Human Molecular Genetics, 2011, Vol. 20, No. 24 ***4991–4999***
doi:10.1093/hmg/ddr405
Advance Access published on September 9, 2011

Genome-wide association study identifies breast cancer risk variant at 10q21.2: results from the Asia Breast Cancer Consortium

Qiuyin Cai[1], Jirong Long[1], Wei Lu[3], Shimian Qu[1], Wanqing Wen[1], Daehee Kang[4,5,6], Ji-Young Lee[5], Kexin Chen[8], Hongbing Shen[9], Chen-Yang Shen[10], Hyuna Sung[4], Keitaro Matsuo[11], Christopher A. Haiman[12], Ui Soon Khoo[13], Zefang Ren[14], Motoki Iwasaki[18], Kai Gu[3], Yong-Bing Xiang[19], Ji-Yeob Choi[4], Sue K. Park[4,5,6], Lina Zhang[8], Zhibin Hu[9], Pei-Ei Wu[10], Dong-Young Noh[6,7], Kazuo Tajima[11], Brian E. Henderson[12], Kelvin Y.K. Chan[13], Fengxi Su[15], Yoshio Kasuga[20], Wenjing Wang[3], Jia-Rong Cheng[19], Keun-Young Yoo[5], Jong-Young Lee[21], Hong Zheng[8], Yao Liu[9], Ya-Lan Shieh[10], Sung-Won Kim[22], Jong Won Lee[23], Hiroji Iwata[24], Loic Le Marchand[25], Sum Yin Chan[13], Xiaoming Xie[16], Shoichiro Tsugane[18], Min Hyuk Lee[26], Shenming Wang[17], Guoliang Li[1], Shawn Levy[27], Bo Huang[1], Jiajun Shi[1], Ryan Delahanty[1], Ying Zheng[3], Chun Li[2], Yu-Tang Gao[19], Xiao-Ou Shu[1] and Wei Zheng[1,*]

[1]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center and [2]Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, TN, USA, [3]Shanghai Municipal Center for Disease Control and Prevention, Shanghai, China, [4]Department of Biomedical Sciences, [5]Department of Preventive Medicine, [6]Cancer Research Institute and [7]Department of Surgery, Seoul National University College of Medicine, Seoul, Korea, [8]Department of Epidemiology and Biostatistics, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China, [9]Department of Epidemiology and Biostatistics, Nanjing Medical University, Nanjing, China, [10]Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan, [11]Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan, [12]Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA, USA, [13]Department of Pathology, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong SAR, China, [14]School of Public Health, [15]Second Affiliated Hospital, [16]Cancer Center and [17]First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, [18]Epidemiology and Prevention Division, Research Center for Cancer Prevention and Screening, National Cancer Center, Tokyo, Japan, [19]Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China, [20]Department of Surgery, Nagano Matsushiro General Hospital, Nagano, Japan, [21]Center for Genome Science, Korea National Institute of Health, KCDC, Osong Health Technology Administration Complex, Chungcheongbuk-do, Korea, [22]Department of Surgery, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Korea, [23]Department of Surgery, Ulsan University College of Medicine, Seoul, Korea, [24]Department of Breast Oncology, Aichi Cancer Center Central Hospital, Nagoya, Japan, [25]Epidemiology Program, Cancer Research Center, University of Hawaii, Honolulu, HI, USA, [26]Department of Surgery, Soonchunhyang University Hospital, Korea and [27]HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA

Received July 7, 2011; Revised and Accepted September 5, 2011

*To whom correspondence should be addressed at: Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, 2525 West End Avenue, 8th Floor, TN 37203-1738, Nashville, USA. Tel: +1 6159360682; Fax: +1 6159368241; Email: wei.zheng@vanderbilt.edu

Although approximately 20 common genetic susceptibility loci have been identified for breast cancer risk through genome-wide association studies (GWASs), genetic risk variants reported to date explain only a small fraction of heritability for this common cancer. We conducted a four-stage GWAS including 17 153 cases and 16 943 controls among East-Asian women to search for new genetic risk factors for breast cancer. After analyzing 684 457 SNPs in 2062 cases and 2066 controls (Stage I), we selected for replication among 5969 Chinese women (4146 cases and 1823 controls) the top 49 SNPs that had neither been reported previously nor were in strong linkage disequilibrium with reported SNPs (Stage II). Three SNPs were further evaluated in up to 13 152 Chinese and Japanese women (6436 cases and 6716 controls) (Stage III). Finally, two SNPs were evaluated in 10 847 Korean women (4509 cases and 6338 controls) (Stage IV). SNP rs10822013 on chromosome 10q21.2, located in the zinc finger protein 365 (*ZNF365*) gene, showed a consistent association with breast cancer risk in all four stages with a combined per-risk allele odds ratio of 1.10 (95% CI: 1.07–1.14) (*P*-value for trend = 5.87×10^{-9}). *In vitro* electrophoretic mobility shift assays demonstrated the potential functional significance of rs10822013. Our results strongly implicate rs10822013 at 10q21.2 as a genetic risk variant for breast cancer among East-Asian women.

INTRODUCTION

Breast cancer, one of the most common malignancies among women worldwide, is a complex polygenic disorder for which genetic factors play a significant role in disease etiology (1,2). To date, approximately 20 loci have been associated with breast cancer risk in genome-wide association studies (GWASs) (3–16). With the exception of our study conducted among Chinese women (7,11), all other published GWASs were conducted among women of European ancestry. Only about half of SNPs initially identified in women with European ancestry can be directly replicated in Chinese women, and further, the association with the replicated SNPs is weaker in Chinese women than in women of European ancestry (17,18). Because linkage disequilibrium (LD) structure and allele frequencies of genetic variations differ between women of Chinese and European ancestry, additional risk loci or additional genetic variants in previously identified loci may remain to be discovered by studies conducted among Chinese and other Asian women.

We recently analyzed 684 457 SNPs in 2062 breast cancer cases and 2066 community controls, recruited as part of the Shanghai Breast Cancer Study (SBCS) (7) to identify novel susceptibility loci for breast cancer. We selected the top 49 SNPs that had not been reported previously and that were not in strong LD with any reported SNPs for a fast-track replication conducted through the Asia Breast Cancer Consortium (Table 1 and Fig. 1). By analyzing data from 17 153 cases and 16 943 controls included in the consortium, we found strong evidence for a genetic variant that may contribute to breast cancer susceptibility among East-Asian women.

RESULTS

Of the 49 successfully genotyped SNPs in Stage II (Supplementary Material, Table S1), highly significant associations with breast cancer risk were found for SNPs rs10822013 (10q21.2) (Table 2) and rs2048672 (7q32.3) (Supplementary Material, Table S2). These two SNPs were selected for further validation in Stage III, which included 6436 cases and 6716 controls of Asian ancestry from eight studies participating in the Asia Breast Cancer Consortium (Table 2 and Supplementary Material, Table S2). A third SNP, rs17823421 (16p13.3), also had an association with breast cancer risk at $P < 0.05$ in Stage II (Supplementary Material, Table S3). This SNP, however, was not replicated in three studies included in Stage III; thus, further replication was not conducted.

With the exception of two small studies, the minor T allele of rs10822013 had a consistent positive association with breast cancer risk in the other six studies (Fig. 2, $P = 0.63$ for the heterogeneity test in Stage III). Adjusted odds ratios (ORs) for Stage III were 1.12 [95% confidence interval (CI): 1.03–1.21] and 1.14 (95% CI: 1.03–1.26), respectively, for the CT and TT genotypes (*P*-value for trend = 8.08×10^{-3}) (Table 2). Similarly, the minor T allele of rs10822013 had a consistent positive association with breast cancer risk in all three Stage-IV studies, which were conducted among Korean women (Fig. 2). Adjusted ORs for Stage IV were 1.10 (95% CI: 0.99–1.22) and 1.21 (95% CI: 1.07–1.37), respectively, for the CT and TT genotypes (*P*-value for trend = 2.07×10^{-3}) (Table 2). Pooled analyses of samples from all stages produced ORs of 1.12 (95% CI: 1.06–1.18) and 1.21 (95% CI: 1.13–1.29) for the CT and TT genotypes, respectively (*P*-value for trend = 5.87×10^{-9}) (Table 2). The heterogeneity test for results among all studies was not statistically significant ($P = 0.4817$).

Stratified analyses suggested that the associations of SNP rs10822013 with breast cancer risk were similar among pre- and post-menopausal women (Table 3). No apparent difference was observed between estrogen receptor (ER) (+) and ER (−) breast cancer in relation to rs10822013 (Table 3).

The minor C allele of rs2048672 was associated with an elevated risk of breast cancer in all but one small study included in Stage III (Supplementary Material, Fig. S1). Although none of the study-specific ORs in Stage III was statistically significant, pooling data from Stage III yielded ORs of 1.08 (95% CI: 1.00–1.17) and 1.10 (95% CI: 0.99–1.21), respectively, for the AC and CC genotypes (*P*-value for trend = 0.0545). The minor C allele of rs2048672 was associated with an elevated risk of breast cancer in all three studies included in Stage IV (Supplementary Material, Fig. S1) with pooled ORs of 1.05

Table 1. Selected characteristics of studies participating in the Asia Breast Cancer Consortium

Study/stage	Case	Control	Ethnicity	Study design[a]	Study period	Age (mean)	Menopause (%)	ER (+)[b] (%)
Stage I	2066	2075	Chinese	Population	1996–2005	49.3/49.4	38.7/41.6	63.6
Stage II	4374	1892	Chinese	Population	1996–2005	54.0/52.8[c]	51.1/55.4[c]	64.3
Stage III	6489	6800						
Tianjin	1515	1581	Chinese	Hospital	2004–2008	51.7/51.9	51.7/55.4	44.1
Nanjing	1439	1437	Chinese	Hospital	2004–2008	51.4/51.3	53.3/52.6	54.8
Taiwan	1057	1059	Chinese	Hospital	2004–2007	51.6/47.4[c]	52.4/39.7[c]	66.2
Hong Kong	432	631	Chinese	Hospital	2003–2009	45.9/45.6	52.0/41.2[c]	70.2
Guangzhou	466	518	Chinese	Hospital	2008–2009	47.6/47.6	42.5/47.6	69.8
Nagoya	643	639	Japanese	Hospital	2003–2005	51.4/51.1	48.6/48.6	72.8
Hawaii (MEC)	536	534	Japanese	Population	1993–2008	65.1/60.3[c]	86.7/82.7	85.4
Nagano	401	401	Japanese	Hospital	2001–2005	53.8/54.0	54.9/65.3[c]	74.6
Stage IV	4516	6344						
SeBCS-I	2359	2052	Korean	Hospital	1995–2006	48.1/51.7	37.9/52.0	61.9
SeBCS-II	768	1098	Korean	Hospital	1995–2006	47.5/47.7	36.5/37.2	62.6
KOHBRA/KoGES	1389	3194	Korean	Hospital	2007–2009	40.5/50.3[c]	23.3/NA	62.8
Total	17 445	17 111						

MEC, Multiethnic Cohort Study; SeBCS-I, Seoul Breast Cancer Study Phase I; SeBCS-II, Seoul Breast Cancer Study Phase II; KOHBRA, Korean Hereditary Breast Cancer Study; KoGES, Korean Genome and Epidemiology Study.
[a]With the exception of the MEC (a cohort study), all other studies used the case–control study design with either a population-based or hospital-based approach.
[b]Among cases with ER data.
[c]Significant at $\alpha = 0.01$ level (t-test for continuous variables, χ^2-test for categorical variables).

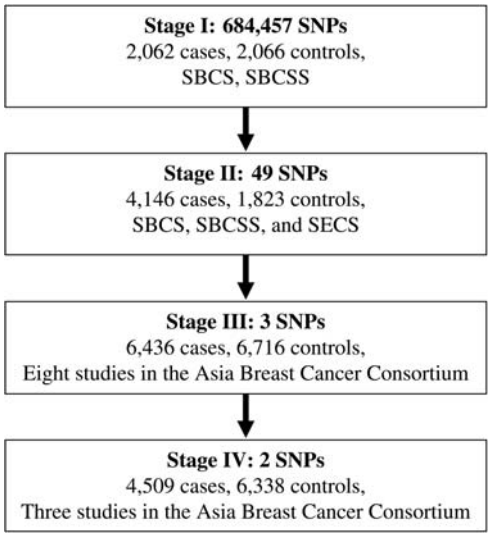

Figure 1. Overview of the study design.

(95% CI: 0.95–1.17) and 1.09 (95% CI: 0.97–1.24), respectively, for the AC and CC genotypes (P-value for trend = 0.1445). Pooled ORs for all stages were 1.11 (95% CI: 1.05–1.17) and 1.15 (95% CI: 1.08–1.23) for the AC and CC genotypes, respectively (P-value for trend $= 6.21 \times 10^{-6}$) (Supplementary Material, Table S2). The heterogeneity test for results among all studies was not statistically significant ($P = 0.5519$).

SNP rs10822013 is located on 10q21.2 (Fig. 3), a region where a genetic risk variant for breast cancer (rs10995190) was reported recently in a study conducted among women of European ancestry (12). These two SNPs are ~26.7 kb apart and are located in different LD blocks in Chinese (CHB, $r^2 = 0.000$), European (CEU, $r^2 = 0.176$) and African (YRI, $r^2 = 0.005$) populations (Supplementary Material, Fig. S2). Furthermore, rs10995190 has a very low minor allele frequency (MAF) in Chinese (2%) and showed no association with breast cancer risk in our study (OR = 0.84, 95%: 0.63–1.12, $P = 0.23$). Because SNP rs10995190 is not included in the Affymetrix Genome-Wide Human SNP Array 6.0, the genotype frequencies for cases and controls included in the Stage I were imputed (RSQR = 0.9154) using the program MACH (www.sph.umich.edu/csg/abecasis/mach) described in detail previously (7,11). Another SNP (rs16917302) in this locus was also recently associated with breast cancer risk among *BRCA2* mutation carriers, although the P-value (3.8×10^{-5}) was not genome-wide significant (15). SNP rs16917302 is ~9.2 kb away from rs10822013 and these two SNPs are located in different LD blocks in Chinese (CHB, $r^2 = 0.271$), European (CEU, $r^2 = 0.108$.) and African (YRI, $r^2 = 0.075$) populations (Supplementary Material, Fig. S2). SNP rs16917302 was associated with breast cancer risk in the Stage I samples (OR = 0.89, 95%: 0.81–0.98, $P = 0.017$). This SNP, however, was not selected for replication due to a relatively large P-value in the Stage-I scan.

Luciferase reporter assays showed no differences between the empty vector and the pGL3 basic or pGL3 enhancer vectors harboring rs10822013 fragments, indicating that rs10822013 fragments may not have intrinsic promoter activity (data not shown). In contrast, in the pGL3 promoter vector, fragments containing rs10822013 showed reduced luciferase activity, and the reduction was slightly more apparent in fragments containing the reference allele C than those containing the risk allele T (9% difference). However, the difference was not statistically significant. To further investigate whether the DNA sequence containing rs10822013 interacts with nuclear proteins, and if so, whether a single-nucleotide change in the rs10822013 site alters protein–DNA interactions, we performed electrophoretic mobility shift assays. In these assays,

Table 2. Association of SNP rs10822013 (10q21.2) with breast cancer risk in Stages I–IV

Stage	Cases/controls	MAF[a] (%)	OR[b] (95% CI) CT	TT	*P*-value for trend[b]
I	2062/2066	47.0	1.13 (0.97–1.31)	1.37 (1.15–1.63)	4.33×10^{-4}
II	4146/1823	46.9	1.13 (0.99–1.29)	1.26 (1.08–1.48)	3.74×10^{-3}
III	6436/6716	47.8	1.12 (1.03–1.21)	1.14 (1.03–1.26)	8.08×10^{-3}
IV	4509/6338	46.4	1.10 (0.99–1.22)	1.21 (1.07–1.37)	2.07×10^{-3}
Combined	17 153/16 943	47.1	1.12 (1.06–1.18)	1.21 (1.13–1.29)	5.87×10^{-9}
Chinese	11 069/9045	47.6	1.12 (1.04–1.19)	1.22 (1.12–1.32)	1.92×10^{-6}
Japanese	1575/1560	47.2	1.17 (0.98–1.38)	1.15 (0.94–1.40)	1.62×10^{-1}
Korean	4509/6338	46.4	1.10 (0.99–1.22)	1.21 (1.07–1.37)	2.07×10^{-3}

[a]Effect allele frequency (T) in controls.
[b]Adjusted for age and study site.

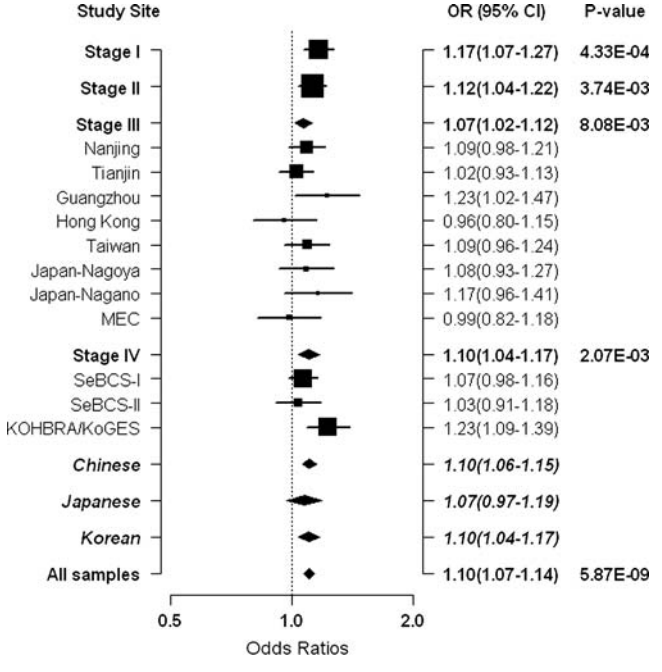

Figure 2. ORs per risk allele of SNP rs10822013 and 95% CIs for breast cancer by study site, ethnicity and study stage. The size of the box is proportional to the number of cases and controls in each study. The heterogeneity test for results among all studies was not statistically significant ($P = 0.4817$).

oligonucleotide probes corresponding to the reference allele C or the risk allele T were incubated with nuclear protein extracts from the MCF7 breast cancer cell line and from HEK293 cells. Compared with the reference allele C, the risk allele T of rs10822013 produced increased DNA–protein complex intensity (II) in both MCF7 and HEK293 cells (Fig. 4).

DISCUSSION

In this large GWAS conducted among East-Asian women and including 17 153 cases and 16 943 controls, we found strong evidence for a novel susceptibility variant for breast cancer, rs10822013 at 10q21.2.

SNP rs10822013 at 10q21.2 (Fig. 3) is located in an intronic region of the zinc finger protein 365 (*ZNF365*). This SNP is ~26.7 kb upstream of SNP rs10995190, which was recently reported in a GWAS conducted among women of European ancestry, and 9.2 kb upstream of SNP rs16917302, which was recently reported in a GWAS conducted among *BRCA2* mutation carriers. However, LD between rs10995190 and rs10822013 is virtually absent in European populations and completely absent in Chinese and African populations (Supplementary Material, Fig. S2). In addition, LD between rs16917302 and rs10822013 is very weak in Chinese and European populations and completely absent in African populations (Supplementary Material, Fig. S2). Furthermore, rs10995190 has a very low MAF in Chinese populations (2%) and we found no association of rs10995190 with breast cancer risk in our study, indicating that rs10822013 is likely a new risk variant for breast cancer.

SNP rs10822013 and its variant could affect alternative splicing or isoform transcription of the *ZNF365* gene. A database search (www.cbrc.jp/research/db/TFSEARCH.html) for transcription-factor-binding sites showed that sequences at rs10822013 have a high degree of similarity with consensus elements that can be differentially recognized by the transcription factor Cap. Our *in vitro* assay data showed that the risk allele T of rs10822013 produced increased DNA–protein complex intensity in both MCF7 breast cancer cells and HEK293 cells. The *ZNF365* gene encodes the zinc finger centrosomal protein, which is essential for cell division. The *ZNF365* gene is highly expressed in a number of cancer cell lines. Ectopic expression of the *ZNF365* gene can cause centrosome alterations and abnormal mitosis, which could lead to abnormal chromosome segregation and subsequently contribute to aneuploidy and malignant transformation (19). Several alternatively spliced variants, encoding distinct proteins, have been identified. These isoforms have different expression patterns, and their functions are largely unknown (www.ncbi.nlm.nih.gov/gene/22891). It has been reported that mutation in the *ZNF365* gene may be associated with uric acid nephrolithiasis (20). Taken together, these data suggest that rs10822013 may be a functional variant.

We also conducted a database search (SCAN, www.scandb.org) for expression quantitative trait locus (eQTL) genes associated with rs10822013 and the 26 SNPs that are in strong LD with rs10822013. Significant associations ($P < 10^{-4}$) between

Table 3. Association of SNP rs10822013 with breast cancer risk by menopausal status and ER status in Stages I–IV[a]

Genotype	Cases/controls	OR (95% CI)
Pre-menopausal women		
CC	1923/1466	1.00 (reference)
CT	3824/2638	1.14 (1.04–1.24)
TT	1907/1232	1.19 (1.07–1.32)
Per T allele	7654/5336	1.09 (1.04–1.15)
P-value for trend		1.14×10^{-3}
Post-menopausal women		
CC	1719/1507	1.00 (reference)
CT	3518/2725	1.11 (1.01–1.21)
TT	1687/1185	1.21 (1.09–1.35)
Per T allele	6924/5417	1.10 (1.05–1.16)
P-value for trend		3.17×10^{-4}
Heterogeneity test (pre- versus post-menopause): $P = 0.7343$		
ER (+) breast cancer		
CC	1956/4742	1.00 (reference)
CT	3961/8440	1.12 (1.04–1.20)
TT	1928/3761	1.21 (1.11–1.31)
Per T allele	7845/16 943	1.10 (1.05–1.15)
P-value for trend		1.05×10^{-5}
ER (−) breast cancer		
CC	1176/4742	1.00 (reference)
CT	2339/8440	1.11 (1.02–1.21)
TT	1154/3761	1.23 (1.12–1.36)
Per T allele	4669/16 943	1.11 (1.06–1.17)
P-value for trend		3.68×10^{-5}

[a]Adjusted for age and study site.

SNPs and gene expression were identified for 32 genes that were associated with two or more of these SNPs. The most interesting of these is a *trans*-eQTL gene, the ferritin, heavy polypeptide 1 (*FTH1*) gene, located on chromosome 11q13, for which small *P*-values were found for three SNPs (rs10822013, $P = 5 \times 10^{-6}$; rs10509168, $P = 1 \times 10^{-5}$; and rs2393886, $P = 2 \times 10^{-5}$). The *FTH1* gene encodes the heavy subunit of ferritin, the major intracellular iron storage protein in prokaryotes and eukaryotes. Overexpression of the *FTH1* gene was associated with a reduction of cellular labile iron, oxidative stress and inhibition of apoptosis (21,22). Gene expression levels of *FTH1* were up-regulated in breast cancer cells with an aggressive mesenchymal phenotype (23). We also found significant associations with the eQTL genes *CCR7* and *CFLAR*, both of which may be involved in cancer development (24–27).

In summary, in this large GWAS conducted as part of the Asia Breast Cancer Consortium, we found strong evidence of an association of a genetic variant, rs10822013 (10q21.2), with breast cancer risk. Our study further demonstrates the utility of conducting GWASs in non-European populations to identify novel genetic risk factors for breast cancer.

MATERIALS AND METHODS

Ethics statement

The study protocol was approved by the institutional review boards at Vanderbilt University Medical Center and at each collaborating institute. Informed consent was obtained from all participants.

Study population

This study consisted of a discovery stage and three validation stages. The overall study design is presented in Figure 1. Fifteen studies contributing a total of 17 153 breast cancer cases and 16 943 controls participated in this consortium. Detailed descriptions of participating studies are included in Supplementary Material. Briefly, the consortium included 20 114 Chinese women from eight studies conducted in the following locations: Shanghai, China [$n = 10\,097$; SBCS Phase I (SBCS-I) (7,28) and Phase II (SBCS-II) (7), Shanghai Breast Cancer Survival Study (SBCSS)/Shanghai Endometrial Cancer Study (SECS, only SECS controls were included in the current study) (7)], Tianjin, China [$n = 3095$ (29)], Nanjing, China [$n = 2816$ (30,31)], Taiwan [$n = 2080$ (32,33)], Hong Kong [$n = 1047$ (34)] and Guangzhou, China ($n = 979$); 3135 Japanese women from three studies conducted in Nagoya, Japan [$n = 1273$, (35)], Hawaii, USA [$n = 1061$; Multiethnic Cohort Study (MEC) (36,37)] and Nagano, Japan [$n = 801$, (38)]; and 10 847 women from three studies conducted in Korea [Seoul Breast Cancer Study (SeBCS) (39,40), Korean Hereditary Breast Cancer Study (KOHBRA) (41) and Korean Genome and Epidemiology Study (KoGES) (42)] (Table 1).

SNP selection for validation in Stage II

In the present study, we selected 49 promising SNPs for Stage-II replication among 4146 cases and 1823 controls recruited in the Shanghai studies. Only SNPs included in the Affymetrix 6.0 SNP arrays were selected for Stage-II evaluation. Selection criteria for these SNPs were: (i) MAF $\geq$ 10%, (ii) very clear genotyping clusters, (iii) not in strong LD ($r^2 \leq 0.5$) with any previously confirmed breast cancer genetic risk variants, (iv) consistent with Hardy–Weinberg equilibrium (HWE) with $P > 0.01$ in controls, and (v) $P < 0.005$ in the SBCS GWA scan data for SNPs on Affymetrix 6.0 array (7,11). If there were multiple SNPs in strong LD with a strong association in one region, the SNP with the smallest *P*-value was selected.

Genotyping

Genotyping for Stage I has been described previously (7,11). Briefly, the initial 300 subjects were genotyped using the Affymetrix GeneChip Mapping 500K Array Set, and the remaining 3918 subjects were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0. We included one negative control and three positive quality-control (QC) samples from Coriell Cell Repositories (http://ccr.coriell.org/) in each of the 96-well plates for Affymetrix SNP Array 6.0 genotyping. A total of 127 positive QC samples were successfully genotyped, and the average concordance rate was 99.9% with a median value of 100%. The sex of all study samples was confirmed to be female. The identity-by-descent analysis based on identity-by-state was conducted to detect first-degree cryptic relationships using PLINK, version 1.06. All samples with a call rate <95% were excluded. SNPs were excluded if: (i) MAF < 1%, (ii) call rate < 95%, or (iii) genotyping

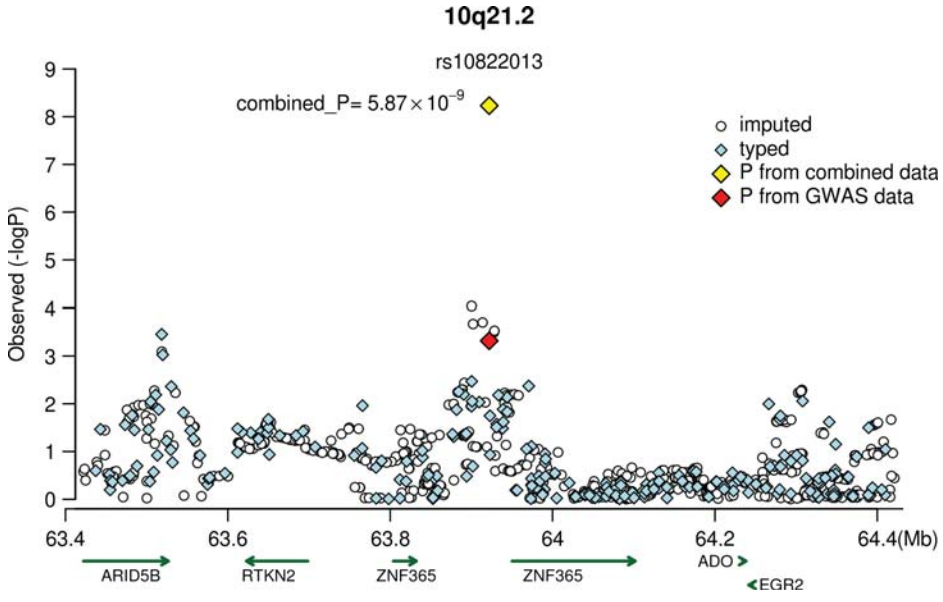

Figure 3. A regional plot of the $-\log P$-values for SNPs at 10q21.2. Results ($-\log P$) are shown for directly genotyped (blue diamonds) and imputed (white circles) SNPs for a 1 Mb region centered on SNP rs10822013. SNP rs10822013 is shown in red diamonds for Stage-I data and yellow diamonds for all four stages combined. Gene locations are from the March 2006 UCSC Genome Browser assembly.

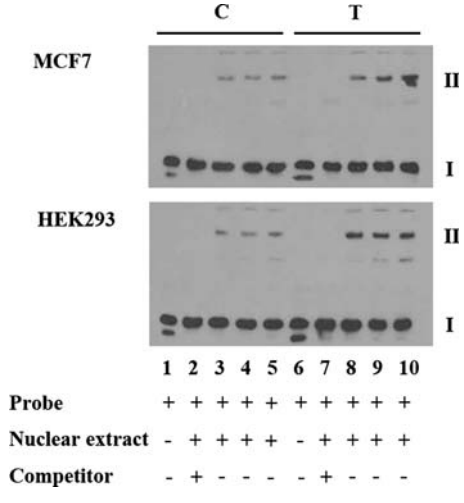

Figure 4. Electrophoretic mobility shift assays of SNP rs10822013. Nuclear protein extracts from MCF7 (top panel) and HEK293 (bottom panel) cells were incubated with biotin-labeled probes corresponding to the reference allele (lanes 1–5) or risk allele (lanes 6–10) of rs10822013 in the absence or presence of competitors. Lanes 1 and 6, no nuclear extracts; lanes 2 and 7, unlabeled competitor in 200-fold molar excess; lanes 3 and 8 (5 mM $MgCl_2$); lanes 4 and 9 (2.5 mM $MgCl_2$); and lanes 5 and 10 (1.25 mM $MgCl_2$), no competitor. I, free biotin-labeled probes; II, specific DNA–protein complex bands.

concordance rate $< 95\%$ in QC samples. The final data set included 2062 cases and 2066 controls for 684 457 markers.

Genotyping for Stage II was completed using the iPLEX Sequenom MassArray platform (7,11). Included in each 96-well plate as QC samples were one negative control (water), two blinded duplicates and two samples from the HapMap project. The mean concordance rate was 99.7% for the blinded duplicates and 98.8% for HapMap samples. To compare the consistency between Affymetrix 6.0 and Sequenom platforms, we also genotyped 124 samples included in Stage I in the Stage II Sequenom genotyping; the concordance rate was 98.6% between the two platforms.

Genotyping for eight of the studies included in Stage III was performed at five different centers using TaqMan assays. The genotyping assay protocol was developed and validated at the Vanderbilt Molecular Epidemiology Laboratory, and TaqMan genotyping assay reagents were provided to the investigators of the Tianjin study (Tianjin Cancer Institute and Hospital), Nanjing study (Nanjing Medical University), Guangzhou study (Sun Yat-sen University) and MEC study (University of Southern California), who conducted genotyping assays at their own laboratories. Samples from the four other studies included in Stage III were genotyped at the Vanderbilt Molecular Epidemiology Laboratory. For TaqMan genotyping assays conducted at the Vanderbilt Molecular Epidemiology Laboratory, one negative control and two samples from the HapMap project were included in each 96-well plate, along with 30 unrelated European and 45 Chinese samples from the HapMap project for QC purposes. The consistency rate was 100% for the HapMap samples comparing the genotyping data obtained from the current study with the data obtained from the HapMap project. Each of the non-Vanderbilt laboratories was asked to genotype a trial plate containing DNA from 70 Chinese-ancestry samples before the main study genotyping was conducted. The consistency rate across all centers for these trial samples was 100% compared with genotypes previously determined at Vanderbilt. In addition, replicate samples comparing 3–7% of all study samples were dispersed among the genotyping plates at all centers.

Data from the SeBCS phase I (SeBCS-I) analyzed in Stage IV were extracted from GWA scan data generated using the Affymetrix Genome-Wide Human SNP Array 6.0. A total of 30 QC samples were successfully genotyped, and the concordance rate was 99.83%. The sex of all samples was

confirmed to be female. SNPs were excluded if: (i) genotype call rate $< 95\%$, (ii) MAF $< 1\%$ in either the cases or controls, (iii) deviation from HWE at $P < 10^{-6}$, and (iv) poor cluster plot in either cases or controls. Genotyping for other Stage IV samples [SeBCS phase II (SeBCS-II), KOHBRA and KoGES] was completed using the iPLEX Sequenom MassArray platform as described above (7,11). The mean concordance rate was 99.7% for the blinded duplicates, 98.9% for HapMap samples and 99.5% between Sequenom and Affymetrix 6.0 genotyping.

Plasmid constructs and luciferase assays

A 2.2 kb DNA fragment was PCR-amplified by using human genomic DNA containing either the major allele (C) or the minor allele (T) of rs10822013 with forward primer 5′-GTTACGCGTCAC CTA TAG AAA AGG GCC TGG TTG-3′ and backward primer 5′-GTTCTCGAGCTC TTT ACC TAG GGC AGA GGA GC-3′. The fragments were cloned upstream of the luciferase reporter vectors pGL3 basic, pGL3 promoter or pGL3 enhancer (Promega, WI, USA) between the *Nhe*1 and *Xho*1 restriction sites. All DNA constructs were verified by sequencing analysis. Enhancer and promoter activity was determined by transient transfection followed by an *in vitro* luciferase reporter assay in HEK293 cells. Transfection was performed with the use of FuGene 6 Transfection Reagent (Roche Diagnostics, Indianapolis, IN, USA) in triplicate for each of the constructs. Briefly, 2×10^5 cells were seeded in 24-well plates and co-transfected with pGL4.73, a *Renilla* expressing vector, which served as a reference for transfection efficiency. Thirty-six to 48 h later, the cells were lysed with passive lysis buffer, and luminescence (relative light units) was measured using the Dual-Luciferase Assay System (Promega). The rs10822013 regulatory activity was measured as a ratio of firefly luciferase activity to *Renilla* luciferase activity, and the mean was calculated from three independent experiments with triplicate assays in each experiment.

Electrophoretic mobility shift assay

Biotin-labeled, double-stranded oligonucleotide probes 5′-TGG CAC AAG AAA ATG CGT TGT GAA CAA ACT-3′ and 5′-AGT TTG TTC ACA ACG CAT TTT CTT GTG CCA-3′, and 5′-TGG CAC AAG AAA ATG TGT TGT GAA CAA ACT-3′ and 5′-AGT TTG TTC ACA ACA CAT TTT CTT GTG CCA-3′ containing either the major or minor allele sequence were synthesized. The probes were incubated with nuclear protein extracts from HEK293 and MCF7 cells, in the presence or absence of competitors, i.e. unlabelled probes. Protein–DNA complexes were resolved by polyacrylamide gel electrophoresis and detected using a LightShift Chemiluminescent EMSA kit (Pierce Biotechnology, Rockford, IL, USA).

Statistical analyses

PLINK version 1.06 was used to analyze genome-wide data obtained in Stage I. A set of 4305 SNPs with an MAF $\geq 35\%$ and a distance ≥ 100 kb between two adjacent SNPs was selected to evaluate the population structure. The inflation factor λ was estimated to be 1.038, suggesting that any population substructure, if present, should not have any appreciable effect on the results.

Individual data were obtained from each study for a pooled analysis. Case–control differences in selected demographic characteristics and major risk factors were evaluated using *t*-tests (for continuous variables) or χ^2-tests (for categorical variables). Associations between SNPs and breast cancer risk were assessed using ORs and 95% CIs derived from logistic regression models. ORs were estimated for heterozygotes and homozygotes for the variant allele compared with homozygotes for the common allele. ORs also were estimated for the variant allele based on a log-additive model and adjusted for age, study site and ethnicity, when appropriate. Heterogeneity across studies and between ethnicities was assessed with likelihood ratio tests. Stratified analyses by ethnicity, menopausal status and ER status were carried out.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

The authors wish to thank the study participants and research staff for their contributions and commitment to this project, Regina Courtney and Jie Wu for DNA preparation, and Bethanie Rammer and Mary Jo Daly for clerical support in the preparation of this manuscript.

Conflict of Interest statement. None declared.

FUNDING

This research was supported in part by US National Institutes of Health grants R01CA124558, R01CA064277, R37CA070867, R01CA090899 and R01CA100374, as well as Ingram Professorship and Research Reward funds awarded to W.Z.; R01CA118229, R01CA092585, and Department of Defense (DOD) Idea Award BC011118 awarded to X.-O.S.; and R01CA122756 and DOD Idea Award BC050791 awarded to Q.C. Participating studies (principal investigator, grant support) in the consortium are as follows: the Shanghai Breast Cancer Study (W.Z., R01CA064277), the Shanghai Breast Cancer Survival Study (X.-O.S., R01CA118229), the Shanghai Endometrial Cancer Study (X.-O.S., R01CA092585, contributed only controls to the consortium), the Seoul Breast Cancer Study [D.K., Basic Research Laboratory (BRL) program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology, 2011-0001564], the Korean Hereditary Breast Cancer Study/Korean Genome and Epidemiology Study (S.-W.K., the National R&D Program for Cancer Control, Ministry for Health, Welfare and Family Affairs, Republic of Korea, 1020350), the Tianjin Study (K.C., the National Natural Science Foundation of China Grant No. 30771844), the Nanjing Study (H.S., 09KJA330001, Jiangsu, China), the Taiwan Biobank Study

(C.-Y.S., DOH97-01), the Hong Kong Study (U.S.K., Research Grant Council, Hong Kong SAR, China, HKU 7520/05M and 76730M), the Guangzhou Breast Cancer Study (Z.R., the National Natural Science Foundation of China, 81072383), the Multiethnic Cohort Study (B.E.H., CA063464; L.K., CA054281; and C.H., CA132839), the Nagano Breast Cancer Study (S.T., Grants-in-Aid for the Third Term Comprehensive Ten-Year Strategy for Cancer Control from the Ministry of Health, Labor and Welfare of Japan, for Scientific Research on Priority Areas, 17015049 and for Scientific Research on Innovative Areas, 221S0001, from the Ministry of Education, Culture, Sports, Science, and Technology of Japan) and the Hospital-based Epidemiologic Research Program at Aichi Cancer Center (K.T., Grants-in-Aid for Scientific Research on Priority Areas, 17015052, from the Ministry of Education, Culture, Sports, Science, and Technology of Japan H.T., Grants-in-Aid for the Third Term Comprehensive Ten-Year Strategy for Cancer Control from the Ministry of Health, Labor and Welfare of Japan, H20-002). 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 Vanderbilt-Ingram Cancer Center (P30 CA068485).

REFERENCES

1. Nathanson, K.L., Wooster, R. and Weber, B.L. (2001) Breast cancer genetics: what we know and what we need. *Nat. Med.*, **7**, 552–556.
2. Balmain, A., Gray, J. and Ponder, B. (2003) The genetics and genomics of cancer. *Nat. Genet.*, **33** (suppl.), 238–244.
3. Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D., Thompson, D., Ballinger, D.G., Struewing, J.P., Morrison, J., Field, H., Luben, R. *et al.* (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, **447**, 1087–1093.
4. Hunter, D.J., Kraft, P., Jacobs, K.B., Cox, D.G., Yeager, M., Hankinson, S.E., Wacholder, S., Wang, Z., Welch, R., Hutchinson, A. *et al.* (2007) A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.*, **39**, 870–874.
5. Stacey, S.N., Manolescu, A., Sulem, P., Thorlacius, S., Gudjonsson, S.A., Jonsson, G.F., Jakobsdottir, M., Bergthorsson, J.T., Gudmundsson, J., Aben, K.K. *et al.* (2008) Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat. Genet.*, **40**, 703–706.
6. Gold, B., Kirchhoff, T., Stefanov, S., Lautenberger, J., Viale, A., Garber, J., Friedman, E., Narod, S., Olshen, A.B., Gregersen, P. *et al.* (2008) Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. *Proc. Natl Acad. Sci. USA*, **105**, 4340–4345.
7. Zheng, W., Long, J., Gao, Y.T., Li, C., Zheng, Y., Xiang, Y.B., Wen, W., Levy, S., Deming, S.L., Haines, J.L. *et al.* (2009) Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat. Genet.*, **41**, 324–328.
8. Thomas, G., Jacobs, K.B., Kraft, P., Yeager, M., Wacholder, S., Cox, D.G., Hankinson, S.E., Hutchinson, A., Wang, Z., Yu, K. *et al.* (2009) A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nat. Genet.*, **41**, 579–584.
9. Ahmed, S., Thomas, G., Ghoussaini, M., Healey, C.S., Humphreys, M.K., Platte, R., Morrison, J., Maranian, M., Pooley, K.A., Luben, R. *et al.* (2009) Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat. Genet.*, **41**, 585–590.
10. Stacey, S.N., Manolescu, A., Sulem, P., Rafnar, T., Gudmundsson, J., Gudjonsson, S.A., Masson, G., Jakobsdottir, M., Thorlacius, S., Helgason, A. *et al.* (2007) Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat. Genet.*, **39**, 865–869.
11. Long, J., Cai, Q., Shu, X.O., Qu, S., Li, C., Zheng, Y., Gu, K., Wang, W., Xiang, Y.B., Cheng, J. *et al.* (2010) Identification of a functional genetic variant at 16q12.1 for breast cancer risk: results from the Asia Breast Cancer Consortium. *PLoS Genet.*, **6**, e1001002.
12. Turnbull, C., Ahmed, S., Morrison, J., Pernet, D., Renwick, A., Maranian, M., Seal, S., Ghoussaini, M., Hines, S., Healey, C.S. *et al.* (2010) Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat. Genet.*, **42**, 504–507.
13. Antoniou, A.C., Wang, X., Fredericksen, Z.S., McGuffog, L., Tarrell, R., Sinilnikova, O.M., Healey, S., Morrison, J., Kartsonaki, C., Lesnick, T. *et al.* (2010) A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. *Nat. Genet.*, **42**, 885–892.
14. Fletcher, O., Johnson, N., Orr, N., Hosking, F.J., Gibson, L.J., Walker, K., Zelenika, D., Gut, I., Heath, S., Palles, C. *et al.* (2011) Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. *J. Natl Cancer Inst.*, **103**, 425–435.
15. Gaudet, M.M., Kirchhoff, T., Green, T., Vijai, J., Korn, J.M., Guiducci, C., Segre, A.V., McGee, K., McGuffog, L., Kartsonaki, C. *et al.* (2010) Common genetic variants and modification of penetrance of BRCA2-associated breast cancer. *PLoS Genet.*, **6**, e1001183.
16. Zhang, B., Beeghly-Fadiel, A., Long, J. and Zheng, W. (2011) Genetic variants associated with breast-cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. *Lancet Oncol.*, **12**, 477–488.
17. Long, J., Shu, X.O., Cai, Q., Gao, Y.T., Zheng, Y., Li, G., Li, C., Gu, K., Wen, W., Xiang, Y.B. *et al.* (2010) Evaluation of breast cancer susceptibility loci in Chinese women. *Cancer Epidemiol. Biomarkers Prev.*, **19**, 2357–2365.
18. Zheng, W., Wen, W., Gao, Y.T., Shyr, Y., Zheng, Y., Long, J., Li, G., Li, C., Gu, K., Cai, Q. *et al.* (2010) Genetic and clinical predictors for breast cancer risk assessment and stratification among Chinese women. *J. Natl Cancer Inst.*, **102**, 972–981.
19. Wang, Q., Du, X., Meinkoth, J., Hirohashi, Y., Zhang, H., Liu, Q., Richter, M. and Greene, M.I. (2006) Characterization of Su48, a centrosome protein essential for cell division. *Proc. Natl Acad. Sci. USA*, **103**, 6512–6517.
20. Gianfrancesco, F., Esposito, T., Ombra, M.N., Forabosco, P., Maninchedda, G., Fattorini, M., Casula, S., Vaccargiu, S., Casu, G., Cardia, F. *et al.* (2003) Identification of a novel gene and a common variant associated with uric acid nephrolithiasis in a Sardinian genetic isolate. *Am. J. Hum. Genet.*, **72**, 1479–1491.
21. Berberat, P.O., Katori, M., Kaczmarek, E., Anselmo, D., Lassman, C., Ke, B., Shen, X., Busuttil, R.W., Yamashita, K., Csizmadia, E. *et al.* (2003) Heavy chain ferritin acts as an antiapoptotic gene that protects livers from ischemia reperfusion injury. *FASEB J.*, **17**, 1724–1726.
22. Epsztejn, S., Glickstein, H., Picard, V., Slotki, I.N., Breuer, W., Beaumont, C. and Cabantchik, Z.I. (1999) H-ferritin subunit overexpression in erythroid cells reduces the oxidative stress response and induces multidrug resistance properties. *Blood*, **94**, 3593–3603.
23. Shpyleva, S.I., Tryndyak, V.P., Kovalchuk, O., Starlard-Davenport, A., Chekhun, V.F., Beland, F.A. and Pogribny, I.P. (2011) Role of ferritin alterations in human breast cancer cells. *Breast Cancer Res. Treat.*, **126**, 63–71.
24. Wu, X., Lee, V.C., Chevalier, E. and Hwang, S.T. (2009) Chemokine receptors as targets for cancer therapy. *Curr. Pharm. Des.*, **15**, 742–757.
25. Mburu, Y.K., Wang, J., Wood, M.A., Walker, W.H. and Ferris, R.L. (2006) CCR7 mediates inflammation-associated tumor progression. *Immunol. Res.*, **36**, 61–72.
26. Yu, J.W. and Shi, Y. (2008) FLIP and the death effector domain family. *Oncogene*, **27**, 6216–6227.
27. Safa, A.R., Day, T.W. and Wu, C.H. (2008) Cellular FLICE-like inhibitory protein (C-FLIP): a novel target for cancer therapy. *Curr. Cancer Drug Targets*, **8**, 37–46.
28. Gao, Y.T., Shu, X.O., Dai, Q., Potter, J.D., Brinton, L.A., Wen, W., Sellers, T.A., Kushi, L.H., Ruan, Z., Bostick, R.M. *et al.* (2000) Association of menstrual and reproductive factors with breast cancer risk: results from the Shanghai Breast Cancer Study. *Int. J. Cancer*, **87**, 295–300.
29. Zhang, L., Gu, L., Qian, B., Hao, X., Zhang, W., Wei, Q. and Chen, K. (2009) Association of genetic polymorphisms of ER-alpha and the estradiol-synthesizing enzyme genes CYP17 and CYP19 with breast cancer risk in Chinese women. *Breast Cancer Res. Treat.*, **114**, 327–338.

30. Liang, J., Chen, P., Hu, Z., Zhou, X., Chen, L., Li, M., Wang, Y., Tang, J., Wang, H. and Shen, H. (2008) Genetic variants in fibroblast growth factor receptor 2 (FGFR2) contribute to susceptibility of breast cancer in Chinese women. *Carcinogenesis*, **29**, 2341–2346.
31. Wang, Y., Hu, Z., Liang, J., Wang, Z., Tang, J., Wang, S., Wang, X., Qin, J., Wang, X. and Shen, H. (2008) A tandem repeat of human telomerase reverse transcriptase (hTERT) and risk of breast cancer development and metastasis in Chinese women. *Carcinogenesis*, **29**, 1197–1201.
32. Ding, S.L., Yu, J.C., Chen, S.T., Hsu, G.C., Kuo, S.J., Lin, Y.H., Wu, P.E. and Shen, C.Y. (2009) Genetic variants of BLM interact with RAD51 to increase breast cancer susceptibility. *Carcinogenesis*, **30**, 43–49.
33. Hsu, H.M., Wang, H.C., Chen, S.T., Hsu, G.C., Shen, C.Y. and Yu, J.C. (2007) Breast cancer risk is associated with the genes encoding the DNA double-strand break repair Mre11/Rad50/Nbs1 complex. *Cancer Epidemiol. Biomarkers Prev.*, **16**, 2024–2032.
34. Chan, K.Y., Liu, W., Long, J.R., Yip, S.P., Chan, S.Y., Shu, X.O., Chua, D.T., Cheung, A.N., Ching, J.C., Cai, H. *et al.* (2009) Functional polymorphisms in the BRCA1 promoter influence transcription and are associated with decreased risk for breast cancer in Chinese women. *J. Med. Genet.*, **46**, 32–39.
35. Hamajima, N., Matsuo, K., Saito, T., Hirose, K., Inoue, M., Takezaki, T., Kuroishi, T. and Tajima, K. (2001) Gene-environment interactions and polymorphism studies of cancer risk in the Hospital-based Epidemiologic Research Program at Aichi Cancer Center II (HERPACC-II). *Asian Pac. J. Cancer Prev.*, **2**, 99–107.
36. Haiman, C.A., Garcia, R.R., Hsu, C., Xia, L., Ha, H., Sheng, X., Le, M.L., Kolonel, L.N., Henderson, B.E., Stallcup, M.R. *et al.* (2009) Screening and association testing of common coding variation in steroid hormone receptor co-activator and co-repressor genes in relation to breast cancer risk: the Multiethnic Cohort. *BMC Cancer*, **9**, 43.
37. Kolonel, L.N., Henderson, B.E., Hankin, J.H., Nomura, A.M., Wilkens, L.R., Pike, M.C., Stram, D.O., Monroe, K.R., Earle, M.E. and Nagamine, F.S. (2000) A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am. J. Epidemiol.*, **151**, 346–357.
38. Itoh, H., Iwasaki, M., Hanaoka, T., Kasuga, Y., Yokoyama, S., Onuma, H., Nishimura, H., Kusama, R. and Tsugane, S. (2009) Serum organochlorines and breast cancer risk in Japanese women: a case-control study. *Cancer Causes Control*, **20**, 567–580.
39. Choi, J.Y., Lee, K.M., Park, S.K., Noh, D.Y., Ahn, S.H., Yoo, K.Y. and Kang, D. (2005) Association of paternal age at birth and the risk of breast cancer in offspring: a case control study. *BMC Cancer*, **5**, 143.
40. Lee, K.M., Choi, J.Y., Park, S.K., Chung, H.W., Ahn, B., Yoo, K.Y., Han, W., Noh, D.Y., Ahn, S.H., Kim, H. *et al.* (2005) Genetic polymorphisms of ataxia telangiectasia mutated and breast cancer risk. *Cancer Epidemiol. Biomarkers Prev.*, **14**, 821–825.
41. Han, S.A., Park, S.K., Hyun, A.S., Hyuk, L.M., Noh, D.Y., Kim, L.S., Noh, W.C., Jung, Y., Sang, K.K. and Kim, S.W., Korean Breast Cancer Study Group (2011) The Korean Hereditary Breast Cancer (KOHBRA) Study: protocols and interim report. *Clin. Oncol. (R. Coll. Radiol.)*, **23**, 434–441.
42. Cho, Y.S., Go, M.J., Kim, Y.J., Heo, J.Y., Oh, J.H., Ban, H.J., Yoon, D., Lee, M.H., Kim, D.J., Park, M. *et al.* (2009) A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat. Genet.*, **41**, 527–534.