwestern China.

Keywords: Breast cancer, *CCDC170*, *ESR1*, single nucleotide polymorphism, case-control study

Introduction

Breast cancer (BC) is the most frequently diagnosed cancer and the leading cause of cancer deaths among females. It now represents one in four of all cancers in women. According to GLOBOCAN 2012, 187,213 individuals were diagnosed with BC in China, and 47,984 individuals died; the incidence was ranked first (age-standardized incidence rate = 22.1 per 100,000), and the mortality rate was ranked sixth (age-standardized mortality rate = 5.4 per 100,000) [1, 2]. BC is a complex disease that is associated with environmental and genetic factors. Previous studies have identified multiple genetic loci with important roles in the etiology of BC [3-5].

The estrogen receptor 1 (*ESR1*) locus, which encodes estrogen receptor (ER)α, has been a focus of attention because of the roles of estrogen in determining the risk of BC, osteoporosis, and other diseases [6, 7]. Recent studies reported that the single nucleotide polymorphism (SNP) rs2046210, which is located 180 kb upstream of the transcription start site of the first exon, 6 kb downstream of *CCDC170*, and 29 kb upstream of the first un-translated region of *ESR1*, increased the BC risk in a genome-wide association study of the population of eastern Asia [5, 8]. Further study confirmed that the SNP at 6q25.1 with the lowest p-value (rs3757318; $p = 2.9 \times 10^{-6}$) lies ~200 kb upstream of *ESR1* in an intron of *CCDC170* and found that rs3757318 is only weakly correlated with rs2046210 in Europeans (r^2 = 0.088), although those two SNPs are more strongly correlated in the East Asian population (r^2 = 0.48 in HapMap CHB) [9, 10]. Recent studies have shown that SNPs near *CCDC170*, an open reading frame immediately upstream of *ESR1*, are associated with an increased risk of

Table 1. Basic characteristics of the control individuals and patients with breast cancer

Characteristic	Control (N=577)	Case (N=551)	*p*-value
Mean age ± SD	48.79±8.29 (N=577)	49.09±11.02 (N=551)	0.613[a]
Mean BMI ± SD	22.95±3.21 (N=549)	22.52±2.84 (N=459)	0.027[b]
Clinical stage (TNM)			
1-2		348 (63.2)	
3-4		148 (26.9)	
Missing		55 (9.9)	
Estrogen receptor (ER) status			
ER+		292 (53.0)	
ER-		136 (24.7)	
Missing		123 (22.3)	
Progesterone receptor (PR) status			
PR+		247 (44.8)	
PR-		180 (32.7)	
Missing		124 (22.5)	

SD: Standard deviation. [a]*p* value was calculated by Welch's t-test. [b]*p* value was calculated by Student's t-test. Clinic stage reference: international unifying new TNM classification from Union for International Cancer Control. BMI: Body mass index (weight [kg]/height [m]^2).

BC in the East Asian population. The risk was also present in the western population, in which *CCDC170* was highly correlated with *ESR1* [5, 6]. Later research found that the *CCDC170-ESR1* intergenic region also influenced the progression and survival of Japanese patients with BC [8].

This case-control study of 551 patients with BC and 577 control individuals from northwestern China was conducted to assess the associations between BC risk and *CCDC170-ESR1* polymorphisms located at 6q25.1, which were previously associated with BC susceptibility in some genome-wide association studies of East Asian or European populations.

Materials and methods

Study participants

This study was based on a case-control study of BC conducted in northwestern China. We recruited patients with BC and healthy women into this molecular epidemiological study at the Tang Du Hospital and the First Affiliated Hospital of Xi'an Jiao Tong University from January 2011 to November 2014. The eligible patients were recently diagnosed with histologically confirmed BC. Patients with other cancers who underwent radiotherapy or chemotherapy were excluded. Finally, 551 patients were included in the study and successfully genotyped. We also recruited a random sample of unrelated healthy individuals during the same period from the same hospitals. The control population was matched with the case population based upon age and ethnicity and had no history of cancer. This minimized the presence of factors that could influence the mutation rate, thus maximizing the study's power. Finally, we selected 577 unrelated healthy women to serve as controls. All participants were women at least 18 years of age with good mental condition and no genetic relation going back three generations.

Clinical and demographic data

A standardized epidemiological questionnaire including residential region, age, smoking status, alcohol use, ethnicity, education status, and family history of cancer was used to collect personal data through in-person interviews. For the patients, related information was collected through a consultation with the treating physicians or a medical chart review. BC staging relies on the TNM system designed jointly by the Union International Cancer Control [11]. The main clinical information and demographic characteristics are shown in **Table 1**. The alpha-fetoprotein and plasma carcinoembryonic antigen were tested to make sure that no participants in the control group suffered from any cancers. Venous blood samples (5 ml) and signed informed consent were obtained from each participant. All blood samples were quick-

Table 2. Basic information on the five SNPs and Pearson Chi-square test at 6q25.1 between *CCDC170-ESR1*

SNPs	Chr	Position	Nearby Gene	Distance	A/B	MAF-control	MAF-case	HWE-*p*	OR	95% CI	*p*-χ^2
rs3757318	6	151955806	CCDC170	98886||28215	A/G	0.286	0.247	0.656	1.30	1.08-1.56	0.005*
rs3734805	6	151981043	CCDC170	124123||2978	C/A	0.315	0.300	0.922	1.28	1.08-1.53	0.006*
rs2046210	6	151990059	CCDC170-ESR1	-6038||-180320	T/C	0.351	0.363	0.858	1.20	1.02-1.42	0.033*
rs9383938	6	152029050	CCDC170-ESR1	-45029||-141329	T/G	0.405	0.375	0.860	1.09	0.92-1.29	0.322
rs9383951	6	152337306	ESR1	166927||128793	C/G	0.089	0.105	0.663	0.89	0.67-1.17	0.392

A/B: Minor/major alleles based on the control sample frequencies. HWE: Hardy-Weinberg equilibrium test among controls. *$p < 0.05$ indicates statistical significance. A: Minor alleles. B: Major alleles. SNPs: Single nucleotide polymorphisms. Chr: Chromosome. MAF: Minor allele frequency. OR: Odds ratio. CI: Confidence interval.

ly frozen in liquid nitrogen and stored at -80°C. This study was approved by the ethics committee of Northwest University.

SNP selection and genotyping

Using the HapMap database, five candidate SNPs (rs3757318, rs3734805, rs2046210, rs9383938, and rs9383951) in *CCDC170-ESR1* with minor allele frequencies > 5% in the Asian population and previously published associations with BC were selected by previous studies [12-14]. The phenol-chloroform extraction method was performed to extract genomic DNA from whole blood [15]. The concentration of DNA was measured by spectrometry (DU530 UV/VIS™ spectrophotometer, Beckman Instruments, Fullerton, CA, USA). We used the Sequenom MassARRAY® Assay Design 3.0 Software to design a Multiplexed SNP MassEXTEND® assay [16]. SNP genotyping using the standard protocol recommended by the manufacturer was performed using Sequenom MassARRAY® RS1000. The SequenomTyper 4.0 Software™ was used to perform data management and analyses [16, 17].

Statistical analysis

Differences in demographic characteristics between the cases and controls were evaluated by the Chi-square test (for categorical variables) or the Student's t-test (for continuous variables) [18]. Hardy-Weinberg equilibrium (HWE) of each SNP was tested by goodness-of-fit χ^2 test to compare the expected frequency of genotypes in controls among the controls. All the minor alleles were deemed risk alleles for BC susceptibility. Associations between the genotypes and BC risk were estimated by computing odds ratios (ORs) and 95% confidence intervals (CIs) evaluated by five genetic models (co-dominant, dominant, recessive, over-dominant, log-additive model) using unconditional logistic regression adjusted for body mass index (BMI) and age [19]. Linkage disequilibrium (LD) analysis was done using genotype data from all the subjects. The pattern of LD was analyzed using D' and the haplotypes of the candidate SNPs were analyzed using Haploview v4.2. All statistical tests were two-sided. A p-value ≤ 0.05 was considered statistically significant, and a p-value < 0.1 was considered a possible trend that could be explored further in larger study groups. Statistical analyses of demographic characteristics were performed using Windows software with the SPSS 17.0 statistical packages (SPSS, Chicago, IL). Other statistical analyses were performed using the PLINK software package (version 1.07).

Results

Characteristics of the study participants

This case-control study included 551 patients with BC and 577 control individuals. Student t test was used to determine the differences in age at diagnosis and BMI between case and control groups. There was no significant difference in the age distributions between the cases and controls ($p = 0.613$). The mean age was 49.09 years and 48.79 years for the case and control groups, respectively. Additionally, there was a significant difference in the BMI distributions ($p = 0.027$). The mean BMI was 22.52 and 22.95 for the case and control groups, respectively; so the BMI was adjusted for the logistic regression models. The basic characteristics and population stratification are listed in **Table 1**.

Association between the SNPs and BC risk

All five SNPs conformed to Hardy-Weinberg proportions among the controls ($p > 0.05$). We compared the frequency distributions of the alleles between the cases and the controls by

Table 3. Single loci association with breast cancer risk (logistic regression adjusted by age and BMI)

SNPs	Model	Genotype	Control	Case	OR (95% CI)	*p*-value	AIC	BIC
rs3757318	Co-dominant	G/G	308 (56.5%)	226 (49.6%)	1	0.01*	1375.4	1400
		A/G	207 (38%)	185 (40.6%)	1.21 (0.93-1.58)			
		A/A	30 (5.5%)	45 (9.9%)	2.07 (1.26-3.39)			
	Dominant	G/G	308 (56.5%)	226 (49.6%)	1	0.03*	1377.8	1397.5
		A/G-A/A	237 (43.5%)	230 (50.4%)	1.32 (1.03-1.69)			
	Recessive	G/G-A/G	515 (94.5%)	411 (90.1%)	1	0.008*	1375.4	1395.1
		A/A	30 (5.5%)	45 (9.9%)	1.91 (1.18-3.09)			
	Over-dominant	G/G-A/A	338 (62%)	271 (59.4%)	1	0.44	1381.9	1401.6
		A/G	207 (38%)	185 (40.6%)	1.11 (0.86-1.43)			
	Log-additive	---	---	---	1.33 (1.09-1.62)	0.005*	1374.6	1394.2
rs3734805	Co-dominant	A/A	269 (49%)	196 (42.7%)	1	0.019*	1386.2	1410.8
		C/A	229 (41.7%)	198 (43.1%)	1.18 (0.90-1.54)			
		C/C	51 (9.3%)	65 (14.2%)	1.79 (1.19-2.71)			
	Dominant	A/A	269 (49%)	196 (42.7%)	1	0.047*	1388.1	1407.8
		C/A-C/C	280 (51%)	263 (57.3%)	1.29 (1.00-1.66)			
	Recessive	A/A-C/A	498 (90.7%)	394 (85.8%)	1	0.011*	1385.7	1405.3
		C/C	51 (9.3%)	65 (14.2%)	1.66 (1.12-2.45)			
	Over-dominant	A/A-C/C	320 (58.3%)	261 (56.9%)	1	0.71	1392	1411.6
		C/A	229 (41.7%)	198 (43.1%)	1.05 (0.81-1.35)			
	Log-additive	---	---	---	1.29 (1.07-1.55)	0.008*	1385	1404.6
rs2046210	Co-dominant	C/C	223 (40.6%)	159 (34.6%)	1	0.038*	1387.6	1412.1
		C/T	252 (45.9%)	217 (47.3%)	1.19 (0.90-1.56)			
		T/T	74 (13.5%)	83 (18.1%)	1.63 (1.12-2.37)			
	Dominant	C/C	223 (40.6%)	159 (34.6%)	1	0.056	1388.4	1408.1
		C/T-T/T	326 (59.4%)	300 (65.4%)	1.29 (0.99-1.66)			
	Recessive	C/C-C/T	475 (86.5%)	376 (81.9%)	1	0.025*	1387.1	1406.7
		T/T	74 (13.5%)	83 (18.1%)	1.48 (1.05-2.09)			
	Over-dominant	C/C-T/T	297 (54.1%)	242 (52.7%)	1	0.81	1392	1411.7
		C/T	252 (45.9%)	217 (47.3%)	1.03 (0.80-1.32)			
	Log-additive	---	---	---	1.26 (1.05-1.50)	0.012*	1385.8	1405.5
rs9383938	Co-dominant	G/G	211 (38.4%)	176 (38.6%)	1	0.34	1387.4	1411.9
		G/T	256 (46.6%)	197 (43.2%)	0.92 (0.70-1.21)			
		T/T	82 (14.9%)	83 (18.2%)	1.21 (0.84-1.75)			
	Dominant	G/G	211 (38.4%)	176 (38.6%)	1	0.96	1387.5	1407.2
		G/T-T/T	338 (61.6%)	280 (61.4%)	0.99 (0.77-1.28)			
	Recessive	G/G-G/T	467 (85.1%)	373 (81.8%)	1	0.17	1385.7	1405.3
		T/T	82 (14.9%)	83 (18.2%)	1.26 (0.90-1.77)			
	Over-dominant	G/G-T/T	293 (53.4%)	259 (56.8%)	1	0.29	1386.4	1406
		G/T	256 (46.6%)	197 (43.2%)	0.87 (0.68-1.12)			
	Log-additive	---	---	---	1.06 (0.89-1.27)	0.5	1387.1	1406.7
rs9383951	Co-dominant	G/G	442 (80.5%)	379 (82.8%)	1	0.74	1391.7	1416.3
		G/C	102 (18.6%)	76 (16.6%)	0.89 (0.64-1.23)			
		C/C	5 (0.9%)	3 (0.7%)	0.77 (0.18-3.28)			
	Dominant	G/G	442 (80.5%)	379 (82.8%)	1	0.45	1389.8	1409.4
		G/C-C/C	107 (19.5%)	79 (17.2%)	0.88 (0.64-1.22)			
	Recessive	G/G-G/C	544 (99.1%)	455 (99.3%)	1	0.75	1390.3	1409.9
		C/C	5 (0.9%)	3 (0.7%)	0.79 (0.19-3.35)			
	Over-dominant	G/G-C/C	447 (81.4%)	382 (83.4%)	1	0.48	1389.9	1409.5
		G/C	102 (18.6%)	76 (16.6%)	0.89 (0.64-1.24)			
	Log-additive	---	---	---	0.89 (0.65-1.20)	0.43	1389.7	1409.4

*$p < 0.05$ indicates statistical significance. SNPs: Single nucleotide polymorphisms. OR: Odds ratio. CI: Confidence interval. AIC: Akaike's information criterion. BIC: Bayesian information criterion.

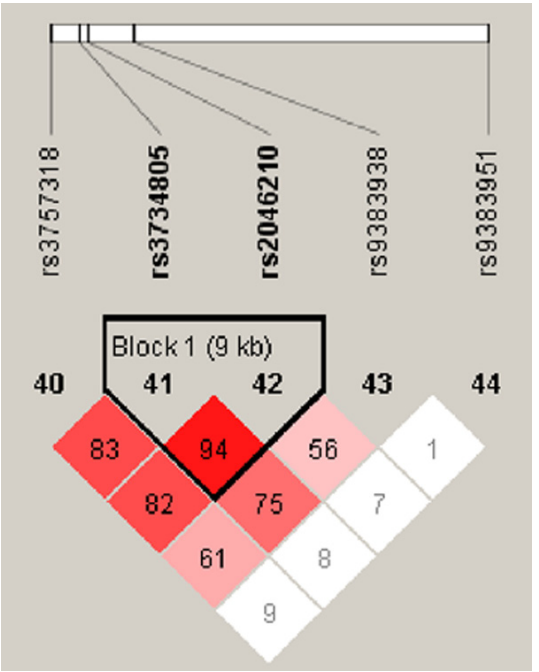

Figure 1. Linkage disequilibrium plots containing five SNPs from 6q21.1. Red squares display statistically significant associations between a pair of SNPs, as measured by D'; darker shades of red indicate higher D'.

Chi-square tests and found that the minor alleles of three SNPs (rs3757318, rs3734805, and rs2046210) were significantly associated with increased BC risk under the allelic model (OR = 1.30, CI = 1.08-1.56, p = 0.005; OR = 1.28, CI = 1.08-1.53, p = 0.006; OR = 1.20, CI = 1.02-1.42, p = 0.033; respectively; **Table 2**).

We further analyzed the association between each SNP and BC risk by unconditional logistic regression analysis using five models (co-dominant model, dominant model, recessive model, over-dominant model, and log-additive model). We found that the risk allele "A" of rs3757318 was associated with an increased risk of BC based on the results of the co-dominant model (OR = 2.07, CI = 1.26-3.39, p = 0.01), the dominant model (OR = 1.32, CI = 1.03-1.69, p = 0.03), the recessive model (OR = 1.91, CI = 1.18-3.09, p = 0.008), and the log-additive model (OR = 1.33, CI = 1.09-1.62, p = 0.005). The minor allele "C" of rs3734805 was associated with an increased BC risk under the co-dominant model (OR = 1.18, CI = 0.90-1.54, p =0.019), the dominant model (OR = 1.29, CI = 1.00-1.66, p = 0.047), the recessive model (OR = 1.66, CI = 1.12-2.45, p = 0.011), and the log-additive model (OR = 1.29, CI = 1.07-1.55, p = 0.008). The minor allele "T" of rs2046210 was also associated with an increased risk of BC based on the results of the co-dominant model (OR = 1.19, CI = 0.90-1.56, p = 0.038), the recessive model (OR = 1.48, CI = 1.05-2.09, p = 0.025), and the log-additive model (OR = 1.26, CI = 1.05-1.50, p = 0.012). No associations were observed under any of the genetic models for rs9383938 and rs9383951 (**Table 3**). Chicago, IL). LD analysis was done using genotype data from all the subjects. One block was detected among the *CCDC170-ESR1* SNPs by haplotype analysis. The block included two SNPs (rs3734805 and rs2046210) with D' = 0.94. The LD between two SNPs is the standardized D' (red schemes) shown in **Figure 1**. By the Chi-square test, we found that haplotypes "AC" (p = 0.032) and "CT" (p = 0.006) were significantly different between the cases and the controls. Additionally, the haplotype "CT" was significantly associated with a 1.31-fold increase in BC risk (OR = 1.31, CI = 1.08-1.58, p = 0.006) by the logistic regression adjusted by BMI and age (**Table 4**).

Subtype analyses based on clinical stage

We analyzed the association between each of the five SNPs and BC risk by unconditional logistic regression by clinical stage subgroups. The results further confirmed that three SNPs (rs3757318, rs3734805, and rs2046210) were associated with an increased BC risk (Table S1). The "CT" haplotype increased the risk of BC in stage 1-2 by 35% (OR = 1.35, CI = 1.09-1.68, p = 0.007) and increased the risk of BC in stage 3-4 by 41% (OR = 1.41, CI = 1.04-1.91, p = 0.026) compared with the "AC" haplotype (Table S2).

Subtype analyses based on estrogen receptor status and progesterone receptor status

Subgroup analyses of rs3757318, rs3734805, rs2046210, rs9383938, and rs9383951 were conducted according to the ER and progesterone receptor (PR) status. Five models were also applied for analyzing the association between polymorphisms and BC, which was adjusted by the age and BMI of the subjects, we found the minor alleles of rs3757318, rs3734805, and rs2046210 increased the BC risk in both the

Table 4. The haplotypes of two SNPs (rs3734805 and rs2046210) and risk of breast cancer (adjusted by age, and BMI)

Haplotype	Freq.	Case-Control Ratio Counts	Case-Control Frequencies	χ^2	p^a	OR (95% CI)	p^b
AC	0.606	653.2:466.8, 731.2:434.8	0.583, 0.627	4.615	0.032*	1	---
CT	0.316	385.2:734.8, 338.2:827.8	0.344, 0.290	7.654	0.006*	1.31 (1.08-1.58)	0.006*
AT	0.068	69.8:1050.2, 84.8:1081.2	0.062, 0.073	0.973	0.324	1.06 (0.75-1.50)	0.74
CC	0.01	11.8:1108.2, 11.8:1154.2	0.011, 0.010	0.012	0.914	0.98 (0.36-2.67)	0.97

SNPs: Single nucleotide polymorphisms. p^a for Chi-square test. OR: Odds ratio. CI: Confidence interval. p^b for logistic regression adjusted by BMI and age. *$p < 0.05$ indicates statistical significance.

Table 5. ER and PR stratified analysis of the associations between each SNP and breast cancer risk (adjusted by BMI and age)

SNP ID	Case (ER-)		Control (ER+)		χ^2	*p*-value	Case (PR-)		Control (PR+)		χ^2	*p*-value
	A count	B count	A count	B count			A count	B count	A count	B count		
rs3757318	89	181	172	406	0.887	0.346	110	248	151	339	0.008	0.978
rs3734805	106	166	207	375	0.925	0.336	130	230	183	311	0.078	0.780
rs2046210	119	153	244	338	0.253	0.615	152	208	211	283	0.020	0.886
rs9383938	116	156	230	348	0.624	0.429	143	213	203	291	0.073	0.787
rs9383951	34	238	42	538	6.302	0.012*	40	320	36	456	3.683	0.055

ER: Estrogen receptor. PR: progesterone receptor. SNP: Single nucleotide polymorphism. BMI: Body mass index (weight [kg]/height [m]^2). *$p < 0.05$ indicates statistical significance. A: Minor alleles. B: Major alleles.

ER subgroup and the PR subgroup. In addition, we found that the minor allele "C" of rs9383951 significantly reduced the risk of BC (OR = 0.69, CI = 0.47-1.00, p = 0.048) under the log-additive model in the ER-positive individuals (Table S3). That phenomenon was not observed, however, in the ER-negative individuals or in the individuals with either PR status (Table S4). We used the Pearson Chi-square method to test whether there was a significant difference within the ER and PR subgroups and found that only rs9383951 was significantly different between the ER-negative and ER-positive patients (p = 0.012; **Table 5**).

Discussion

In this case-control study of samples (551 cases and 577 controls) in drawn from the population of northwestern China, we found that three SNPs (rs3757318, rs3734805, and rs2046210) were associated with an increased risk of BC. Those findings are consistent with the findings of our previous study and genome-wide association study of East Asian women [13, 20]. A previous study found that rs9383951 was significantly associated with reduced BC risk in East Asians [14]. In our study, we found that the minor allele "C" of rs9383951 was significantly associated with a reduced the risk of BC (p = 0.048) in ER-positive patients but not in ER-negative patients, and the allele had no association with PR status, clinic stages, and the overall case-control. That result might be due to differences among different regions of Asia. Another study [21] found that rs3757318 and rs2046210 are in weak LD in the Asian population (r^2 = 0.50) and in the Caucasian population (r^2 = 0.09). Haplotype analyses of those SNPs in the Chinese population found one protective haplotype, "GC", which had a higher frequency in the cases (53%) and in the controls (62%), and that haplotype showed a stronger association with BC in the Chinese population ($p = 8.3 \times 10^{-11}$, OR = 0.69). In our study, an increased risk of BC was observed for the haplotype "CT" involving rs3734805 and rs2046210 (OR = 1.31, p = 0.006), and stratification analysis also found that effect in stage 1-2 (OR = 1.35 p = 0.007) and in stage 3-4 (OR = 1.41 p = 0.026). We therefore conjecture that there is a complex relationship between rs2046210 and BC.

ESR1, which encodes ERα, is located on chromosome 6q and includes seven introns and eight exons over a 140-kb span. Indeed, polymorphisms of the *CCDC170-ESR1* region are correlated with endometriosis, uterine fibroids, BC, and osteoporosis [22-24]. In clinical practice, ER status and PR status remain essential in determining the need for and type of adju-

vant therapy [25]. The prevailing theory is that estrogens increase the rate of cell proliferation by stimulating ER-mediated transcription, thus increasing the number of errors occurring during DNA replication [26].

Despite the current study possessing enough statistical power, some limitations should be considered. Because of our limited sample size, we did not perform further analysis of *HER2* status, *CK5/6* status, and *EGFR* status in the tumor tissues. Although our study included analysis with stratification of the clinical level (stage 1-2 and stage 3-4), we did not have enough samples to analyze each of the four clinical stages separately. In addition, we did not perform a stratified analysis of demographic characteristics including BMI and age. Because the sample was newly collected from January 2011 to November 2014, we did not assess the association between the 5-year survival rate or prognosis factor and the five SNPs. Although we identified significant associations for three SNPs (rs3757318, rs3734805, and rs2046210), the mechanisms are still unclear.

Conclusion

In conclusion, we identified and validated associations between three SNPs (rs3757318, rs3734805, and rs2046210) and BC and between a new haplotype, "CT", of rs3734805 and rs2046210 and BC in the Han population in Xi'an City located in northwestern China. Our data provide additional evidence for the relationships between genetic variants and BC progression. Determining whether there are any biological associations among those SNPs and clinical indicators needs further investigation. The identification and functional characterization of such genes will have a significant impact on BC research and early detection. A major challenge for researchers will be to understand the complicated mechanisms and changes that lead to the development and progression of BC and to apply that knowledge to BC prevention, detection, and treatment. Nevertheless, further studies are needed to clarify the underlying biological mechanisms.

Acknowledgements

This study was funded by the State Project for Essential Drug Research and Development (Grant: 2012ZX09506001-007). We are grateful to the patients and control individuals for their participation in this study. We also thank the clinicians and hospital staff who contributed to the sample and data collection for this study. We would also like to thank BioScience Writers for their assistance in the preparation of this manuscript's grammar.

Disclosure of conflict of interest

None.

Address correspondence to: Drs. Tianbo Jin and Chao Chen, School of Life Sciences, Northwest University, 386, #229 North Taibai Road, Xi'an 710069, Shaanxi, China. Tel: +86-29-88302831; Fax: +86-29-88302831; E-mail: tianbojinprofessor@163.com (TBJ); Tel: +86-29-88303800; Fax: +86-29-88303800; E-mail: cchen898@nwu.edu.cn (CC)

References

[1] Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D and Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015; 136: E359-386.

[2] Bray F, Ren JS, Masuyer E and Ferlay J. Global estimates of cancer prevalence for 27 sites in the adult population in 2008. Int J Cancer 2013; 132: 1133-1145.

[3] Gaudet MM, Milne RL, Cox A, Camp NJ, Goode EL, Humphreys MK, Dunning AM, Morrison J, Giles GG, Severi G, Baglietto L, English DR, Couch FJ, Olson JE, Wang X, Chang-Claude J, Flesch-Janys D, Abbas S, Salazar R, Mannermaa A, Kataja V, Kosma VM, Lindblom A, Margolin S, Heikkinen T, Kampjarvi K, Aaltonen K, Nevanlinna H, Bogdanova N, Coinac I, Schurmann P, Dork T, Bartram CR, Schmutzler RK, Tchatchou S, Burwinkel B, Brauch H, Torres D, Hamann U, Justenhoven C, Ribas G, Arias JI, Benitez J, Bojesen SE, Nordestgaard BG, Flyger HL, Peto J, Fletcher O, Johnson N, Dos Santos Silva I, Fasching PA, Beckmann MW, Strick R, Ekici AB, Broeks A, Schmidt MK, van Leeuwen FE, Van't Veer LJ, Southey MC, Hopper JL, Apicella C, Haiman CA, Henderson BE, Le Marchand L, Kolonel LN, Kristensen V, Grenaker Alnaes G, Hunter DJ, Kraft P, Cox DG, Hankinson SE, Seynaeve C, Vreeswijk MP, Tollenaar RA, Devilee P, Chanock S, Lissowska J, Brinton L, Peplonska B, Czene K, Hall P, Li Y, Liu J, Balasubramanian S, Rafii S, Reed MW, Pooley KA, Conroy D, Baynes C, Kang D, Yoo KY, Noh DY, Ahn SH, Shen CY, Wang HC, Yu JC, Wu PE, Anton-Culver H, Ziogoas A, Egan K,

Newcomb P, Titus-Ernstoff L, Trentham Dietz A, Sigurdson AJ, Alexander BH, Bhatti P, Allen-Brady K, Cannon-Albright LA, Wong J, Australian Ovarian Cancer Study G, Chenevix-Trench G, Spurdle AB, Beesley J, Pharoah PD, Easton DF, Garcia-Closas M and Breast Cancer Association C. Five polymorphisms and breast cancer risk: results from the Breast Cancer Association Consortium. Cancer Epidemiol Biomarkers Prev 2009; 18: 1610-1616.

[4] Sehrawat B, Sridharan M, Ghosh S, Robson P, Cass CE, Mackey JR, Greiner R and Damaraju S. Potential novel candidate polymorphisms identified in genome-wide association study for breast cancer susceptibility. Hum Genet 2011; 130: 529-537.

[5] Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, Wen W, Levy S, Deming SL, Haines JL, Gu K, Fair AM, Cai Q, Lu W and Shu XO. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nat Genet 2009; 41: 324-328.

[6] Dunbier AK, Anderson H, Ghazoui Z, Lopez-Knowles E, Pancholi S, Ribas R, Drury S, Sidhu K, Leary A, Martin LA and Dowsett M. ESR1 is co-expressed with closely adjacent uncharacterised genes spanning a breast cancer susceptibility locus at 6q25.1. PLoS Genet 2011; 7: e1001382.

[7] Ichikawa S, Koller DL, Padgett LR, Lai D, Hui SL, Peacock M, Foroud T and Econs MJ. Replication of previous genome-wide association studies of bone mineral density in premenopausal American women. J Bone Miner Res 2010; 25: 1821-1829.

[8] Yamamoto-Ibusuki M, Yamamoto Y, Fujiwara S, Sueta A, Yamamoto S, Hayashi M, Tomiguchi M, Takeshita T and Iwase H. C6ORF97-ESR1 breast cancer susceptibility locus: influence on progression and survival in breast cancer patients. Eur J Hum Genet 2015; 23: 949-56.

[9] Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, Seal S, Ghoussaini M, Hines S, Healey CS, Hughes D, Warren-Perry M, Tapper W, Eccles D, Evans DG, Breast Cancer Susceptibility C, Hooning M, Schutte M, van den Ouweland A, Houlston R, Ross G, Langford C, Pharoah PD, Stratton MR, Dunning AM, Rahman N and Easton DF. Genome-wide association study identifies five new breast cancer susceptibility loci. Nat Genet 2010; 42: 504-507.

[10] Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, Hankinson SE, Hutchinson A, Wang Z, Yu K, Chatterjee N, Garcia-Closas M, Gonzalez-Bosquet J, Prokunina-Olsson L, Orr N, Willett WC, Colditz GA, Ziegler RG, Berg CD, Buys SS, McCarty CA, Feigelson HS, Calle EE, Thun MJ, Diver R, Prentice R, Jackson R, Kooperberg C, Chlebowski R, Lissowska J, Peplonska B, Brinton LA, Sigurdson A, Doody M, Bhatti P, Alexander BH, Buring J, Lee IM, Vatten LJ, Hveem K, Kumle M, Hayes RB, Tucker M, Gerhard DS, Fraumeni JF Jr, Hoover RN, Chanock SJ and Hunter DJ. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nat Genet 2009; 41: 579-584.

[11] Uehiro N, Horii R, Iwase T, Tanabe M, Sakai T, Morizono H, Kimura K, Iijima K, Miyagi Y, Nishimura S, Makita M, Ito Y and Akiyama F. Validation study of the UICC TNM classification of malignant tumors, seventh edition, in breast cancer. Breast Cancer 2014; 21: 748-753.

[12] Chen W, Song H, Zhong R, Zhu B, Guo H, Lou J, Shen N, Li J, Chen X, Liu C, Ming J, Huang T and Miao X. Risk of GWAS-identified genetic variants for breast cancer in a Chinese population: a multiple interaction analysis. Breast Cancer Res Treat 2013; 142: 637-644.

[13] Fletcher O, Johnson N, Orr N, Hosking FJ, Gibson LJ, Walker K, Zelenika D, Gut I, Heath S, Palles C, Coupland B, Broderick P, Schoemaker M, Jones M, Williamson J, Chilcott-Burns S, Tomczyk K, Simpson G, Jacobs KB, Chanock SJ, Hunter DJ, Tomlinson IP, Swerdlow A, Ashworth A, Ross G, dos Santos Silva I, Lathrop M, Houlston RS and Peto J. Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. J Natl Cancer Inst 2011; 103: 425-435.

[14] Long J, Cai Q, Sung H, Shi J, Zhang B, Choi JY, Wen W, Delahanty RJ, Lu W, Gao YT, Shen H, Park SK, Chen K, Shen CY, Ren Z, Haiman CA, Matsuo K, Kim MK, Khoo US, Iwasaki M, Zheng Y, Xiang YB, Gu K, Rothman N, Wang W, Hu Z, Liu Y, Yoo KY, Noh DY, Han BG, Lee MH, Zheng H, Zhang L, Wu PE, Shieh YL, Chan SY, Wang S, Xie X, Kim SW, Henderson BE, Le Marchand L, Ito H, Kasuga Y, Ahn SH, Kang HS, Chan KY, Iwata H, Tsugane S, Li C, Shu XO, Kang DH and Zheng W. Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer. PLoS Genet 2012; 8: e1002532.

[15] Kochl S, Niederstatter H and Parson W. DNA extraction and quantitation of forensic samples using the phenol-chloroform method and real-time PCR. Methods Mol Biol 2005; 297: 13-30.

[16] Gabriel S, Ziaugra L and Tabbaa D. SNP genotyping using the Sequenom MassARRAY iPLEX platform. Curr Protoc Hum Genet 2009; Chapter 2: Unit 2 12.

[17] Thomas RK, Baker AC, Debiasi RM, Winckler W, Laframboise T, Lin WM, Wang M, Feng W, Zander T, MacConaill L, Lee JC, Nicoletti R, Hatton C, Goyette M, Girard L, Majmudar K, Ziaugra L, Wong KK, Gabriel S, Beroukhim R, Peyton M, Barretina J, Dutt A, Emery C, Greulich H,

Shah K, Sasaki H, Gazdar A, Minna J, Armstrong SA, Mellinghoff IK, Hodi FS, Dranoff G, Mischel PS, Cloughesy TF, Nelson SF, Liau LM, Mertz K, Rubin MA, Moch H, Loda M, Catalona W, Fletcher J, Signoretti S, Kaye F, Anderson KC, Demetri GD, Dummer R, Wagner S, Herlyn M, Sellers WR, Meyerson M and Garraway LA. High-throughput oncogene mutation profiling in human cancer. Nat Genet 2007; 39: 347-351.

[18] Adamec C. [Example of the Use of the Nonparametric Test. Test X2 for Comparison of 2 Independent Examples]. Cesk Zdrav 1964; 12: 613-619.

[19] Bland JM and Altman DG. Statistics notes. The odds ratio. BMJ 2000; 320: 1468.

[20] Mizoo T, Taira N, Nishiyama K, Nogami T, Iwamoto T, Motoki T, Shien T, Matsuoka J, Doihara H, Ishihara S, Kawai H, Kawasaki K, Ishibe Y, Ogasawara Y, Komoike Y and Miyoshi S. Effects of lifestyle and single nucleotide polymorphisms on breast cancer risk: a case-control study in Japanese women. BMC Cancer 2013; 13: 565.

[21] Barzan D, Veldwijk MR, Herskind C, Li Y, Zhang B, Sperk E, Du WD, Zhang XJ and Wenz F. Comparison of genetic variation of breast cancer susceptibility genes in Chinese and German populations. Eur J Hum Genet 2013; 21: 1286-1292.

[22] Lamp M, Peters M, Reinmaa E, Haller-Kikkatalo K, Kaart T, Kadastik U, Karro H, Metspalu A and Salumets A. Polymorphisms in ESR1, ESR2 and HSD17B1 genes are associated with fertility status in endometriosis. Gynecol Endocrinol 2011; 27: 425-433.

[23] Hassan MH, Fouad H, Bahashwan S and Al-Hendy A. Towards non-surgical therapy for uterine fibroids: catechol-O-methyl transferase inhibitor shrinks uterine fibroid lesions in the Eker rat model. Hum Reprod 2011; 26: 3008-3018.

[24] Luo L, Xia W, Nie M, Sun Y, Jiang Y, Zhao J, He S and Xu L. Association of ESR1 and C6orf97 gene polymorphism with osteoporosis in postmenopausal women. Mol Biol Rep 2014; 41: 3235-3243.

[25] Calhoun BC and Collins LC. Predictive markers in breast cancer: An update on ER and HER2 testing and reporting. Semin Diagn Pathol 2015; 32: 362-369.

[26] Yue W, Wang JP, Li Y, Bocchinfuso WP, Korach KS, Devanesan PD, Rogan E, Cavalieri E and Santen RJ. Tamoxifen versus aromatase inhibitors for breast cancer prevention. Clin Cancer Res 2005; 11: 925s-930s.

Table S1. Clinic stage stratified analysis on the associations each SNP with breast cancer risk (adjusted by BMI and age)

SNP	Model	Genotype	Control	Case (stage 1-2)	OR (95% CI)	*p*-value	Control	Case (stage 3-4)	OR (95% CI)	*p*-value
rs3757318	Co-dominant	G/G	308 (56.5%)	146 (48.8%)	1	0.032*	308 (56.5%)	58 (48.7%)	1	0.046*
		A/G	207 (38%)	125 (41.8%)	1.27 (0.94-1.71)		207 (38%)	47 (39.5%)	1.22 (0.80-1.86)	
		A/A	30 (5.5%)	28 (9.4%)	1.98 (1.14-3.45)		30 (5.5%)	14 (11.8%)	2.49 (1.24-4.98)	
	Dominant	G/G	308 (56.5%)	146 (48.8%)	1	0.033*	308 (56.5%)	58 (48.7%)	1	0.11
		A/G-A/A	237 (43.5%)	153 (51.2%)	1.36 (1.02-1.81)		237 (43.5%)	61 (51.3%)	1.38 (0.93-2.05)	
	Recessive	G/G-A/G	515 (94.5%)	271 (90.6%)	1	0.036*	515 (94.5%)	105 (88.2%)	1	0.021*
		A/A	30 (5.5%)	28 (9.4%)	1.79 (1.04-3.06)		30 (5.5%)	14 (11.8%)	2.29 (1.17-4.47)	
	Over-dominant	G/G-A/A	338 (62%)	174 (58.2%)	1	0.29	338 (62%)	72 (60.5%)	1	0.73
		A/G	207 (38%)	125 (41.8%)	1.17 (0.88-1.56)		207 (38%)	47 (39.5%)	1.07 (0.71-1.62)	
	Log-additive	---	---	---	1.34 (1.07-1.68)	0.01*	---	---	1.43 (1.05-1.94)	0.025*
rs3734805	Co-dominant	A/A	269 (49%)	123 (40.7%)	1	0.021*	269 (49%)	49 (41.2%)	1	0.14
		C/A	229 (41.7%)	137 (45.4%)	1.31 (0.97-1.77)		229 (41.7%)	53 (44.5%)	1.28 (0.84-1.97)	
		C/C	51 (9.3%)	42 (13.9%)	1.86 (1.17-2.96)		51 (9.3%)	17 (14.3%)	1.87 (0.99-3.51)	
	Dominant	A/A	269 (49%)	123 (40.7%)	1	0.019*	269 (49%)	49 (41.2%)	1	0.11
		C/A-C/C	280 (51%)	179 (59.3%)	1.41 (1.06-1.87)		280 (51%)	70 (58.8%)	1.39 (0.93-2.08)	
	Recessive	A/A-C/A	498 (90.7%)	260 (86.1%)	1	0.03*	498 (90.7%)	102 (85.7%)	1	0.11
		C/C	51 (9.3%)	42 (13.9%)	1.63 (1.05-2.53)		51 (9.3%)	17 (14.3%)	1.66 (0.92-2.99)	
	Over-dominant	A/A-C/C	320 (58.3%)	165 (54.6%)	1	0.33	320 (58.3%)	66 (55.5%)	1	0.55
		C/A	229 (41.7%)	137 (45.4%)	1.15 (0.87-1.53)		229 (41.7%)	53 (44.5%)	1.13 (0.76-1.69)	
	Log-additive	---	---	---	1.35 (1.09-1.66)	0.0056*	---	---	1.34 (1.00-1.80)	0.05*
rs2046210	Co-dominant	C/C	223 (40.6%)	105 (34.8%)	1	0.081	223 (40.6%)	37 (31.1%)	1	0.077
		C/T	252 (45.9%)	143 (47.4%)	1.20 (0.88-1.63)		252 (45.9%)	59 (49.6%)	1.42 (0.90-2.23)	
		T/T	74 (13.5%)	54 (17.9%)	1.62 (1.06-2.48)		74 (13.5%)	23 (19.3%)	1.92 (1.07-3.46)	
	Dominant	C/C	223 (40.6%)	105 (34.8%)	1	0.086	223 (40.6%)	37 (31.1%)	1	0.046*
		C/T-T/T	326 (59.4%)	197 (65.2%)	1.29 (0.96-1.73)		326 (59.4%)	82 (68.9%)	1.53 (1.00-2.34)	
	Recessive	C/C-C/T	475 (86.5%)	248 (82.1%)	1	0.053	475 (86.5%)	96 (80.7%)	1	0.095
		T/T	74 (13.5%)	54 (17.9%)	1.47 (1.00-2.16)		74 (13.5%)	23 (19.3%)	1.58 (0.94-2.66)	
	Over-dominant	C/C-T/T	297 (54.1%)	159 (52.6%)	1	0.78	297 (54.1%)	60 (50.4%)	1	0.47
		C/T	252 (45.9%)	143 (47.4%)	1.04 (0.79-1.38)		252 (45.9%)	59 (49.6%)	1.16 (0.78-1.73)	
	Log-additive	---	---	---	1.26 (1.03-1.54)	0.027*	---	---	1.39 (1.05-1.85)	0.024*
rs9383938	Co-dominant	G/G	211 (38.4%)	103 (34.2%)	1	0.23	211 (38.4%)	52 (44.1%)	1	0.24
		G/T	256 (46.6%)	141 (46.8%)	1.13 (0.83-1.55)		256 (46.6%)	45 (38.1%)	0.72 (0.46-1.11)	
		T/T	82 (14.9%)	57 (18.9%)	1.44 (0.95-2.17)		82 (14.9%)	21 (17.8%)	1.05 (0.59-1.86)	
	Dominant	G/G	211 (38.4%)	103 (34.2%)	1	0.21	211 (38.4%)	52 (44.1%)	1	0.27
		G/T-T/T	338 (61.6%)	198 (65.8%)	1.21 (0.90-1.62)		338 (61.6%)	66 (55.9%)	0.80 (0.53-1.19)	
	Recessive	G/G-G/T	467 (85.1%)	244 (81.1%)	1	0.13	467 (85.1%)	97 (82.2%)	1	0.43
		T/T	82 (14.9%)	57 (18.9%)	1.34 (0.92-1.94)		82 (14.9%)	21 (17.8%)	1.24 (0.73-2.11)	

rs9383951	Over-dominant	G/G-T/T	293 (53.4%)	160 (53.2%)	1	0.94	293 (53.4%)	73 (61.9%)	1	0.095
		G/T	256 (46.6%)	141 (46.8%)	1.01 (0.76-1.34)		256 (46.6%)	45 (38.1%)	0.71 (0.47-1.07)	
	Log-additive	---	---	---	1.19 (0.97-1.45)	0.095	---	---	0.95 (0.71-1.26)	0.72
	Co-dominant	G/G	442 (80.5%)	245 (81.4%)	1	0.65	442 (80.5%)	102 (85.7%)	1	0.21
		G/C	102 (18.6%)	55 (18.3%)	0.97 (0.67-1.40)		102 (18.6%)	15 (12.6%)	0.63 (0.35-1.13)	
		C/C	5 (0.9%)	1 (0.3%)	0.40 (0.05-3.42)		5 (0.9%)	2 (1.7%)	1.75 (0.33-9.21)	
	Dominant	G/G	442 (80.5%)	245 (81.4%)	1	0.76	442 (80.5%)	102 (85.7%)	1	0.17
		G/C-C/C	107 (19.5%)	56 (18.6%)	0.95 (0.66-1.36)		107 (19.5%)	17 (14.3%)	0.68 (0.39-1.19)	
	Recessive	G/G-G/C	544 (99.1%)	300 (99.7%)	1	0.36	544 (99.1%)	117 (98.3%)	1	0.47
		C/C	5 (0.9%)	1 (0.3%)	0.40 (0.05-3.43)		5 (0.9%)	2 (1.7%)	1.90 (0.36-9.98)	
	Over-dominant	G/G-C/C	447 (81.4%)	246 (81.7%)	1	0.91	447 (81.4%)	104 (87.4%)	1	0.1
		G/C	102 (18.6%)	55 (18.3%)	0.98 (0.68-1.41)		102 (18.6%)	15 (12.6%)	0.63 (0.35-1.12)	
	Log-additive	---	---	---	0.92 (0.65-1.30)	0.65	---	---	0.76 (0.46-1.27)	0.28

SNP: Single nucleotide polymorphism. BMI: Body mass index (weight [kg]/height [m]^2). *$p < 0.05$ indicates statistical significance. OR: Odds Ratio. CI: Confidence interval.

Table S2. Chi-squaretest and logistic regression models to test the risk between haplotype and breast cancer adjusted by BMI and age including two SNPs: 3734805, and rs2046210 on chromosome 6

Haplotype	Case (Stage 1-2) - Control						Case (Stage 3-4) - Control					
	Freq.	Case-Control Frequencies	χ^2	p^a	OR (95% CI)	p^b	Freq.	Case-Control Frequencies	χ^2	p^a	OR (95% CI)	p^b
AC	0.609	0.578, 0.627	4.369	0.037	1	---	0.492	0.482, 0.494	0.132	0.716	1	---
CT	0.316	0.360, 0.290	9.821	0.002	1.35 (1.09-1.68)	0.007*	0.321	0.313, 0.322	0.092	0.761	1.41 (1.04-1.91)	0.026*
AT	0.066	0.054, 0.073	2.511	0.113	0.89 (0.59-1.34)	0.58	0.183	0.194, 0.180	0.284	0.594	1.27 (0.75-2.18)	0.38
**	0.0082	--	--	--	0.79 (0.24-2.63)	0.71	0.008	--	--	--	0.55 (0.07-4.41)	0.57

BMI: Body mass index (weight [kg]/height [m]^2). SNP: Single nucleotide polymorphism. *$p < 0.05$ indicates statistical significance. p^a for Chi-squarer test. p^b for logistic regression models adjusted by BMI and age. **: stands for rare haplotype with frequency < 0.01. OR: Odds ratio. CI: Confidence interval.

Table S3. SNP association with the ER+ and ER- groups (adjusted by age and BMI)

SNP	Model	Genotype	ER+				ER-			
			Control	Case (ER+)	OR (95% CI)	*P*-value	Control	Case (ER-)	OR (95% CI)	*P*-value
rs3757318	Co-dominant	G/G	308 (56.5%)	149 (51.4%)	1	0.022*	308 (56.5%)	58 (43%)	1	0.011*
		A/G	207 (38%)	110 (37.9%)	1.09 (0.80-1.48)		207 (38%)	65 (48.1%)	1.71 (1.15-2.55)	
		A/A	30 (5.5%)	31 (10.7%)	2.16 (1.25-3.70)		30 (5.5%)	12 (8.9%)	2.13 (1.03-4.42)	
	Dominant	G/G	308 (56.5%)	149 (51.4%)	1	0.17	308 (56.5%)	58 (43%)	1	0.003*
		A/G-A/A	237 (43.5%)	141 (48.6%)	1.22 (0.92-1.63)		237 (43.5%)	77 (57%)	1.77 (1.20-2.59)	
	Recessive	G/G-A/G	515 (94.5%)	259 (89.3%)	1	0.007*	515 (94.5%)	123 (91.1%)	1	0.17
		A/A	30 (5.5%)	31 (10.7%)	2.08 (1.23-3.52)		30 (5.5%)	12 (8.9%)	1.66 (0.83-3.35)	
	Over-dominant	G/G-A/A	338 (62%)	180 (62.1%)	1	0.94	338 (62%)	70 (51.9%)	1	0.024*
		A/G	207 (38%)	110 (37.9%)	0.99 (0.74-1.33)		207 (38%)	65 (48.1%)	1.56 (1.06-2.28)	
	Log-additive	---	---	---	1.29 (1.03-1.62)	0.026*	---	---	1.56 (1.16-2.11)	0.004*
rs3734805	Co-dominant	A/A	269 (49%)	125 (42.8%)	1	0.061	269 (49%)	51 (37.5%)	1	0.021*
		C/A	229 (41.7%)	127 (43.5%)	1.18 (0.87-1.60)		229 (41.7%)	64 (47.1%)	1.50 (0.99-2.25)	
		C/C	51 (9.3%)	40 (13.7%)	1.76 (1.10-2.81)		51 (9.3%)	21 (15.4%)	2.18 (1.21-3.95)	
	Dominant	A/A	269 (49%)	125 (42.8%)	1	0.09	269 (49%)	51 (37.5%)	1	0.013*
		C/A-C/C	280 (51%)	167 (57.2%)	1.28 (0.96-1.71)		280 (51%)	85 (62.5%)	1.62 (1.10-2.39)	
	Recessive	A/A-C/A	498 (90.7%)	252 (86.3%)	1	0.034*	498 (90.7%)	115 (84.6%)	1	0.047*
		C/C	51 (9.3%)	40 (13.7%)	1.62 (1.04-2.53)		51 (9.3%)	21 (15.4%)	1.78 (1.03-3.08)	
	Over-dominant	A/A-C/C	320 (58.3%)	165 (56.5%)	1	0.72	320 (58.3%)	72 (52.9%)	1	0.23
		C/A	229 (41.7%)	127 (43.5%)	1.06 (0.79-1.41)		229 (41.7%)	64 (47.1%)	1.26 (0.86-1.84)	
	Log-additive	---	---	---	1.28 (1.03-1.58)	0.024*	---	---	1.48 (1.12-1.96)	0.006*
rs2046210	Co-dominant	C/C	223 (40.6%)	102 (34.9%)	1	0.056	223 (40.6%)	42 (30.9%)	1	0.073
		C/T	252 (45.9%)	136 (46.6%)	1.16 (0.84-1.58)		252 (45.9%)	69 (50.7%)	1.48 (0.97-2.27)	
		T/T	74 (13.5%)	54 (18.5%)	1.69 (1.10-2.59)		74 (13.5%)	25 (18.4%)	1.80 (1.02-3.16)	
	Dominant	C/C	223 (40.6%)	102 (34.9%)	1	0.11	223 (40.6%)	42 (30.9%)	1	0.029*
		C/T-T/T	326 (59.4%)	190 (65.1%)	1.27 (0.95-1.71)		326 (59.4%)	94 (69.1%)	1.55 (1.04-2.33)	
	Recessive	C/C-C/T	475 (86.5%)	238 (81.5%)	1	0.026*	475 (86.5%)	111 (81.6%)	1	0.17
		T/T	74 (13.5%)	54 (18.5%)	1.56 (1.06-2.30)		74 (13.5%)	25 (18.4%)	1.43 (0.87-2.37)	
	Over-dominant	C/C-T/T	297 (54.1%)	156 (53.4%)	1	0.96	297 (54.1%)	67 (49.3%)	1	0.26
		C/T	252 (45.9%)	136 (46.6%)	0.99 (0.74-1.32)		252 (45.9%)	69 (50.7%)	1.24 (0.85-1.81)	
	Log-additive	---	---	---	1.27 (1.03-1.56)	0.023*	---	---	1.36 (1.04-1.79)	0.026*
rs9383938	Co-dominant	G/G	211 (38.4%)	115 (39.7%)	1	0.21	211 (38.4%)	44 (32.4%)	1	0.34
		G/T	256 (46.6%)	120 (41.4%)	0.86 (0.63-1.18)		256 (46.6%)	68 (50%)	1.29 (0.84-1.96)	
		T/T	82 (14.9%)	55 (19%)	1.23 (0.81-1.86)		82 (14.9%)	24 (17.6%)	1.45 (0.83-2.54)	
	Dominant	G/G	211 (38.4%)	115 (39.7%)	1	0.72	211 (38.4%)	44 (32.4%)	1	0.16
		G/T-T/T	338 (61.6%)	175 (60.3%)	0.95 (0.71-1.27)		338 (61.6%)	92 (67.7%)	1.33 (0.89-1.98)	

	Recessive	G/G-G/T	467 (85.1%)	235 (81%)	1	0.14	467 (85.1%)	112 (82.3%)	1	0.39
		T/T	82 (14.9%)	55 (19%)	1.34 (0.91-1.95)		82 (14.9%)	24 (17.6%)	1.25 (0.76-2.07)	
	Over-dominant	G/G-T/T	293 (53.4%)	170 (58.6%)	1	0.14	293 (53.4%)	68 (50%)	1	0.48
		G/T	256 (46.6%)	120 (41.4%)	0.81 (0.60-1.07)		256 (46.6%)	68 (50%)	1.15 (0.79-1.67)	
	Log-additive	---	---	---	1.06 (0.86-1.29)	0.59	---	---	1.22 (0.93-1.60)	0.15
rs9383951	Co-dominant	G/G	442 (80.5%)	249 (85.6%)	1	0.047*	442 (80.5%)	105 (77.2%)	1	0.5
		G/C	102 (18.6%)	42 (14.4%)	0.74 (0.50-1.09)		102 (18.6%)	28 (20.6%)	1.10 (0.69-1.77)	
		C/C	5 (0.9%)	0 (0%)	0.00 (0.00-NA)		5 (0.9%)	3 (2.2%)	2.39 (0.56-10.23)	
	Dominant	G/G	442 (80.5%)	249 (85.6%)	1	0.078	442 (80.5%)	105 (77.2%)	1	0.52
		G/C-C/C	107 (19.5%)	42 (14.4%)	0.71 (0.48-1.05)		107 (19.5%)	31 (22.8%)	1.16 (0.74-1.84)	
	Recessive	G/G-G/C	544 (99.1%)	291 (100%)	1	0.052	544 (99.1%)	133 (97.8%)	1	0.27
		C/C	5 (0.9%)	0 (0%)	0.00 (0.00-NA)		5 (0.9%)	3 (2.2%)	2.34 (0.55-9.98)	
	Over-dominant	G/G-C/C	447 (81.4%)	249 (85.6%)	1	0.14	447 (81.4%)	108 (79.4%)	1	0.74
		G/C	102 (18.6%)	42 (14.4%)	0.75 (0.50-1.11)		102 (18.6%)	28 (20.6%)	1.09 (0.68-1.74)	
	Log-additive	---	---	---	0.69 (0.47-1.00)	0.048*	---	---	1.21 (0.80-1.83)	0.38

SNP: Single nucleotide polymorphism. ER: Estrogen receptor. BMI: Body mass index (weight [kg]/height [m]^2). *$p < 0.05$ indicates statistical significance. OR: Odds ratio; CI: Confidence interval.

Table S4. SNP association with the PR+ and PR- groups (adjusted by age and BMI)

SNP	Model	PR+					PR-			
		Genotype	Control	Case (PR+)	OR (95% CI)	*p*-value	Control	Case (PR-)	OR (95% CI)	*p*-value
rs3757318	Co-dominant	G/G	308 (56.5%)	121 (49.4%)	1	0.016*	308 (56.5%)	85 (47.5%)	1	0.065
		A/G	207 (38%)	97 (39.6%)	1.18 (0.85-1.63)		207 (38%)	78 (43.6%)	1.38 (0.97-1.97)	
		A/A	30 (5.5%)	27 (11%)	2.30 (1.31-4.04)		30 (5.5%)	16 (8.9%)	1.91 (0.99-3.67)	
	Dominant	G/G	308 (56.5%)	121 (49.4%)	1	0.074	308 (56.5%)	85 (47.5%)	1	0.033*
		A/G-A/A	237 (43.5%)	124 (50.6%)	1.32 (0.97-1.79)		237 (43.5%)	94 (52.5%)	1.45 (1.03-2.03)	
	Recessive	G/G-A/G	515 (94.5%)	218 (89%)	1	0.007*	515 (94.5%)	163 (91.1%)	1	0.13
		A/A	30 (5.5%)	27 (11%)	2.14 (1.24-3.71)		30 (5.5%)	16 (8.9%)	1.66 (0.88-3.12)	
	Over-dominant	G/G-A/A	338 (62%)	148 (60.4%)	1	0.73	338 (62%)	101 (56.4%)	1	0.17
		A/G	207 (38%)	97 (39.6%)	1.06 (0.77-1.44)		207 (38%)	78 (43.6%)	1.28 (0.91-1.80)	
	Log-additive	---	---	---	1.36 (1.08-1.73)	0.011*	---	---	1.38 (1.06-1.81)	0.019*
rs3734805	Co-dominant	A/A	269 (49%)	100 (40.5%)	1	0.021*	269 (49%)	75 (41.7%)	1	0.11
		C/A	229 (41.7%)	111 (44.9%)	1.28 (0.93-1.78)		229 (41.7%)	80 (44.4%)	1.26 (0.88-1.81)	
		C/C	51 (9.3%)	36 (14.6%)	1.97 (1.21-3.22)		51 (9.3%)	25 (13.9%)	1.76 (1.02-3.05)	
	Dominant	A/A	269 (49%)	100 (40.5%)	1	0.028*	269 (49%)	75 (41.7%)	1	0.083
		C/A-C/C	280 (51%)	147 (59.5%)	1.41 (1.04-1.91)		280 (51%)	105 (58.3%)	1.35 (0.96-1.90)	
	Recessive	A/A-C/A	498 (90.7%)	211 (85.4%)	1	0.019*	498 (90.7%)	155 (86.1%)	1	0.09
		C/C	51 (9.3%)	36 (14.6%)	1.75 (1.10-2.77)		51 (9.3%)	25 (13.9%)	1.58 (0.94-2.64)	

	Over-dominant	A/A-C/C	320 (58.3%)	136 (55.1%)	1	0.49	320 (58.3%)	100 (55.6%)	1	0.5
		C/A	229 (41.7%)	111 (44.9%)	1.11 (0.82-1.51)		229 (41.7%)	80 (44.4%)	1.12 (0.80-1.58)	
	Log-additive	---	---	---	1.37 (1.09-1.71)	0.006*	---	---	1.31 (1.02-1.68)	0.037*
rs2046210	Co-dominant	C/C	223 (40.6%)	83 (33.6%)	1	0.038*	223 (40.6%)	60 (33.3%)	1	0.14
		C/T	252 (45.9%)	117 (47.4%)	1.22 (0.87-1.71)		252 (45.9%)	88 (48.9%)	1.31 (0.90-1.90)	
		T/T	74 (13.5%)	47 (19%)	1.80 (1.15-2.82)		74 (13.5%)	32 (17.8%)	1.61 (0.97-2.68)	
	Dominant	C/C	223 (40.6%)	83 (33.6%)	1	0.065	223 (40.6%)	60 (33.3%)	1	0.075
		C/T-T/T	326 (59.4%)	164 (66.4%)	1.34 (0.98-1.85)		326 (59.4%)	120 (66.7%)	1.38 (0.96-1.96)	
	Recessive	C/C-C/T	475 (86.5%)	200 (81%)	1	0.022*	475 (86.5%)	148 (82.2%)	1	0.17
		T/T	74 (13.5%)	47 (19%)	1.62 (1.08-2.43)		74 (13.5%)	32 (17.8%)	1.39 (0.88-2.20)	
	Over-dominant	C/C-T/T	297 (54.1%)	130 (52.6%)	1	0.9	297 (54.1%)	92 (51.1%)	1	0.46
		C/T	252 (45.9%)	117 (47.4%)	1.02 (0.75-1.38)		252 (45.9%)	88 (48.9%)	1.14 (0.81-1.59)	
	Log-additive	---	---	---	1.32 (1.06-1.64)	0.013*	---	---	1.28 (1.00-1.63)	0.05*
rs9383938	Co-dominant	G/G	211 (38.4%)	90 (36.4%)	1	0.44	211 (38.4%)	68 (38.2%)	1	0.45
		G/T	256 (46.6%)	111 (44.9%)	1.01 (0.72-1.41)		256 (46.6%)	77 (43.3%)	0.93 (0.64-1.36)	
		T/T	82 (14.9%)	46 (18.6%)	1.31 (0.84-2.03)		82 (14.9%)	33 (18.5%)	1.27 (0.78-2.08)	
	Dominant	G/G	211 (38.4%)	90 (36.4%)	1	0.63	211 (38.4%)	68 (38.2%)	1	0.93
		G/T-T/T	338 (61.6%)	157 (63.6%)	1.08 (0.79-1.48)		338 (61.6%)	110 (61.8%)	1.02 (0.72-1.44)	
	Recessive	G/G-G/T	467 (85.1%)	201 (81.4%)	1	0.2	467 (85.1%)	145 (81.5%)	1	0.23
		T/T	82 (14.9%)	46 (18.6%)	1.30 (0.87-1.94)		82 (14.9%)	33 (18.5%)	1.32 (0.85-2.07)	
	Over-dominant	G/G-T/T	293 (53.4%)	136 (55.1%)	1	0.63	293 (53.4%)	101 (56.7%)	1	0.42
		G/T	256 (46.6%)	111 (44.9%)	0.93 (0.69-1.26)		256 (46.6%)	77 (43.3%)	0.87 (0.62-1.22)	
	Log-additive	---	---	---	1.12 (0.90-1.39)	0.31	---	---	1.09 (0.86-1.39)	0.49
rs9383951	Co-dominant	G/G	442 (80.5%)	210 (85.4%)	1	0.083	442 (80.5%)	143 (79.4%)	1	0.75
		G/C	102 (18.6%)	36 (14.6%)	0.76 (0.50-1.15)		102 (18.6%)	34 (18.9%)	0.99 (0.64-1.53)	
		C/C	5 (0.9%)	0 (0%)	0.00 (0.00-NA)		5 (0.9%)	3 (1.7%)	1.78 (0.42-7.57)	
	Dominant	G/G	442 (80.5%)	210 (85.4%)	1	0.12	442 (80.5%)	143 (79.4%)	1	0.89
		G/C-C/C	107 (19.5%)	36 (14.6%)	0.73 (0.48-1.10)		107 (19.5%)	37 (20.6%)	1.03 (0.68-1.57)	
	Recessive	G/G-G/C	544 (99.1%)	246 (100%)	1	0.072	544 (99.1%)	177 (98.3%)	1	0.45
		C/C	5 (0.9%)	0 (0%)	0.00 (0.00-NA)		5 (0.9%)	3 (1.7%)	1.78 (0.42-7.56)	
	Over-dominant	G/G-C/C	447 (81.4%)	210 (85.4%)	1	0.2	447 (81.4%)	146 (81.1%)	1	0.94
		G/C	102 (18.6%)	36 (14.6%)	0.77 (0.50-1.16)		102 (18.6%)	34 (18.9%)	0.98 (0.64-1.52)	
	Log-additive	---	---	---	0.71 (0.47-1.05)	0.082	---	---	1.06 (0.72-1.57)	0.75

SNP: Single nucleotide polymorphism. PR: Progesterone receptor. BMI: Body mass index (weight [kg]/height [m]^2). *$p < 0.05$ indicates statistical significance. OR: Odds ratio. CI: Confidence interval.