American Journal of Epidemiology
Vol. 177, No. 2
DOI: 10.1093/aje/kws238
Advance Access publication:
December 6, 2012

Original Contribution

Gene-Environment Interactions for Breast Cancer Risk Among Chinese Women: A Report From the Shanghai Breast Cancer Genetics Study

Haixin Li, Alicia Beeghly-Fadiel, Wanqing Wen, Wei Lu, Yu-Tang Gao, Yong-Bing Xiang, Qiuyin Cai, Jirong Long, Jiajun Shi, Kexin Chen, Ying Zheng, Xiao Ou Shu, and Wei Zheng*

* Correspondence to Dr. Wei Zheng, Vanderbilt Epidemiology Center and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, 2525 West End Avenue, 8th Floor, Nashville, TN 37203-1738 (e-mail: wei.zheng@vanderbilt.edu).

Initially submitted February 1, 2012; accepted for publication April 27, 2012.

Genome-wide association studies have identified approximately 20 susceptibility loci for breast cancer. A cumulative genetic risk score (GRS) was constructed from 10 variants with replicated associations among participants of the Shanghai Breast Cancer Genetics Study (Shanghai, China, 1996–1998 and 2002–2005). Interactions between the GRS and 11 breast cancer risk factors were evaluated. Among the 6,408 study participants, no evidence of effect modification was found with the GRS for age at menarche, age at menopause, age at first live birth/parity, total months of breastfeeding, family history of breast cancer, history of benign breast disease, hormone replacement therapy, body mass index, waist/hip ratio, or regular physical activity. The effect of the GRS was least homogeneous by duration of menstruation; further analysis indicated a nominally significant interaction with one genetic variant. The mitochondrial ribosomal protein S30 gene (*MRPS30*) rs10941679 was associated with breast cancer risk only among women with more than 30 years of menstruation (odds ratio = 1.15, 95% confidence interval: 1.05, 1.26). Although this multiplicative interaction reached a nominal significance level ($P = 0.037$), it did not withstand correction for multiple comparisons. In conclusion, this study revealed no apparent interactions between genome-wide association study-identified genetic variants and breast cancer risk factors in the etiology of this common cancer.

breast cancer risk; effect measure modification; gene-environment interaction; genetic variants; genome-wide association study

Abbreviations: CI, confidence interval; GRS, genetic risk score; OR, odds ratio; SNP, single nucleotide polymorphism.

The etiology of breast cancer is complex and multifactorial; genetic, environmental, and lifestyle factors all contribute to disease risk (1). Established reproductive and lifestyle breast cancer risk factors include age at menarche, age at menopause, parity, age at first live birth, alcohol consumption, postmenopausal obesity, sedentary behavior, and hormone replacement therapy (2–4). Genetic determinants including several high and moderate penetrance genes (breast cancer 1, early onset gene (*BRCA1*); breast cancer 2, early onset gene (*BRCA2*); BRCA1 interacting protein carboxy-terminal helicase 1 gene (*BRIP1*); checkpoint kinase 2 gene (*CHEK2*); partner and localizer of *BRCA2* gene (*PALB2*); phosphatase and tensin homolog gene (*PTEN*); and tumor protein p53 gene (*TP53*)) have been identified as breast cancer susceptibility genes. However, these explain only a small fraction of breast cancer cases in the general population (5). Since 2007, approximately 20 novel loci have been revealed by genome-wide association studies to be associated with breast cancer risk (6–15). Genetic variants in these loci are common but confer only small effects. However, it remains unclear if these genetic variants interact with environmental factors in the etiology of breast cancer.

Identification of interactions between genetic variants and environmental exposures or lifestyle factors will not only provide insight into the etiology of breast cancer but also may serve to identify women at higher risk of the disease that can benefit from targeted interventions. To date, 3 large-scale studies have been conducted to evaluate

interactions between genome-wide association study-identified single nucleotide polymorphisms (SNPs) and breast cancer risk factors (16–18). One evaluated 10 risk factors and 12 SNPs among 7,610 women in the Million Women Study (16). Another evaluated 9 established breast cancer risk factors and 17 SNPs among more than 20,000 women nested within the National Cancer Institute's Breast and Prostate Cancer Cohort Consortium (18). The largest study, using pooled data from 21 studies (17), evaluated 5 risk factors and 12 SNPs among more than 58,000 women. In all 3 reports, no significant interactions were observed after correction for the number of comparisons conducted. Two of these analyses used a Bonferroni-corrected *P* value threshold to determine statistical significance (16, 18), while the third used a parametric bootstrap test to accommodate the number of interactions evaluated (17). Notably, all 3 analyses were conducted among study populations that were either completely or predominantly of European ancestry, and all 3 evaluated each included genetic variant separately in their analyses. To our knowledge, no studies of interactions for breast cancer risk to date have utilized a cumulative genetic risk score (GRS) for the combined effect of multiple genetic variants on breast cancer risk. Further, differences in both the genetic architecture and lifestyle and reproductive characteristics exist between Asian and European women. As yet, no large-scale study of gene-environment interactions among Asian women has been reported. Therefore, we systematically evaluated multiplicative interactions between 11 established breast cancer risk factors and genetic variants previously shown to be associated with breast cancer risk by genome-wide association studies in a population-based, case-control study conducted among Chinese women. Of the 20 genome-wide association study variants evaluated, 10 variants in independent loci had significant or marginal associations with breast cancer risk in our study population and were used to construct the GRS.

MATERIALS AND METHODS

Study population

In the current analysis, a total of 6,477 breast cancer cases and 3,981 controls from the population-based, multistage, case-control Shanghai Breast Cancer Genetics Study were included. Previous reports have described the study population and enrollment process in detail (12, 19, 20). Briefly, breast cancer cases were identified via the Shanghai Cancer Registry; controls were randomly selected by using the Shanghai Resident Registry. Recruitment occurred between August 1996 and March 1998 and again between April 2002 and February 2005. Demographic and breast cancer-related information was obtained by in-person interviews and included age at menarche, age at menopause, age at first live birth and total months of breastfeeding among parous women, oral contraceptive use, hormone replacement therapy use, family history of breast cancer, and prior diagnosis of benign breast disease. Trained personnel measured all participants for weight, height, and circumferences of the waist and hips. The study was approved by the institutional review boards at all participating institutes, and all participants provided written, informed consent before participating in the study.

DNA extraction and genotyping

Laboratory protocols for the DNA extraction and genotyping methods used by the Shanghai Breast Cancer Genetics Study have been previously described in detail (12, 20). Briefly, genomic DNA was extracted from either buffy coat blood fractions or exfoliated buccal cells. Eight genetic variants (rs2180341, rs2046210, rs1219648, rs2981582, rs3817198, rs3803662, rs4784227, and rs8051542) were genotyped by using the Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara, California) among 5,242 included participants. These 8 variants, as well as 4 additional variants (rs11249433, rs4973768, rs999737, and rs6504950), were also genotyped by using the Sequenom iPLEX MassARRAY (Sequenom, San Diego, California) among 6,028 included participants. Five SNPs (rs13387042, rs10941679, rs889312, rs13281615, and rs12443621) were genotyped by using TaqMan allelic discrimination assays (Applied Biosystems, Life Technologies Corporation, Carlsbad, California) for 5,956 included participants. Genotypes for the 5 remaining SNPs (rs1011970, rs2380205, rs10995190, rs704010, and rs614367) were derived by imputation by use of MaCH 1.0 for 4,484 included participants (21). HapMap Han Chinese in Beijing, China (termed "CHB"), and Japanese in Tokyo, Japan (termed "JPT"), phase II (release 24) and phase III (release 2) samples were used as the reference for imputation (Coriell Institute for Medical Research, Camden, New Jersey); a minimum squared correlation between imputed and true genotypes (RSQ) of 0.3 was required to ensure quality imputation data. Quality-control protocols for genotyping assays have been previously described (22). Briefly, Affymetrix assays included 4 controls (3 positive and 1 negative) per 96-well plate. Genotyping data from the 3 quality-control samples, each genotyped approximately 45 times, showed an average concordance rate of 99.85%. Sequenom and TaqMan assays included 6 controls (2 positive controls in duplicate and 2 negative controls) per 96-well plate. Positive controls included blinded duplicate samples, as well as either European ($n = 60$) or Chinese ($n = 45$) HapMap DNA samples. Blinded duplicate samples had mean concordance rates of 99.6% on Sequenom assays and 96.7% on TaqMan assays. HapMap samples had mean concordance rates of 100% on Sequenom assays and 99.3% on TaqMan assays.

Statistical analysis

Means and standard deviations were computed for continuous variables; counts and proportions were computed for categorical variables. Associations with breast cancer risk were evaluated by using logistic regression to derive adjusted odds ratios and 95% confidence intervals. The established breast cancer risk factors evaluated included early age at menarche, late age at menopause, long duration of menstruation (years), late age at first live birth/parity, longer duration of breastfeeding, family history of breast

cancer among first-degree relatives, history of benign breast disease (fibroadenoma or lobular proliferation), hormone replacement therapy use, high body mass index, high waist/hip ratio, and lack of regular physical activity. Associations with breast cancer for genetic variants utilized gene-dose effects; genotypes were coded as having 0, 1, or 2 risk alleles, and allelic odds ratios and 95% confidence intervals were estimated. Associations with breast cancer for established risk factors and genetic variants were adjusted for age and education. A GRS was created to measure the cumulative effect of multiple genetic variants. The GRS was calculated by summing the number of risk alleles of variants that had a significant or marginally significant association with breast cancer risk in our study, weighted by the effect size of the association for each variant in the current analysis, and then scaled by a factor of 10. One variant from each independent locus was selected for inclusion in the creation of GRS; independence of loci was determined by linkage disequilibrium, using a minimum r^2 of 0.3. In total, 10 variants were included in the GRS (rs4973768, rs10941679, rs889312, rs2046210, rs13281615, rs704010, rs1219648, rs3817198, rs3803662, and rs4784227). As not all variants were genotyped among all women, the mean allelic count specific to each variant for either cases or controls was included when the GRS was constructed; however, if more than 5 GRS components were unavailable, no GRS was calculated, and the participant was not included in the interaction analyses. When not used as a continuous variable, GRS categorization was based on the distribution of the variable among controls. Effect measure modification on the multiplicative scale of GRS on breast cancer risk was evaluated for 11 breast cancer risk factors, by both continuous and categorical measures of GRS. All potential effect modifiers were dichotomized, with the lower breast cancer risk group given first: age at menarche (>13, ≤13 years), age at menopause (≤50, >50 years), duration of menstruation (≤30, >30 years), age at first live birth/parity (≤25, >25 years or nulliparous), duration of breastfeeding (>12, ≤12 months), family history of breast cancer among first-degree relatives (no, yes), history of benign breast disease (no, yes), hormone replacement therapy use (no, yes), body mass index (≤25, >25), waist/hip ratio (≤0.81, >0.81), and regular physical activity (yes, no). Interactions were evaluated by using likelihood ratio tests comparing nested models with only main effects and models with main effects plus the relevant interaction term. Statistical significance was defined by $P \leq 0.05$; marginal significance was defined by $P \leq 0.15$. All statistical tests were 2 sided, and all analyses were conducted with SAS, version 9.2, software (SAS Institute, Inc., Cary, North Carolina).

RESULTS

Associations with breast cancer risk for demographic characteristics and breast cancer risk factors are shown in Table 1. The mean age among cases (49.8 years) did not significantly differ from that of controls (50.1 years). Increased risks of breast cancer were associated with older ages at menopause, longer durations of menstruation, older ages at first live birth or not having given birth, having a first-degree relative with breast cancer, having a history of benign breast disease, and having a higher body mass index or waist/hip ratio; decreased risks of breast cancer were associated with older age at menarche, longer durations of breastfeeding, and regular physical activity. Notably, the association with hormone replacement therapy was not statistically significant in this analysis, most likely because of a low frequency of use in this study population. Thus, a total of 11 risk factors were selected for evaluation of interactions with genetic variants for breast cancer risk.

Genome-wide association studies have identified 22 SNPs that are associated with breast cancer risk (6–15). Associations with breast cancer risk for these SNPs among 10,458 Chinese women are shown in Table 2. At a nominal significance level ($P \leq 0.05$), 8 SNPs (rs4973768, rs2046210, rs1219648, rs2981582, rs3817198, rs3803662, rs4784227, and rs8051542) were associated with breast cancer risk. Five additional SNPs (rs10941679, rs889312, rs2180341, rs1328161, and rs704010) were associated at a marginal significance level ($P < 0.15$). With the exception of one (rs2180341/6q22.33), these variants had associations with breast cancer risk that were in the same direction as previously reported. Thus, 12 variants were considered for inclusion in constructing the GRS. Two SNPs (rs1219648 and rs2981582) in moderate linkage disequilibrium ($r^2 = 0.64$) were located in the fibroblast growth factor receptor 2 gene (*FGFR2*); rs1219648 was selected for inclusion in the GRS because of a higher minor allele frequency and stronger *P* value for association with breast cancer risk than rs2981582. Of 3 significantly associated SNPs located in the TOX high mobility group box family member 3 gene (*TOX3*), 2 (rs4784227 and rs8051542) share modest linkage disequilibrium ($r^2 = 0.40$); rs3803662 and rs4784227 were selected for inclusion in the GRS because of higher minor allele frequencies and stronger *P* values for their associations with breast cancer risk than rs8051542.

Of the 10,458 women in our initial analyses, sufficient genotyping information was available to construct a GRS for 6,408. For the 10 SNPs included in the GRS, the number of risk alleles weighted by the effect size for associations with breast cancer risk was summed and scaled by 10; values ranged from 1.04 to 21.2, with a mean of 9.02 and a median of 8.93 (data not shown). Breast cancer cases had significantly higher mean GRS values than controls ($P = 8.01 \times 10^{-29}$), such that each unit increase in GRS was associated with a 13% increased risk of breast cancer (odds ratio (OR) = 1.13, 95% confidence interval (CI): 1.11, 1.14) (Table 3). Compared with women with GRS values of ≤7.3, those with values of 7.3–9.7 had modest increases in breast cancer risk (OR = 1.38, 95% CI: 1.21, 1.56), while women with values ≥9.7 had a nearly 2-fold increased risk of breast cancer (OR = 1.89, 95% CI: 1.68, 2.14). This dose-response relation between GRS and increasing breast cancer risk was highly significant ($P = 4.65 \times 10^{-25}$).

Multiplicative effect measure modification of GRS by 11 breast cancer risk factors was evaluated, by using both continuous and categorical measures of GRS. All breast cancer risk factors were initially dichotomized.

Table 1. Breast Cancer Risk Factors and Breast Cancer Risk Among Chinese Women, the Shanghai Breast Cancer Genetics Study, 1996–1998 and 2002–2005

Characteristics and Risk Factors	Cases ($n = 6,477$)		Controls ($n = 3,981$)		OR[a]	95% CI	*P* Value
	No.	%	No.	%			
Demographic characteristics							
Age, >50 years	3,292	50.8	1,920	51.8	NA		9.9×10^{-3}*
Education, high school or higher	3,385	52.3	1,768	44.4	1.48*	1.37, 1.61	2.4×10^{-21}*
Reproductive risk factors							
Age at menarche, >13 years	4,342	67.1	2,865	72.0	0.80*	0.73, 0.87	8.2×10^{-19}*
Age at menopause, <50 years[b]	1,237	40.6	619	32.4	1.21*	1.07, 1.37	0.003*
Years of menstruation, >30	4,268	67.5	2,371	59.7	1.30*	1.19, 1.42	1.2×10^{-8}*
Parity/age at first live birth, >25 years or nulliparous	4,278	66.1	2,278	57.2	1.61*	1.48, 1.76	2.7×10^{-27}*
Months of breastfeeding, >12[c]	1,715	27.9	1,187	31.1	0.56*	0.50, 0.63	6.7×10^{-22}*
Medical history risk factors							
Family history of breast cancer	141	4.6	108	2.7	1.75*	1.36, 2.26	1.8×10^{-5}*
History of benign breast disease	1,345	44.2	1,204	30.2	1.73*	1.56, 1.91	7.2×10^{-27}*
Other risk factors							
Hormone replacement therapy[b]	212	7.0	105	5.5	1.27	0.99, 1.63	0.058
Body mass index, >25[d]	2,183	33.7	1,132	28.5	1.25*	1.14, 1.37	7.7×10^{-7}*
Waist/hip ratio, >0.81	413	63.6	1,964	49.4	1.76*	1.62, 1.91	1.8×10^{-40}*
Regular physical activity[e]	770	25.3	1,254	31.5	0.79*	0.71, 0.88	2.7×10^{-5}*

Abbreviations: CI, confidence interval; NA, not appropriate to estimate because of the frequency-matched case-control study design utilized; OR, odds ratio.

* $P \leq 0.05$ (significant).

[a] Derived from logistic regression for breast cancer risk, for women described, compared with remaining women, adjusted for age and education.

[b] Among postmenopausal women (3,048 cases and 1,910 controls).

[c] Among parous women (6,140 cases and 3,813 controls).

[d] Body mass index: weight (kg)/height $(m)^2$.

[e] Among 3,043 cases and 3,981 controls with data available for this variable.

Heterogeneity of the effect of GRS was suggested to vary by total years of menstruation ($P = 0.194$), such that the effect of GRS on breast cancer risk was stronger among women with 30 years or more of menstruation. However, neither this interaction nor any others reached a level of nominal statistical significance. Additional analyses, using breast cancer risk factors in tertiles, also indicated no evidence for interactions (Appendix Table 1). When the relation between GRS and total years of menstruation was further evaluated, one variant included in the GRS was found to have a heterogeneous effect by total years of menstruation (data not shown). Among women with fewer than 30 years' total duration of menstruation, the mitochondrial ribosomal protein S30 gene (*MRPS30*) rs10941679 had no effect on breast cancer risk (OR = 0.97, 95% CI: 0.86, 1.09), whereas, among women with 30 years or more of menstruation, this variant was associated with a significantly increased risk of breast cancer (OR = 1.15, 95% CI: 1.05, 1.26). Although this interaction reached a nominal significance level, it did not withstand correction for the number of variants evaluated ($P = 0.037$). Further, when women with nonnatural causes of menopause ($n = 339$) were excluded from this analysis, the interaction was also attenuated ($P = 0.059$).

DISCUSSION

In this large, population-based study, 22 genome-wide association study-identified genetic variants (6–15) were evaluated for associations with breast cancer risk among 10,458 Chinese women. Eight SNPs in 6 independent loci were found to be significantly associated with breast cancer, including 2 SNPs originally identified among Chinese women (rs4784227 and rs2046210) (12, 14). The cumulative effect of genetic variants on breast cancer susceptibility was examined by using a GRS that was constructed using 6 significant and 4 marginally significant SNPs, each reflecting an independent genetic locus, all with consistent directions of association between women of Chinese and European ancestry. As expected, cases had significantly higher GRS values than controls. Effect measure modification of GRS by 11 breast cancer risk factors was systematically evaluated. A multiplicative interaction with GRS was suggested for total years of

Table 2. Association with Breast Cancer Risk for 22 Genome-wide Association Study Single Nucleotide Polymorphisms Among Chinese Women, the Shanghai Breast Cancer Genetics Study, 1996–1998 and 2002–2005

SNP	Chromosomal Location	Gene/ Locus[a]	Data Source[b]	Alleles[c]	RAF, %		Association With Breast Cancer[d]			GWAS Reference[e]	Association Agreement[f]	Included in GRS
					Cases	Controls	OR	95% CI	*P* value			
rs11249433	1p11.2	*NOTCH2*	G1	A/G	3.3	2.7	1.20	0.93, 1.54	0.1682	Thomas, 2009 (13)	Yes	No
rs13387042	2q35	Unknown	G2	G/A	11.6	11.3	1.03	0.92, 1.15	0.6446	Stacey, 2007 (8)	Yes	No
rs4973768	3p24.1	*SLC4A7*	G1	C/T	20.3	17.8	1.16*	1.06, 1.28	0.0011*	Ahmed, 2009 (11)	Yes	Yes
rs10941679	5p12	*MRPS30*	G2	A/G	51.9	50.1	1.07	1.00, 1.16	0.0527	Stacey, 2008 (10)	Yes	Yes
rs889312	5q11.2	*MAP3K1*	G2	A/C	53.3	51.7	1.07	0.99, 1.15	0.0820	Easton, 2007 (6)	Yes	Yes
rs2180341	6q22.33	*ECHDC1*	G1, G3	G/A	75.3	74.1	1.07	0.98, 1.16	0.1344	Gold, 2008 (9)	No	No
rs2046210	6q25.1	*C6orf97*	G1, G3	G/A	41.9	36.4	1.27*	1.20, 1.35	1.15×10^{-15}*	Zheng, 2009 (12)	Yes	Yes
rs13281615	8q24.21	Unknown	G2	A/G	51.6	50.2	1.06	0.98, 1.14	0.1293	Easton, 2007 (6)	Yes	Yes
rs1011970	9p21.3	Unknown	I	G/T	9.5	9.0	1.05	0.91, 1.21	0.4808	Turnbull, 2010 (15)	Yes	No
rs2380205	10p15.1	Unknown	I	T/C	89.2	88.9	1.04	0.91, 1.18	0.6102	Turnbull, 2010 (15)	Yes	No
rs10995190	10q21.2	Unknown	I	A/G	98.1	97.8	1.16	0.86, 1.57	0.3382	Turnbull, 2010 (15)	Yes	No
rs704010	10q22.3	Unknown	I	C/T	31.8	30.3	1.07	0.98, 1.17	0.1482	Turnbull, 2010 (15)	Yes	Yes
rs1219648	10q26.13	*FGFR2*	G1, G3	A/G	42.2	38.9	1.14*	1.07, 1.21	1.36×10^{-5}*	Hunter, 2007 (7)	Yes	Yes
rs2981582	10q26.13	*FGFR2*	G1, G3	G/A	34.8	32.0	1.13*	1.06, 1.20	0.0001*	Easton, 2007 (6)	Yes	No
rs3817198	11p15.5	*LSP1*	G1, G3	T/C	13.2	12.3	1.10*	1.01, 1.19	0.0325*	Easton, 2007 (6)	Yes	Yes
rs614367	11q13.3	Unknown	I	C/T	0.2	0.1	1.96	0.02,167.5	0.7661	Turnbull, 2010 (15)	Yes	No
rs999737	14q24.1	*RAD51B*	G1	C/T	0.3	0.2	1.81	0.73, 4.48	0.2029	Thomas, 2009 (13)	No	No
rs12443621	16q12.1	*TOX3*	G2	A/G	57.4	57.2	1.01	0.94, 1.09	0.7431	Easton, 2007 (6)	No	No
rs3803662	16q12.1	*TOX3*	G1, G3	G/A	67.7	65.1	1.13*	1.06, 1.20	1.18×10^{-4}*	Easton, 2007 (6)	Yes	Yes
rs4784227	16q12.1	*TOX3*	G1, G3	C/T	28.6	24.1	1.27*	1.17, 1.38	9.47×10^{9}*	Long, 2010 (22)	Yes	Yes
rs8051542	16q12.1	*TOX3*	G1, G3	C/T	20.0	18.1	1.12*	1.04, 1.21	0.0020*	Easton, 2007 (6)	Yes	No
rs6504950	17q22	*COX11*	G1	G/A	8.0	7.7	1.04	0.91, 1.18	0.5863	Ahmed, 2009 (11)	Yes	No

Abbreviations: CI, confidence interval; GRS, genetic risk score; GWAS, genome-wide association study; OR, odds ratio; RAF, risk allele frequency; SNP, single nucleotide polymorphism.

* $P \leq 0.05$ (significant).

[a] *C6orf97*, chromosome 6 open reading frame 97 gene; *COX11*, cytochrome *c* oxidase assembly homolog 11 (yeast) gene; *ECHDC1*, enoyl CoA hydratase domain containing 1 gene; *FGFR2,* fibroblast growth factor receptor 2 gene; *LSP1*, lymphocyte-specific protein 1 gene; *MAP3K1*, mitogen-activated protein kinase kinase kinase 1, E3 ubiquitin protein ligase, gene; *MRPS30*, mitochondrial ribosomal protein S30 gene; *NOTCH2*, notch 2 gene; *RAD51B,* the RAD51 homolog B (*Saccharomyces cerevisiae*) gene; *SLC4A7*, the solute carrier family 4, sodium bicarbonate cotransporter, member 7 gene; *TOX3,* TOX high mobility group box family member 3 gene.

[b] Data source: G1, genotyped by Sequenom, Inc. (San Diego, California), among 6,028 Chinese women; G2, genotyped by TaqMan (Applied Biosystems, Life Technologies Corporation, Carlsbad, California), among 5,956 Chinese women; G3, genotyped by Affymetrix (Santa Clara, California) among 5,242 Chinese women; or I, imputed by MACH (a Markov chain-based haplotyper) (21) for 4,484 Chinese women.

[c] Reference allele/risk allele among controls.

[d] Derived from logistic regression, adjusted for age and education, calculated for increasing risk allele compared with reference allele; P_{trend}.

[e] First author, year (reference no.).

[f] Agreement between increasing or decreasing risk alleles between current analysis and GWAS reference.

Table 3. Interaction Analyses of Genetic Risk Score and Breast Cancer Risk Factors Among 6,408 Chinese Women, the Shanghai Breast Cancer Genetics Study, 1996–1998 and 2002–2005

Characteristic or Risk Factor	Genetic Risk Score[a]									
	Continuous			By Tertiles						
				<7.3		7.3–9.7		>9.7		
	OR[b]	95% CI	*P* value[c]	OR[b]	95% CI	OR[b]	95% CI	OR[b]	95% CI	*P* value[c]
All women	1.13*	1.11, 1.14	8.01×10^{-29}*	1.00	Referent	1.38*	1.21, 1.56	1.89*	1.68, 2.14	4.65×10^{-25}*
Age at menarche, years										
>13	1.10*	1.08, 1.12	0.743	1.00	Referent	1.35*	1.16, 1.58	1.90*	1.64, 2.20	0.893
≤13	1.11*	1.07, 1.14		1.23*	1.00, 1.50	1.77*	1.45, 2.14	2.30*	1.92, 2.77	
Age at menopause, years[d]										
≤50	1.09*	1.06, 1.13	0.768	1.00	Referent	1.11	0.88, 1.41	1.55*	1.24, 1.93	0.343
>50	1.10*	1.05, 1.15		1.03	0.76, 1.39	1.34	1.00, 1.79	1.93*	1.44, 2.57	
Years of menstruation										
30	1.09*	1.06, 1.12	0.194	1.00	Referent	1.35*	1.10, 1.65	1.77*	1.45, 2.15	0.301
>30	1.11*	1.09, 1.14		1.16	0.95, 1.41	1.63*	1.35, 1.99	2.30*	1.91, 2.78	
Parity/age at first live birth, years										
≤25	1.12*	1.09, 1.15	0.570	1.00	Referent	1.12	0.92, 1.37	1.79*	1.48, 2.17	0.644
>25 or nulliparous	1.11*	1.08, 1.13		1.12	0.93, 1.35	1.76*	1.46, 2.11	2.19*	1.83, 2.62	
Total breastfeeding, months										
>12 total	1.10*	1.06, 1.13	0.621	1.00	Referent	1.18	0.91, 1.51	1.73*	1.36, 2.21	0.456
≤12	1.11*	1.08, 1.13		1.18	0.94, 1.47	1.72*	1.37, 2.15	2.31*	1.85, 2.87	
Family history of breast cancer										
No	1.10*	1.08, 1.12	0.897	1.00	Referent	1.33*	1.16, 1.52	1.86*	1.64, 2.11	0.811
Yes	1.10*	1.01, 1.21		2.02*	1.18, 3.48	1.65*	1.02, 2.68	3.29*	2.13, 5.07	
History of benign breast disease										
No	1.09*	1.07, 1.12	0.789	1.00	Referent	1.28*	1.08, 1.51	1.80*	1.53, 2.10	0.824
Yes	1.11*	1.07, 1.13		1.66*	1.36, 2.02	2.22*	1.84, 2.70	3.09*	2.59, 3.70	
Hormone replacement therapy										
No	1.10*	1.08, 1.12	0.836	1.00	Referent	1.38*	1.22, 1.57	1.91*	1.68, 2.16	0.401
Yes	1.10	1.00, 1.21		1.53	0.92, 2.56	1.76*	1.07, 2.89	2.24*	1.47, 3.43	
Body mass index[e]										
≤25	1.10*	1.08, 1.13	0.735	1.00	Referent	1.49*	1.28, 1.73	1.86*	1.61, 2.15	0.303
>25	1.11*	1.07, 1.14		1.36*	1.11, 1.66	1.58*	1.30, 1.93	2.81*	2.32, 3.40	

Table continues

Table 3. Continued

Characteristic or Risk Factor	Genetic Risk Score[a]									
	Continuous			By Tertiles						
				<7.3		7.3–9.7		>9.7		
	OR[b]	95% CI	*P* value[c]	OR[b]	95% CI	OR[b]	95% CI	OR[b]	95% CI	*P* value[c]
Waist/hip ratio										
≤0.81	1.11*	1.08, 1.14	0.779	1.00	Referent	1.49*	1.23, 1.80	2.00*	1.66, 2.40	0.570
>0.81	1.10*	1.08, 1.13		1.75*	1.45, 2.11	2.33*	1.93, 2.81	3.24*	2.70, 3.89	
Regular physical activity[f]										
Yes	1.11*	1.08, 1.13	0.990	1.00	Referent	1.00	0.77, 1.28	1.76*	1.39, 2.22	0.682
No	1.10*	1.08, 1.12		1.17	0.95, 1.45	1.70*	1.37, 2.09	2.24*	1.82, 2.74	

Abbreviations: CI, confidence interval; OR, odds ratio.

* $P \leq 0.05$ (significant).

[a] Genetic risk score based on genotypes for rs4973768, rs10941679, rs889312, rs2046210, rs13281615, rs704010, rs1219648, rs3817198, rs3803662, and rs4784227.

[b] Derived from logistic regression, adjusted for age and education.

[c] *P* value for association or trend for genetic risk score among all women; *P* values for interaction in stratified analyses.

[d] Among postmenopausal women.

[e] Body mass index: weight (kg)/height $(m)^2$.

[f] Among women from case-control studies.

menstruation, and a nominally significant interaction was found for one variant (*MRPS30* rs10941679). However, this interaction was no longer statistically significant after correction for multiple comparisons or exclusion of women with nonnatural causes of menopause. These results are similar to those reported by previous studies conducted among women of European ancestry.

Most of the established breast cancer risk factors evaluated among Chinese women in the current study had strong associations with breast cancer risk. An exception was hormone replacement therapy use, which was not significantly associated with breast cancer risk in this study population, most likely because of a low frequency of use. To be comprehensive, hormone replacement therapy was still evaluated as a potential effect modifier of genetic risk on breast cancer susceptibility. In order to provide a cumulative measure of the effect of genetic variants on breast cancer risk, a GRS was constructed, which was also very strongly associated with breast cancer risk. Use of a GRS also served to minimize the number of statistical comparisons made during our analysis. Generally, in case-control studies, a potential gene-environment interaction, or epidemiologic effect measure modification, is assessed by comparing nested models that include the genetic and environmental risk factor in one model and these main effects plus the product of the 2 factors in a second model (23). Although alternate approaches for interaction analyses have been suggested and used (23–25), they are most suitable for genome-wide data and very large-scale analyses. Given that only 22 variants were considered in the current analysis and only 10 were actually evaluated for interactions, a more traditional analytical approach was maintained. Multiplicative interactions with 11 breast cancer risk factors were first evaluated by using a GRS, and then for one potential interaction, each of the 10 included genetic variants was evaluated. However, even using this approach, we found that the suggested interaction between *MRPS30* rs10941679 and duration of menstruation did not maintain statistical significance after considering the number of variants evaluated.

To date, 3 large studies have evaluated interactions between genome-wide association study-identified genetic variants and traditional breast cancer risk factors; none found significant interactions after correction for the number of comparisons made, and all were conducted among completely or predominantly European-ancestry study populations (16–18). However, other studies have reported evidence that is generally supportive of gene-environment interactions for breast cancer risk, most abundantly for variants in *FGFR2* (26–28). A significant interaction between *FGFR2* rs1219648 and hormone replacement therapy was found among European-American women (26). A small study of Japanese women found that the risk of breast cancer associated with *FGFR2* rs2420946 varied by age at menarche ($P = 0.019$), parity ($P = 0.026$), and family history of breast cancer ($P = 0.003$) (27). A small study of Chinese women found that the combined effect of 3 *FGFR2* SNPs was greater among women with later age at first live birth compared with women with a first live birth by age 25; however, the test for an additive interaction was not statistically significant (29). Finally, among a subset of the current study population, significant

interactions with oral contraceptive use were observed for 12 linked *FGFR2* SNPs, such that increased breast cancer risk for the variants was observed only for women with oral contraceptive use (28). Most relevant to the current study, using data from the Women's Health Initiative, Huang et al. (25) analyzed nearly 5,000 SNPs for interactions with the hormone and dietary intervention arms of the trial; SNPs in the *MRPS30* region of chromosome 5p12 had highly suggestive evidence for interactions with diet modification and vitamin intake.

Several key differences between our study and prior studies deserve attention. First, we created a composite variable, GRS, rather than evaluate all genetic variants independently. This served to focus our analysis and limit the number of statistical comparisons conducted. A limitation of this approach is that interactions with variants comprising specific biologic pathways were not evaluated. However, current biologic understanding of the pathways implicated by genome-wide association study-identified variants for breast cancer risk is still limited. Second, we also created composite variables for reproductive risk factors, such as total years of menstruation, and parity/age at first live birth. This was to maximize the effect of reproductive factors on breast cancer risk and also to reduce the number of comparisons made. Third, we did not conduct tests for all the possible interactions between the genetic variants and breast cancer risk factors evaluated. We first evaluated 11 interactions with GRS and then tested 10 genetic variants for their contributions to one possible interaction. However, even with this approach, our one nominal association did not maintain significance after considering the number of tests conducted. It is possible that the current approaches and methods to identify interactions between genetic variants and breast cancer risk factors are not optimal. For a genetic variant with a minor allele frequency of 0.10, an additive OR of 1.2, and a binary environmental exposure with a main effect of 1.2 to detect an interaction with 80% power would require over 50,000 cases and controls each.

In conclusion, the effect of the GRS was most heterogeneous by total years of menstruation, and a nominally significant multiplicative interaction was found for *MRPS30* rs10941679, such that an increased risk of breast cancer was evident only among women with more than 30 years of menstruation. In agreement with other large studies conducted among predominantly European-ancestry study populations, in this large case-control study of Chinese women, the significance of our findings did not withstand correction for multiple comparisons. Therefore, either gene-environment interactions are unlikely to play a large role in breast cancer susceptibility, or current approaches to identify such interactions are insufficient, and new methodologies are needed.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology and Biostatistics, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China (Haixin Li, Kexin Chen); Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee (Alicia Beeghly-Fadiel, Wanqing Wen, Qiuyin Cai, Jirong Long, Jiajun Shi, Xiao Ou Shu, Wei Zheng); Shanghai Municipal Center for Disease Control and Prevention, Shanghai, China (Wei Lu, Ying Zheng); and Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China (Yu-Tang Gao, Yong-Bing Xiang).

Authors Haixin Li and Alicia Beeghly-Fadiel contributed equally to this work.

This research was supported in part by US National Institutes of Health grants R01 CA064277, R01 CA124558, R01 CA148667, and R37 CA070867 (to W. Z.) and grants R01 CA118229 and R01 CA092585 (to X. O. S.). This work was conducted while H. X. L. was a visiting research fellow trainee at the Vanderbilt Epidemiology Center and supported by a Fogarty International Institute grant D43 TW008313 to X. O. S. Sample preparation and genotyping assays at Vanderbilt were conducted at the Survey and Biospecimen Shared Resources and Vanderbilt Microarray Shared Resource, which are supported in part by the Vanderbilt-Ingram Cancer Center (P30 CA068485).

The authors thank the research staff for their contributions and commitment to this project, Regina Courtney for DNA preparation, and Samantha Stansel for clerical support in the preparation of this manuscript.

The content is solely the responsibility of the authors and does not necessarily represent the official views of any listed funding agencies.

Conflict of interest: none declared.

REFERENCES

1. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer: analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med*. 2000;343(2):78–85.
2. Feigelson HS, Henderson BE. Estrogens and breast cancer. *Carcinogenesis*. 1996;17(11):2279–2284.
3. Baselga J, Norton L. Focus on breast cancer. *Cancer Cell*. 2002;1(4):319–322.
4. Colditz GA, Baer HJ, Tamini RM. Breast cancer. In: Schottenfeld D, Fraumeni Joseph F, eds. *Cancer Epidemiology and Prevention*. New York, NY: Oxford University Press; 2011.
5. Zhang B, Beeghly-Fadiel A, Long J, et al. Genetic variants associated with breast-cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. *Lancet Oncol*. 2011;12(5):477–488.
6. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*. 2007;447(7148):1087–1093.
7. Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer. *Nat Genet*. 2007;39(7):870–874.
8. Stacey SN, Manolescu A, Sulem P, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet*. 2007; 39(7):865–869.

9. Gold B, Kirchhoff T, Stefanov S, et al. Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. *Proc Natl Acad Sci U S A*. 2008; 105(11):4340–4345.
10. Stacey SN, Manolescu A, Sulem P, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet*. 2008;40(6): 703–706.
11. Ahmed S, Thomas G, Ghoussaini M, et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet*. 2009;41(5):585–590.
12. Zheng W, Long J, Gao YT, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet*. 2009;41(3):324–328.
13. Thomas G, Jacobs KB, Kraft P, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (*RAD51L1*). *Nat Genet*. 2009;41(5):579–584.
14. Long J, Cai Q, Shu XO, et al. Identification of a functional genetic variant at 16q12.1 for breast cancer risk: results from the Asia Breast Cancer Consortium. *PLoS Genet*. 2010; 6:pe1001002. (doi:10.1371/journal.pgen.1001002).
15. Turnbull C, Ahmed S, Morrison J, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet*. 2010;42(6):504–507.
16. Travis RC, Reeves GK, Green J, et al. Gene-environment interactions in 7610 women with breast cancer: prospective evidence from the Million Women Study. *Lancet*. 2010; 375(9732):2143–2151.
17. Milne R, Gaudet M, Spurdle A, et al. Assessing interactions between the associations of common genetic susceptibility variants, reproductive history and body mass index with breast cancer risk in the Breast Cancer Association Consortium: a combined case-control study. *Breast Cancer Res*. 2010;12:pR110. (doi:10.1186/bcr2797).
18. Campa D, Kaaks R, Le Marchand L, et al. Interactions between genetic variants and breast cancer risk factors in the Breast and Prostate Cancer Cohort Consortium. *J Natl Cancer Inst*. 2011;103(16):1252–1263.
19. Gao YT, Shu XO, Dai Q, et al. Association of menstrual and reproductive factors with breast cancer risk: results from the Shanghai Breast Cancer Study. *Int J Cancer*. 2000;87(2): 295–300.
20. Zheng W, Wen W, Gao YT, et al. Genetic and clinical predictors for breast cancer risk assessment and stratification among Chinese women. *J Natl Cancer Inst*. 2010; 102(13):972–981.
21. Li Y, Willer CJ, Ding J, et al. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol*. 2010;34(8):816–834.
22. Long J, Shu XO, Cai Q, et al. Evaluation of breast cancer susceptibility loci in Chinese women. *Cancer Epidemiol Biomarkers Prev*. 2010;19(9):2357–2365.
23. Thomas D. Gene-environment-wide association studies: emerging approaches. *Nat Rev Genet*. 2010;11(4):259–272.
24. Murcray CE, Lewinger JP, Gauderman WJ. Gene-environment interaction in genome-wide association studies. *Am J Epidemiol*. 2009;169(2):219–226.
25. Huang Y, Ballinger D, Dai J, et al. Genetic variants in the *MRPS30* region and postmenopausal breast cancer risk. *Genome Med*. 2011;3(6):p42. (doi:10.1186/gm258).
26. Rebbeck TR, DeMichele A, Tran TV, et al. Hormone-dependent effects of *FGFR2* and *MAP3K1* in breast cancer susceptibility in a population-based sample of post-menopausal African-American and European-American women. *Carcinogenesis*. 2009;30(2):269–274.
27. Kawase T, Matsuo K, Suzuki T, et al. *FGFR2* intronic polymorphisms interact with reproductive risk factors of breast cancer: results of a case control study in Japan. *Int J Cancer*. 2009;125(8):1946–1952.
28. Xu WH, Shu XO, Long J, et al. Relation of *FGFR2* genetic polymorphisms to the association between oral contraceptive use and the risk of breast cancer in Chinese women. *Am J Epidemiol*. 2011;173(8):923–931.
29. Liang J, Chen P, Hu Z, et al. Genetic variants in fibroblast growth factor receptor 2 (FGFR2) contribute to susceptibility of breast cancer in Chinese women. *Carcinogenesis*. 2008; 29(12):2341–2346.

(Appendix follows)

Appendix Table 1. Additional Interaction Analyses of Genetic Risk Scores and Breast Cancer Risk Factors Among 6,408 Chinese Women, the Shanghai Breast Cancer Genetics Study, 1996–1998 and 2002–2005

Characteristic or Risk Factor	Genetic Risk Score[a]							
	Continuous			Dichotomized				
				<8.5		≥8.5		
	OR[b]	95% CI	*P* Value	OR[b]	95% CI	OR[b]	95% CI	*P* Value[c]
All women	1.13*	1.11, 1.14	8.01×10^{-29}*	1.00	Referent	1.58*	1.43, 1.75	$<1.1 \times 10^{-16}$
Years of menstruation								
≤25	1.10*	1.05, 1.15	0.603	1.00	Referent	1.52*	1.15, 2.00	0.834
>25–≤35	1.11*	1.08, 1.13		1.31*	1.04, 1.66	2.12*	1.68, 2.67	
>35	1.11*	1.07, 1.15		1.51*	1.14, 1.99	2.37*	1.81, 3.11	
Parity/age at first live birth, years								
≤25	1.10*	1.07, 1.13	0.754	1.00	Referent	1.50*	1.28, 1.76	0.679
>25	1.12*	1.09, 1.14		1.21*	1.03, 1.41	2.01*	1.73, 2.33	
Nulliparous	1.02	0.94, 1.10		1.65*	1.15, 2.36	2.24*	1.59, 3.14	
Total breastfeeding, months								
>12 (total)	1.10*	1.06, 1.13	0.254	1.00	Referent	1.53*	1.25, 1.87	0.351
>6–≤12	1.09*	1.06, 1.12		1.33*	1.09, 1.63	2.00*	1.64, 2.44	
≤6	1.12*	1.09, 1.15		1.25*	1.01, 1.54	2.16*	1.77, 2.65	
Body mass index[d]								
≤18.5	1.17*	1.07, 1.27	0.780	1.00	Referent	1.54	0.95, 2.49	0.473
>18.5–≤25	1.10*	1.08, 1.12		1.11	0.76, 1.61	1.73*	1.19, 2.51	
>25	1.11*	1.07, 1.14		1.38	0.94, 2.04	2.32*	1.58, 3.43	
Waist/hip ratio								
≤0.78	1.13*	1.09, 1.17	0.500	1.00	Referent	1.71*	1.39, 2.09	0.874
>0.78–≤0.83	1.09*	1.06, 1.12		1.57*	1.28, 1.91	2.35*	1.94, 2.85	
>0.83	1.11*	1.08, 1.14		2.16*	1.77, 2.63	3.58*	2.94, 4.35	

Abbreviations: CI, confidence interval; OR, odds ratio.

* $P \leq 0.05$ (significant).

[a] Genetic risk score based on genotypes for rs4973768, rs10941679, rs889312, rs2046210, rs13281615, rs704010, rs1219648, rs3817198, rs3803662, and rs4784227.

[b] Derived from logistic regression, adjusted for age and education.

[c] *P* value for association or trend for genetic risk score among all women; *P* values for interaction in stratified analyses.

[d] Body mass index: weight (kg)/height $(m)^2$.