

ORIGINAL ARTICLE

Identification of novel susceptibility markers for the risk of overall breast cancer as well as subtypes defined by hormone receptor status in the Chinese population

Zhiping Deng^{1,2}, Hua Yang³, Qiufang Liu², Zhouquan Wang¹, Tian Feng³, Yongri Ouyang³, Tianbo Jin³ and Hong Ren¹

Genome-wide association studies (GWAS) have identified common variants associated with breast cancer (BC) risk at multiple genetic loci. Above all, accumulated evidence suggests that inherited risk variants may vary in BC subtypes defined by estrogen receptor (ER) or progesterone receptor (PR) status. However, the underlying susceptibility of some variants for BC subtypes has not been well investigated in the Chinese population. Our objective was to explore the association among 23 GWAS-identified single-nucleotide polymorphisms (SNPs) and overall BC incidence, as well as its subtypes, in Chinese women. An extensive association analysis using the Sequenom MassARRAY® platform was conducted in a case–control study, including 551 BC patients and 577 healthy controls. Using the chi-squared (χ^2) test and genetic model analysis, we found an association with BC for four SNPs (rs616488 (1p36.22/*PEX14*), rs6678914 (1q32.1/*LGR6*), rs17530068 (6q14/unknown) and rs6001930 (22q13.1/*MKL1*)) at a 5% level. Stratified analyses under this genetic model determined that rs616488 and rs6001930 were specific to ER positive and PR positive, rs17530068 was specific to ER positive and PR negative, rs3817198 (11p15.5/*LSP1*) was specific to ER negative and rs4784227 (16q12.1/*CASC16*) was specific to PR positive. Our study provides powerful new evidence for the relationship between SNPs and BC susceptibility.

Journal of Human Genetics (2016) 61, 1027–1034; doi:10.1038/jhg.2016.97; published online 8 September 2016

INTRODUCTION

Breast cancer (BC), a substantial global public health concern, is one of the most common malignancies diagnosed in women, with an estimated 1.4 million new cases and over 450 000 deaths annually worldwide.¹ BC is primarily a hormone-dependent disease that can be regulated by the status of steroid hormones such as estrogen and progesterone.² BC can be divided into five subtypes, which vary in their treatment options and survival outcomes based on gene expression profiles.^{3–5} Among the five subtypes of BC, estrogen receptor (ER)-positive and progesterone receptor (PR)-positive tumors account for ~70% of all cases.^{6,7}

Genetic factors have an important role in the etiology of both sporadic and familial BC.^{8,9} High-penetrance BC susceptibility genes, such as *BRCA1* and *BRCA2*, account for only a small proportion of BC in the general population because of their low mutation rates.¹⁰ To date, genome-wide association studies (GWAS) in European, African-American and East-Asian descendants have identified common variants associated with BC risk at multiple genetic loci.^{11–18} Above all, accumulated epidemiologic data suggest substantial heterogeneity in breast tumor subtypes, defined by hormone receptor status, for association with gene polymorphisms.^{12,15,17,19–22} Therefore, detailed

stratification of tumors may deepen our understanding of BC etiology, facilitate the discovery of novel risk factors and potentially enable risk prediction for specific tumor types. Although all observations indicate that inherited risk variants may vary in diverse BC subtypes, the underlying susceptibility of some variants for BC subtype has not been well investigated in the Chinese population.

In this study, we selected 23 GWAS-identified single-nucleotide polymorphisms (SNPs) to investigate and verify their putative association with BC and specific tumor subtypes defined by ER and PR status in Chinese women. We conducted an association analysis in a case–control study in women from the northwest of China (Shaanxi province). Our data provided considerable evidence for the association among common SNPs and the overall risk of BC as well as tumor subtypes. These results may eventually further improve prevention, early detection and treatment of BC.

SUBJECTS AND METHODS

Subjects

In this case–control study, all participants were Chinese women. A total of 551 unrelated subjects with BC (mean age, 49 ± 11) were recruited from the First Affiliated Hospital of the Medical College of Xi'an Jiao tong University from

¹Department of Oncology Surgery, First Affiliated Hospital, Xi'an Jiaotong University, Shaanxi, China; ²Department of Breast Surgery, Tumor Hospital of Shaanxi Province, Shaanxi, China and ³Department of School of Life Sciences, School of Life Sciences, Northwest University, Shaanxi, China
Correspondence: Professor H Ren, Department of Oncology Surgery, First Affiliated Hospital, Xi'an Jiaotong University, Mailbox 96 #277 West Yanta Road, Xi'an, 710069 Shaanxi, China.

E-mail: renhong2014@163.com

Received 24 February 2016; revised 2 July 2016; accepted 4 July 2016; published online 8 September 2016

January 2011 to November 2014. Within BC cases, there were 292 patients with ER-positive tumor, and 136 patients with ER-negative tumor; additionally, there were 247 patients with PR-positive tumors and 180 patients with PR-negative tumors. BC was defined according to the patient's surgical and pathological symptoms, and their disease information was obtained from their medical files.

In sum, 577 healthy blood donors (mean age of 49 ± 8) were recruited from Han origin women living in the city of Xi'an and its surrounding areas. The control group was matched for age and ethnicity with patients, without a history of cancer. Additionally, we selected patients with a body mass index ($\text{BMI} = \text{weight (kg)}/\text{height (m}^2\text{)}$) in the normal range of 18.5–24.9 in both the case and control groups.

Body size is an important modifiable risk factor for BC. The participants were not genetically related within three generations. Written informed consent was obtained from all participants, and the study protocol was approved by the Ethical Committee of the Medical College, Xi'an Jiaotong University.

SNP selection and genotyping

In all, 23 SNPs with minor allele frequencies greater than 0.05 were selected from GWAS based on a review of published literature and a search of HapMap and dbSNP (Han Chinese population).^{11,12,16,23–29} Table 2 lists the 23 SNPs that were selected and outlines other relevant characteristics. Genomic DNA was extracted from peripheral blood using the Qiagen Blood Kit (Qiagen, Chatsworth, CA, USA) according to the manufacturer's protocol. Amplification and extension primers were designed using MassARRAY[®] Assay Design 3.0 software (Sequenom, San Diego, CA, USA). SNP genotyping was performed using matrix-assisted laser desorption ionization-time of flight (MassARRAY system, Sequenom Inc.) mass spectrometry. Genotype calling was performed in real time with the MassARRAY RT software version 3.0.0.4 and analyzed using the MassARRAY Typer software version 3.4 (Sequenom). The experimenters were blinded to the case/control status of the samples.

Statistical analysis

Differences among cases and controls in demographic characteristics, including age and BMI, were evaluated by the Student's *t*-test. The allele and genotype frequencies for each SNP were compared, and Hardy–Weinberg equilibrium was evaluated using the Chi-squared (χ^2) test among the controls. Associations between SNPs and BC were assessed by the Pearson χ^2 test or the Fisher's exact test. The four genetic models (codominant, dominant, recessive and additive) were applied by PLINK software (<http://pngu.mgh.harvard.edu/purcell/plink/>) to assess the association of each locus with the risk of BC. Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used to determine the best-fitting model for each SNP. These measures can weigh the estimated complexity of the model and the goodness of fit to the data. The model with the smallest AIC or BIC value should be selected. In stratified analyses, case patients were classified into subgroups of ER-positive, ER-

negative, PR-positive and PR-negative tumor types. We further assessed the association of the genotypes of these SNPs with the risk of tumor subtypes under four different genetic models (codominant, dominant, recessive and additive models). Unconditional logistic regression with adjustments for age and BMI was used to calculate the odds ratio (OR) and 95% confidence interval (CI) in independent association between each SNP and the overall BC risk as well as subtypes. In this study, all *P*-values were two-sided and $P \leq 0.05$ was the standard for statistical significance. Statistical analyses were performed using Microsoft Excel and SPSS 17.0 software (SPSS Inc., Chicago, IL, USA).

RESULTS

Subject characteristics

Basic characteristics of the 551 case patients and 577 healthy controls were presented in Table 1. Age (age at diagnosis for cases and age at recruitment for controls) was equally distributed among BC cases (overall BC and subtype defined by ER/PR status) and controls. However, comparing with controls, overall cases, ER-positive cases and PR-positive cases, respectively, showed distribution differences in terms of BMI using Student's *t*-test, while no significant distribution differences of BMI were found between remaining cases and controls.

Association between SNPs and overall BC risk

In the current study, 23 SNPs were genotyped, and the average SNP call rate was 99.78% (98.16–100%) in both cases and controls. Among them, SNPs rs13281615 and rs2380205 deviated from Hardy–Weinberg equilibrium test with a *P*-value of < 0.05 (Table 2). We analyzed the association of SNP genotypes and the overall susceptibility to BC by χ^2 test, finding four significant SNPs (rs616488, rs6678914, rs17530068 and rs6001930) at a 5% level. The frequency of the C allele in rs616488 ($P = 0.0003$, odds ratio (OR) = 0.722, 95% confidence interval (CI) = 0.605–0.861) and the A allele in rs6678914 ($P = 0.029$, OR = 0.796, 95% CI = 0.648–0.977) in BC cases was significantly lower than that in controls (Table 2). The rs17530068 C allele frequency ($P = 0.009$, OR = 1.289, 95% CI = 1.064–1.561) and the rs6001930 C allele frequency ($P = 0.041$, OR = 1.209, 95% CI = 1.008–1.450) in BC cases were significantly higher than that in the controls (Table 2). Seventeen additional SNPs yielded negative results (Table 2). To reduce the potential of spurious findings due to multiple testings, a strict Bonferroni correction analysis was applied; we found that one SNP (rs616488) was significant, while SNPs rs678914, rs6001930 and rs17530068 were not significant (Table 2).

We then analyzed the genotype effects of these SNPs by unconditional logistic regression analysis with adjustments for age and BMI

Table 1 Basic characteristics of the cases and controls

Group	Age			BMI		
	Statistical number	Mean age \pm s.d. (years)	<i>P</i> -value	Statistical number	Mean BMI \pm s.d. (kg m^{-2})	<i>P</i> -value
Controls	577	48.790 \pm 8.294		549	22.952 \pm 3.214	
Cases	551	49.09 \pm 11.022	0.506	459	22.52 \pm 2.837	0.027 ^a
<i>ER status</i>						
Positive	292	49.164 \pm 10.781	0.603	292	22.409 \pm 2.811	0.015 ^a
Negative	136	49.691 \pm 10.145	0.337	136	23.242 \pm 2.815	0.335
Missing	123					
<i>PR status</i>						
Positive	247	49.077 \pm 10.773	0.709	247	22.340 \pm 2.848	0.010 ^a
Negative	180	49.656 \pm 10.341	0.307	180	23.154 \pm 2.748	0.449
Missing	124					

Abbreviations: BMI, body mass index; ER, estrogen receptor; PR, progesterone receptor.
^a*P*-value ≤ 0.05 indicates statistical significance.

Table 2 Basic information of candidate SNPs in this study

SNP ID	Ethnicity	Gene name	Chromosome	Position	Alleles A/B	HWE	P-value ^a	P-value ^b	OR (95% CI) ^c	OR (95% CI) ^d	Call rate (%)
rs616488	European ³⁸	PEX14	1p36.22	10566215	C/T	0.718	0.0003*	0.0345**	0.72 (0.60-0.86)	1.06 (1.04-1.09)	99.91
rs6678914	European ²³	LGR6	1q32.1	202187176	A/G	0.904	0.029*	1	0.8 (0.65-0.98)	1.1 (1.06-1.13)	99.83
rs13387042	European ²⁴	/	2q35	217905832	A/G	0.41	0.123	1	1.22 (0.95-1.57)	1.21 (1.14-1.29)	99.48
rs16857609	European ³⁸	DIRC3	2q35	218296508	C/T	0.542	0.79	1	0.98 (0.83-1.16)	1.08 (1.06-1.10)	100
rs1092913	European ²⁵	ROPNIL	5p15.2	10467702	C/T	0.16	0.506	1	0.94 (0.78-1.13)	1.45 (1.24-1.69)	99.91
rs4415084	European ³⁸	/	5p12	44662515	A/C	0.095	0.837	1	0.98 (0.83-1.16)	1.17 (1.11-1.22)	100
rs10941679	European ³⁸	/	5p12	44706498	G/T	0.281	0.824	1	1.02 (0.86-1.20)	1.13 (1.10-1.15)	99.83
rs7716600	European ²⁶	/	5p12	44875005	G/T	0.804	0.55	1	1.05 (0.89-1.24)	1.24 (1.14-1.34)	100
rs11242675	European ³⁸	/	6p25.3	1318878	T/C	0.054	0.992	1	1.00 (0.85-1.18)	1.06 (1.04-1.09)	99.74
rs17530068	European ¹⁶	/	6q14.1	82193109	C/T	0.471	0.009*	1	1.29 (1.06-1.56)	1.16 (1.10-1.23)	99.48
rs2048672	Asia ²⁷	LINC-PINT	7q32.3	130653851	G/T	1	0.348	1	0.92 (0.78-1.09)	1.11 (1.05-1.17)	99.83
rs13281615	European ³⁸	CASC8	8q24.21	128355618	A/G	3.196E-27 ^e	0.664	1	0.96 (0.81-1.14)	1.09 (1.07-1.12)	99.83
rs1562430	European ²⁴	CASC8	8q24.21	128387852	G/A	0.891	0.992	1	1.00 (0.81-1.23)	1.17 (1.10-1.25)	99.91
rs1011970	European ³⁸	CDKN2B-AS1	9p21.3	22062134	A/C	1	0.473	1	1.11 (0.83-1.49)	1.06 (1.03-1.08)	99.48
rs2380205	European ²⁴	/	10p15.1	5886734	T/C	0.025 ^e	0.09	1	1.26 (0.96-1.65)	1.06 (1.02-1.10)	100
rs3817198	European ¹¹	LSP1	11p15.5	1909006	C/T	1	0.261	1	1.15 (0.90-1.48)	1.07 (1.04-1.11)	100
rs12422552	European ³⁸	/	12p13.1	14413931	C/G	0.484	0.815	1	1.02 (0.85-1.22)	1.05 (1.03-1.07)	99.91
rs1292011	European ³⁸	/	12q24.21	115836522	G/A	0.083	0.265	1	0.89 (0.73-1.09)	1.09 (1.06-1.11)	99.74
rs3803662	European ³⁸	CASC16	16q12.1	52586341	C/T	1	0.23	1	0.90 (0.75-1.07)	1.24 (1.21-1.27)	100
rs4784227	Asia ¹⁴	CASC16	16q12.1	52599188	T/C	0.494	0.078	1	1.19 (0.98-1.44)	1.24 (1.20-1.29)	98.16
rs3112612	European ²⁹	CASC16	16q12.1	52635164	C/T	0.701	0.904	1	1.01 (0.83-1.24)	1.15 (1.10-1.21)	100
rs17817449	European ³⁸	FTO	16q12.2	53813367	G/T	0.641	0.419	1	1.12 (0.85-1.46)	1.09 (1.05-1.12)	100
rs6001930	European ³⁸	MKL1	22q13.1	40876234	C/T	0.457	0.041*	1	1.21 (1.01-1.45)	1.12 (1.09-1.16)	100

Abbreviations: CI, confidence interval; HWE, Hardy-Weinberg equilibrium; ORs, odds ratios; SNPs, single-nucleotide polymorphisms.
^aP-value ≤ 0.05 indicates statistical significance. ^{**}P-value ≤ 0.05 indicates statistical significance by Bonferroni correction. A/B stands for minor/major alleles on the control sample frequencies.
^bP-value was calculated by Pearson's χ^2 test.
^cP-value was adjusted by Bonferroni corrections.
^dReported OR.
^eSite with HWE P ≤ 0.05 is excluded.

Table 3 Relationship between significant SNPs and overall breast cancer risk (adjusted by age+BMI)

SNP ID	Model	Genotype	Control (N, %)	Case (N, %)	OR (95% CI)	P-value	AIC	BIC
rs616488	Codominant	T/T	245 (42%)	287 (51.3%)	1.00	0.0015*	1575.6	1590.7
		C/T	263 (45.1%)	227 (40.6%)	0.74 (0.58–0.94)			
		C/C	75 (12.9%)	45 (8.1%)	0.51 (0.34–0.77)			
	Dominant	T/T	245 (42%)	287 (51.3%)	1.00	0.0016*	1576.7	1586.8
		C/T-C/C	338 (58%)	272 (48.7%)	0.69 (0.54–0.87)			
	Recessive	T/T-C/T	508 (87.1%)	514 (92%)	1.00	0.0077*	1579.5	1589.6
C/C		75 (12.9%)	45 (8.1%)	0.59 (0.40–0.88)				
	Log-additive	—	—	—	0.72 (0.61–0.86)	3 × 10 ⁻⁴ *	1573.6	1583.7
rs6678914	Codominant	G/G	355 (60.9%)	368 (66%)	1.00	0.056	1581.5	1596.6
		A/G	199 (34.1%)	175 (31.4%)	0.85 (0.66–1.09)			
		A/A	29 (5%)	15 (2.7%)	0.50 (0.26–0.95)			
	Dominant	G/G	355 (60.9%)	368 (66%)	1.00	0.076	1582.1	1592.1
		A/G-A/A	228 (39.1%)	190 (34%)	0.80 (0.63–1.02)			
	Recessive	G/G-A/G	554 (95%)	543 (97.3%)	1.00	0.043	1581.1	1591.2
A/A		29 (5%)	15 (2.7%)	0.53 (0.28–1.00)				
	Log-additive	—	—	—	0.79 (0.64–0.98)	0.028*	1580.4	1590.4
rs17530068	Codominant	T/T	348 (59.9%)	297 (53.4%)	1.00	0.026*	1574.4	1589.5
		T/C	208 (35.8%)	219 (39.4%)	1.23 (0.97–1.58)			
		C/C	25 (4.3%)	40 (7.2%)	1.87 (1.11–3.16)			
	Dominant	T/T	348 (59.9%)	297 (53.4%)	1.00	0.028*	1574.8	1584.9
		T/C-C/C	233 (40.1%)	259 (46.6%)	1.30 (1.03–1.65)			
	Recessive	T/T-T/C	556 (95.7%)	516 (92.8%)	1.00	0.035*	1575.2	1585.3
C/C		25 (4.3%)	40 (7.2%)	1.72 (1.03–2.88)				
	Log-additive	—	—	—	1.29 (1.07–1.57)	0.0088*	1572.8	1582.9
rs6001930	Codominant	T/T	318 (54.5%)	271 (48.4%)	1.00	0.11	1585.6	1600.8
		C/T	220 (37.7%)	237 (42.3%)	1.26 (0.99–1.61)			
		C/C	45 (7.7%)	52 (9.3%)	1.36 (0.88–2.09)			
	Dominant	T/T	318 (54.5%)	271 (48.4%)	1.00	0.037*	1583.7	1593.8
		C/T-C/C	265 (45.5%)	289 (51.6%)	1.28 (1.01–1.61)			
	Recessive	T/T-C/T	538 (92.3%)	508 (90.7%)	1.00	0.34	1587.2	1597.3
C/C		45 (7.7%)	52 (9.3%)	1.22 (0.81–1.86)				
	Log-additive	—	—	—	1.20 (1.01–1.44)	0.043*	1584	1594

Abbreviations: AIC, Akaike's Information criterion; BIC, Bayesian Information criterion; BMI, body mass index; CI, confidence interval; OR, odd ratio; SNPs, single-nucleotide polymorphisms. *P-value ≤ 0.05 indicates statistical significance; ORs in bold indicate statistical significance.

under four different genetic models (codominant, dominant, recessive and additive). Four significant SNPs (rs616488, rs6678914, rs17530068 and rs6001930) were shown in Table 3. The SNP rs616488 C allele provided a protective effect against BC ($P=0.0015$ for codominant model, $P=0.0016$ for dominant model, $P=0.0077$ for recessive model and $P=3 \times 10^{-4}$ for additive model). As for SNP rs6678914, the A allele also conferred a protective effect for tumor ($P=0.028$ for additive model). On the contrary, the SNP rs17530068 C allele increased BC risk ($P=0.026$ for codominant model, $P=0.028$ for dominant model, $P=0.035$ for recessive model and $P=0.009$ for additive model). Likewise, the SNP rs6001930 C allele increased disease risk ($P=0.037$ for dominant model and $P=0.043$ for additive model).

Associations among SNPs and the risk of tumor defined ER status

As listed in Table 4, 292 patients exhibited ER-positive tumors, 136 patients exhibited ER-negative tumors and 3 SNPs (rs616488, rs17530068 and rs6001930) exhibited significant associations with ER-positive tumor risk. The SNP rs616488 C allele ($P=0.014$ for codominant model, $P=0.007$ for dominant model and $P=0.003$ for additive model) decreased the ER-positive risk; the SNP rs17530068 C allele ($P=0.039$ for additive model) and rs6001930 C allele ($P=0.016$ for dominant model and $P=0.019$ for additive model) increased the

ER-positive risk. At the same time, only one SNP (rs3817198) exhibited a significant association with ER-negative tumor risk. The SNP rs3817198 C allele ($P=0.024$ for recessive model) increased the ER-negative tumor risk.

Associations between SNPs and the risk of tumor defined PR status

Table 5 illustrates that 247 patients had PR-positive tumor, and 177 patients had PR-negative tumors and 3 SNPs (rs616488, rs4784227 and rs6001930) had significant association with PR-positive tumor risk. The SNP rs616488 C allele ($P=0.015$ for codominant model, $P=0.013$ for dominant model, $P=0.026$ for recessive model and $P=0.004$ for additive model) decreased risk, whereas the SNP rs4784227 T allele ($P=0.037$ for dominant model) and rs6001930 C allele ($P=0.011$ for codominant model, $P=0.003$ for dominant model and $P=0.004$ for additive model) increased risk. In contrast, only one SNP (rs17530068) exhibited significant association with PR-negative tumor risk. The rs17530068 C allele ($P=0.024$ for dominant model and $P=0.018$ for additive model) increased the risk.

DISCUSSION

In this study, we conducted a thorough association analysis among 23 GWAS-identified SNPs and BC as well as its subtypes in Chinese women by using matrix-assisted laser desorption ionization-time of

Table 4 Relationship between significant SNPs and ER status tumor risk (adjusted by age + BMI)

SNP ID	Model	Genotype	Control (N, %)	Case (N, %)	OR (95%CI)	P-value	AIC	BIC					
<i>ER positive</i>													
rs616488	Codominant	T/T	245 (42%)	151 (51.7%)	1.00	0.014*	1111.8	1126.1					
		C/T	263 (45.1%)	116 (39.7%)	0.72 (0.53–0.96)								
		C/C	75 (12.9%)	25 (8.6%)	0.54 (0.33–0.89)								
	Dominant	T/T	245 (42%)	151 (51.7%)	1.00	0.0067*	1111	1120.6					
		C/T-C/C	338 (58%)	141 (48.3%)	0.68 (0.51–0.90)								
	Recessive	T/T-C/T	508 (87.1%)	267 (91.4%)	1.00	0.054	1114.6	1124.2					
C/C		75 (12.9%)	25 (8.6%)	0.63 (0.39–1.02)									
rs17530068	Log-additive	—	—	—	0.73 (0.59–0.90)	0.0034*	1109.8	1119.4					
		Codominant	T/T	348 (59.9%)	156 (53.6%)				1.00	0.11	1112.1	1126.4	
			T/C	208 (35.8%)	115 (39.5%)				1.23 (0.92–1.66)				
	C/C		25 (4.3%)	20 (6.9%)	1.78 (0.96–3.31)								
	Dominant	T/T	348 (59.9%)	156 (53.6%)	1.00	0.077	1111.4	1121					
		T/C-C/C	233 (40.1%)	135 (46.4%)	1.29 (0.97–1.72)								
	Recessive	T/T-T/C	556 (95.7%)	271 (93.1%)	1.00	0.11	1112	1121.6					
		C/C	25 (4.3%)	20 (6.9%)	1.64 (0.90–3.01)								
	rs6001930	Log-additive	—	—	—	1.28 (1.01–1.62)	0.039*	1110.3	1119.8				
			Codominant	T/T	318 (54.5%)	134 (45.9%)				1.00	0.051	1114.4	1128.7
				C/T	220 (37.7%)	129 (44.2%)				1.39 (1.03–1.87)			
		C/C		45 (7.7%)	29 (9.9%)	1.53 (0.92–2.54)							
Dominant		T/T	318 (54.5%)	134 (45.9%)	1.00	0.016*	1112.5	1122.1					
		C/T-C/C	265 (45.5%)	158 (54.1%)	1.41 (1.07–1.88)								
Recessive		T/T-C/T	538 (92.3%)	263 (90.1%)	1.00	0.27	1117.2	1126.7					
		C/C	45 (7.7%)	29 (9.9%)	1.32 (0.81–2.15)								
Log-additive		—	—	—	—	1.30 (1.04–1.61)	0.019*	1112.9	1122.4				
<i>ER negative</i>													
rs3817198		Codominant	T/T	455 (78%)	103 (75.7%)	1.00	0.078	698.3	712.1				
			C/T	121 (20.8%)	27 (19.9%)	0.99 (0.62–1.58)							
	C/C		7 (1.2%)	6 (4.4%)	3.79 (1.25–11.50)								
	Dominant	T/T	455 (78%)	103 (75.7%)	1.00	0.56	701.1	710.2					
		C/T-C/C	128 (22%)	33 (24.3%)	1.14 (0.73–1.77)								
	Recessive	T/T-C/T	576 (98.8%)	130 (95.6%)	1.00	0.024*	696.3	705.5					
		C/C	7 (1.2%)	6 (4.4%)	3.80 (1.26–11.49)								
	Log-additive	—	—	—	—	1.27 (0.87–1.85)	0.22	699.9	709.1				

Abbreviations: AIC, Akaike's Information criterion; BIC, Bayesian Information criterion; BMI, body mass index; CI, confidence interval; ER, estrogen receptor; OR, odd ratio; SNPs, single-nucleotide polymorphisms.

*P-value \leq 0.05 indicates statistical significance; ORs in bold indicate statistical significance.

flight mass spectrometry. Among 23 SNPs, 4 SNPs (rs616488, rs6678914, rs17530068 and rs6001930) were identified to be significantly associated with overall BC risk. Stratified analyses identified that rs616488 and rs6001930 were specific to ER positive and PR positive, rs17530068 was specific to ER positive and PR negative, rs3817198 was specific to ER negative and rs4784227 was specific to PR positive. Overall, these findings provided strong evidence of genetic susceptibility to overall BC and BC subtypes.

The C allele of rs616488 (1p36.22/PEX14) and the A allele of rs6678914 (1q32.1/LGR6) were protective factors for BC in the current study. PEX14 is one of ~15–20 genes implicated in the biogenesis of mammalian peroxisomes. Peroxisomes are part of several metabolic pathways; they perform oxidative reactions catalyzed by amino-acid oxidases and catalases, most notably the β -oxidation of very long chain fatty acids, which are then fully degraded in mitochondria.^{30,31} The PEX gene fragment codes for the degradation of the C-terminus of matrix metalloproteinase-2, and studies have shown that this fragment can inhibit matrix metalloproteinase-2 extracellular matrix degradation and tumor angiogenesis.³² Extracellular matrix degradation and tumor angiogenesis have a vital role in tumor cell growth, invasion

and metastasis. PEX gene *in vivo* can inhibit tumor cell growth. The rs6678914, on chromosome 1q32.1 is located in intron 1 of the LGR6 gene, which is expressed in breast tumors along with several other genes in this region, including UBE2T and PTPN7.³³ Previous studies have reported that the SNP rs616488 has statistically significant association with BC risk in European¹² and East-Asian¹⁸ descendants but not in African-American women,¹⁴ which indicated that there are large differences in genetic architecture between the African-ancestry genome and genomes of Asians and Europeans. Interestingly, our subtype analyses showed that the C allele of this locus was a protective factor for both ER-positive and PR-positive cancers. However, Zheng *et al.*¹⁸ found that this SNP had no statistically significant association with BC defined ER status in East-Asian descendants. This contradiction requires further investigation and validation in a larger population. As for rs6678914, Garcia-Closas *et al.*¹² confirmed that this SNP was associated with ER-negative but not ER-positive BC in populations of European ancestry. Subsequently, Sawyer *et al.*³⁴ suggested that rs6678914 was more strongly associated with lobular carcinoma *in situ* than with invasive lobular breast cancer. Along with our findings, these reports contribute to an association between LGR6

Table 5 Relationship between significant SNPs and PR status tumor risk (adjusted by age + BMI)

SNP ID	Model	Genotype	Control (N, %)	Case (N, %)	OR (95% CI)	P-value	AIC	BIC	
<i>PR positive</i>									
rs616488	Codominant	T/T	245 (42%)	127 (51.4%)	1.00	0.015*	1008.2	1022.3	
		C/T	263 (45.1%)	101 (40.9%)	0.74 (0.54–1.01)				
		C/C	75 (12.9%)	19 (7.7%)	0.49 (0.28–0.84)				
	Dominant	T/T	245 (42%)	127 (51.4%)	1.00	0.013*	1008.5	1017.9	
		C/T-C/C	338 (58%)	120 (48.6%)	0.68 (0.51–0.92)				
	Recessive	T/T-C/T	508 (87.1%)	228 (92.3%)	1.00	0.026*	1009.7	1019.1	
C/C		75 (12.9%)	19 (7.7%)	0.56 (0.33–0.96)					
rs4784227	Log-additive	—	—	—	0.72 (0.57–0.90)	0.0038*	1006.3	1015.7	
	Codominant	C/C	340 (58.5%)	123 (50.6%)	1.00	0.11	1001.2	1015.3	
		T/C	205 (35.3%)	102 (42%)	1.38 (1.00–1.88)				
		T/T	36 (6.2%)	18 (7.4%)	1.38 (0.76–2.52)				
	Dominant	C/C	340 (58.5%)	123 (50.6%)	1.00	0.037*	999.2	1008.6	
		T/C-T/T	241 (41.5%)	120 (49.4%)	1.38 (1.02–1.86)				
	Recessive	C/C-T/C	545 (93.8%)	225 (92.6%)	1.00	0.53	1003.1	1012.5	
		T/T	36 (6.2%)	18 (7.4%)	1.21 (0.67–2.18)				
		—	—	—	1.26 (1.00–1.60)				0.055
	rs6001930	Codominant	T/T	318 (54.5%)	107 (43.3%)	1.00	0.011*	1007.5	1021.7
			C/T	220 (37.7%)	113 (45.8%)	1.53 (1.11–2.09)			
			C/C	45 (7.7%)	27 (10.9%)	1.78 (1.05–3.01)			
Dominant		T/T	318 (54.5%)	107 (43.3%)	1.00	0.0031*	1005.9	1015.3	
		C/T-C/C	265 (45.5%)	140 (56.7%)	1.57 (1.16–2.12)				
Recessive		T/T-C/T	538 (92.3%)	220 (89.1%)	1.00	0.14	1012.4	1021.9	
		C/C	45 (7.7%)	27 (10.9%)	1.47 (0.89–2.42)				
		—	—	—	1.40 (1.12–1.76)				0.0035*
<i>PR negative</i>									
rs17530068		Codominant	T/T	348 (59.9%)	89 (50.3%)	1.00	0.06	824.3	838.2
			T/C	208 (35.8%)	76 (42.9%)	1.43 (1.01–2.03)			
			C/C	25 (4.3%)	12 (6.8%)	1.88 (0.91–3.88)			
	Dominant	T/T	348 (59.9%)	89 (50.3%)	1.00	0.024*	822.8	832.1	
		T/C-C/C	233 (40.1%)	88 (49.7%)	1.48 (1.05–2.07)				
	Recessive	T/T-T/C	556 (95.7%)	165 (93.2%)	1.00	0.2	826.3	835.5	
		C/C	25 (4.3%)	12 (6.8%)	1.62 (0.80–3.29)				
	Log-additive	—	—	—	1.40 (1.06–1.85)	0.018*	822.3	831.6	

Abbreviations: AIC, Akaike's Information criterion; BIC, Bayesian Information criterion; BMI, body mass index; CI, confidence interval; PR, progesterone receptor; OR, odd ratio; SNPs, single-nucleotide polymorphisms.

*P-value ≤ 0.05 indicates statistical significance; ORs in bold indicate statistical significance.

rs6678914 and BC susceptibility, our stratified analysis was not significant in this study because the sample size was relatively small.

In this study, the C alleles in the SNPs rs17530068 (6q14/unknown) and rs6001930 (22q13.1/*MKLI*) were risk factors for BC. The SNP rs17530068 at 6q14 is located in a gene desert with no evidence of an open/active regulatory region in human microvascular endothelial cells (HMEC). The closest gene (~262 kb), family with sequence similarity 46, member A (*FAM46A/C6orf37*), encodes a protein of unknown function. The *FAM46A* gene is located at chromosome 614.1 and was first identified and cloned from human retina tissue as a retinal disease candidate gene.³⁵ Studies showed that the *FAM46A* gene is expressed in ameloblast nuclei of developing teeth and hypothesized that it might act together with morphogenetic factors involved in cell proliferation, apoptosis and differentiation activities in tooth buds, and perhaps in enamel production.³⁶ Data suggest that patients with C allele of rs17530068 are associated with increased BC. On the basis of these facts, we hypothesized that the *FAM46A* gene may be involved in BC. However, there might be other biological pathways that are not yet reported, and further functional studies will be necessary to elucidate the precise role of the *FAM46A* gene. A previous meta-

analysis of GWAs, with regard to Japanese, Latino and European descendants, showed that rs17530068 was associated with BC, and with both ER-positive and ER-negative diseases.¹⁶ In East-Asian women (Chinese, Korean and Japanese) population, Zheng *et al.*¹⁸ found that SNP rs6001930 had no statistically significant association with BC, but subgroup analysis revealed that rs6001930 had a stronger association with ER-positive than ER-negative cancer. These results were partially consistent with ours, and to our knowledge, our data demonstrated the relationship between rs17530068 and PR-negative risk and between rs6001930 and PR-positive disease risk in the Chinese population for the first time.

Most intriguingly, although no overall association of BC was found for rs3817198 (11p15.5/*LSP1*) and rs4784227 (16q12.1/*CASC16*), analyses by ER/PR status revealed a statistically significant association between rs3817198 and ER-negative tumors, as well as between rs4784227 and PR-positive tumors. In contradiction with our results, rs3817198 showed stronger associations with ER-positive than ER-negative tumors in populations of European ancestry.¹⁹ Located in the 16q12.1,²⁸ rs4784227 has been predicted to interfere with the affinity of FOXA1, an essential component of ESRα signaling,³⁷ to its binding

site.³⁸ Its position in a regulatory region that interacts with the TOX3 promoter enables it to disrupt the expression of this gene, which in turn alters chromatin structure and DNA-protein binding patterns essential for cell survival.²⁸ Although rs4784227 had a statistically significant association with BC risk in European and Asian women, Lin *et al.*³⁹ did not find a correlation between rs4784227 and PR-positive tumors in a stratified interaction analysis of the Chinese population. Although the evidence for these associations was not very strong, additional analyses, involving a much larger number of BC patients, are needed to independently confirm these associations and assess whether their risks vary by tumor subtype.

Population stratification is an important factor to consider when conducting human genetic surveys.⁴⁰ In the current study, cases and controls were matched for ethnicity by enrolling subjects from a homogeneous population. In addition, our results were in line with the hypothesis that 'different biological mechanisms underlie diverse BC subtypes', and indicated that tumor stratification might help in the identification of novel susceptibility markers for diverse BC subtype.⁴¹ Our results eventually help improve BC prevention, early detection and treatment. However, our analysis data for ER and PR were available for only a portion of the subjects. Therefore, the statistical power of our study was limited in the stratified analyses because of the small sample sizes. Some of the null associations observed in this study could be due to inadequate statistical power.

In summary, our study provided powerful new evidence for the relationship among SNPs and the risk of overall BC as well as the subtypes defined by ER and PR status in Chinese women. Our results shed light on the heterogeneity of different tumor subtypes according to protein expression of ER and PR, and the SNPs we detected can be applied in clinical diagnosis. However, the exact biological mechanism of how the polymorphisms regulate overall as well as tumor subtypes needs subsequent functional studies.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This study was funded by the 2013 Science and Technology Department of Shaanxi Province Natural Science Foundation Project (No. 2013JM4035) and 2013 National Natural Science Foundation of China. We are grateful to the BC patients and control subjects for their participation in this study. We also thank the clinicians and hospital staff who contributed to the sample and data collection.

- Zhang, B., Beeghly-Fadiel, A., Long, J. & Zheng, W. Genetic variants associated with breast-cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. *Lancet Oncol.* **12**, 477–488 (2011).
- Benson, J. R. & Jatoi, I. The global breast cancer burden. *Fut. Oncol.* **8**, 697–702 (2012).
- Easton, D. F., Pooley, K. A., Dunning, A. M., Pharoah, P. D., Thompson, D., Ballinger, D. G. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087–1093 (2007).
- Garcia-Closas, M., Couch, F. J., Lindstrom, S., Michailidou, K., Schmidt, M. K., Brook, M. N. *et al.* Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nat. Genet.* **45**, 392–398 (2013).
- He, X., Yao, G., Li, F., Li, M. & Yang, X. Risk-association of five SNPs in TOX3/LOC643714 with breast cancer in southern China. *Int. J. Mol. Sci.* **15**, 2130–2141 (2014).
- Long, J., Zhang, B., Signorello, L. B., Cai, Q., Deming-Halverson, S., Shrubsole, M. J. *et al.* Evaluating genome-wide association study-identified breast cancer risk variants in African-American women. *PLoS ONE* **8**, e58350 (2013).
- Purrington, K. S., Slager, S., Eccles, D., Yannoukakis, D., Fasching, P. A., Miron, P. *et al.* Genome-wide association study identifies 25 known breast cancer susceptibility loci as risk factors for triple-negative breast cancer. *Carcinogenesis* **35**, 1012–1019 (2014).
- Siddiq, A., Couch, F. J., Chen, G. K., Lindström, S., Eccles, D., Millikan, R. C. *et al.* A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. *Hum. Mol. Genet.* **21**, 5373–5384 (2012).
- Yu, Y., Chen, Z., Wang, H. & Zhang, Y. Quantitative assessment of common genetic variants on chromosome 5p12 and hormone receptor status with breast cancer risk. *PLoS ONE* **8**, e72154 (2013).
- Zheng, W., Zhang, B., Cai, Q., Sung, H., Michailidou, K., Shi, J. *et al.* Common genetic determinants of breast-cancer risk in East Asian women: a collaborative study of 23 637 breast cancer cases and 25 579 controls. *Hum. Mol. Genet.* **22**, 2539–2550 (2013).
- Broeks, A., Schmidt, M. K., Sherman, M. E., Couch, F. J., Hopper, J. L., Dite, G. S. *et al.* Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium. *Hum. Mol. Genet.* **20**, 3289–3303 (2011).
- Garcia-Closas, M. & Chanock, S. Genetic susceptibility loci for breast cancer by estrogen receptor status. *Clin. Cancer Res.* **14**, 8000–8009 (2008).
- Reeves, G. K., Travis, R. C., Green, J., Bull, D., Tipper, S., Baker, K. *et al.* Incidence of breast cancer and its subtypes in relation to individual and multiple low-penetrance genetic susceptibility loci. *JAMA* **304**, 426–434 (2010).
- Stacey, S. N., Manolescu, A., Sulem, P., Rafnar, T., Gudmundsson, J., Gudjonsson, S. A. *et al.* Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat. Genet.* **39**, 865–869 (2007).
- Michailidou, K., Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J., Milne, R. L. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* **45**, 353–361 (2013).
- Turnbull, C., Ahmed, S., Morrison, J., Pernet, D., Renwick, A., Maranian, M. *et al.* Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat. Genet.* **42**, 504–507 (2010).
- Sehrawat, B., Sridharan, M., Ghosh, S., Robson, P., Cass, C. E., Mackey, J. R. *et al.* Potential novel candidate polymorphisms identified in genome-wide association study for breast cancer susceptibility. *Hum. Genet.* **130**, 529–537 (2011).
- Li, J., Humphreys, K., Heikkinen, T., Aittomäki, K., Blomqvist, C., Pharoah, P. D. *et al.* A combined analysis of genome-wide association studies in breast cancer. *Breast Cancer Res. Treat.* **126**, 717–727 (2011).
- Cai, Q., Long, J., Lu, W., Qu, S., Wen, W., Kang, D. *et al.* Genome-wide association study identifies breast cancer risk variant at 10q21. 2: results from the Asia Breast Cancer Consortium. *Hum. Mol. Genet.* **20**, 4991–4999 (2011).
- Long, J., Cai, Q., Shu, X.-O., Qu, S., Li, C., Zheng, Y. *et al.* Identification of a functional genetic variant at 16q12. 1 for breast cancer risk: results from the Asia Breast Cancer Consortium. *PLoS Genet.* **6**, e1001002 (2010).
- Fletcher, O., Johnson, N., Orr, N., Hosking, F. J., Gibson, L. J., Walker, K. *et al.* Novel breast cancer susceptibility locus at 9q31. 2: results of a genome-wide association study. *J. Natl Cancer Inst.* **103**, 425–435 (2011).
- Brooks, P. C., Silletti, S., von Schalscha, T. L., Friedlander, M. & Cheresch, D. A. Disruption of angiogenesis by PEX, a noncatalytic metalloproteinase fragment with integrin binding activity. *Cell* **92**, 391–400 (1998).
- Shimizu, N., Itoh, R., Hirono, Y., Otera, H., Ghaedi, K., Tateishi, K. *et al.* The peroxin Pex14p cDNA cloning by functional complementation on a Chinese hamster ovary cell mutant, characterization, and functional analysis. *J. Biol. Chem.* **274**, 12593–12604 (1999).
- Ohira, M., Kageyama, H., Mihara, M., Furuta, S., Machida, T., Shishikura, T. *et al.* Identification and characterization of a 500-kb homozygously deleted region at 1p36. 2-p36. 3 in a neuroblastoma cell line. *Oncogene* **19**, 4302–4307 (2000).
- Turashvili, G., Bouchal, J., Baumforth, K., Wei, W., Dziechciarova, M., Ehrmann, J. *et al.* Novel markers for differentiation of lobular and ductal invasive breast carcinomas by laser microdissection and microarray analysis. *BMC Cancer* **7**, 1 (2007).
- Sawyer, E., Roylance, R., Petridis, C., Brook, M. N., Nowinski, S., Papouli, E. *et al.* Genetic predisposition to *in situ* and invasive lobular carcinoma of the breast. *PLoS Genet.* **10**, e1004285 (2014).
- Lagali, P., Kakuk, L., Griesinger, I., Wong, P. & Ayyagari, R. Identification and characterization of C6orf37, a novel candidate human retinal disease gene on chromosome 6q14. *Biochem. Biophys. Res. Commun.* **293**, 356–365 (2002).

- Siegel, R., Naishadham, D. & Jemal, A. Cancer statistics, 2013. *CA Cancer J. Clin.* **63**, 11–30 (2013).
- Ring, A. & Dowsett, M. Mechanisms of tamoxifen resistance. *Endocr. Relat. Cancer* **11**, 643–658 (2004).
- Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
- Reis-Filho, J., Westbury, C. & Pierga, J. The impact of expression profiling on prognostic and predictive testing in breast cancer. *J. Clin. Pathol.* **59**, 225–231 (2006).
- Rouzier, R., Wagner, P., Morandi, P. & Pusztai, L. Gene expression profiling of primary breast cancer. *Curr. Oncol. Rep.* **7**, 38–44 (2005).
- Anderson, W. F., Chatterjee, N., Ershler, W. B. & Brawley, O. W. Estrogen receptor breast cancer phenotypes in the Surveillance, Epidemiology, and End Results database. *Breast Cancer Res. Treat.* **76**, 27–36 (2002).
- Niemeier, L. A., Dabbs, D. J., Beriwal, S., Striebel, J. M. & Bhargava, R. Androgen receptor in breast cancer: expression in estrogen receptor-positive tumors and in estrogen receptor-negative tumors with apocrine differentiation. *Mod. Pathol.* **23**, 205–212 (2010).
- Nathanson, K. N., Wooster, R. & Weber, B. L. Breast cancer genetics: what we know and what we need. *Nat. Med.* **7**, 552–556 (2001).

- 36 Etokebe, G. E., Kuchler, A. M., Haraldsen, G., Landin, M., Osmundsen, H. & Dembic, Z. Family-with-sequence-similarity-46, member A (Fam46a) gene is expressed in developing tooth buds. *Arch. Oral Biol.* **54**, 1002–1007 (2009).
- 37 Bernardo, G. M. & Keri, R. A. FOXA1: a transcription factor with parallel functions in development and cancer. *Biosci. Rep.* **32**, 113–130 (2012).
- 38 Cowper-Sal, R., Zhang, X., Wright, J. B., Bailey, S. D., Cole, M. D., Eeckhoute, J. *et al.* Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat. Genet.* **44**, 1191–1198 (2012).
- 39 Lin, Y., Fu, F., Chen, M., Huang, M. & Wang, C. Associations of two common genetic variants with breast cancer risk in a Chinese population: a stratified interaction analysis. *PLoS ONE* **9** (2014).
- 40 Thomas, D. C. & Witte, J. S. Point: population stratification: a problem for case-control studies of candidate-gene associations? *Cancer Epidemiol. Biomarkers Prev.* **11**, 505–512 (2002).
- 41 Milne, R. L., Benítez, J., Nevanlinna, H., Heikkinen, T., Aittomäki, K., Blomqvist, C. *et al.* Risk of estrogen receptor-positive and -negative breast cancer and single-nucleotide polymorphism 2q35-rs13387042. *J. Natl Cancer Inst.* **101**, 1012–1018 (2009).