

## ORIGINAL ARTICLE

# Genetic variants at 5p12 and risk of breast cancer in Han Chinese

Xiao'an Liu<sup>1,4</sup>, Zhenzhen Qin<sup>2,3,4</sup>, Hao Shen<sup>2,3</sup>, Jialei Xue<sup>1</sup>, Yue Jiang<sup>2,3</sup>, Zhibin Hu<sup>2,3</sup>, Hongbing Shen<sup>2,3</sup> and Shui Wang<sup>1</sup>

**A genome-wide association study, conducted among women of European ancestry, has identified two single-nucleotide polymorphisms (SNPs) rs4415084 (T > C) and rs10941679 (A > G) at chromosome 5p12 were associated with risk of breast cancer, suggesting that genetic variants in this region may have a role in the development of breast cancer. To investigate the associations between SNPs at 5p12 and risk of breast cancer in the Chinese population, we conducted a fine-mapping in 5p12 using a haplotype-tagging SNP approach and genotyped these SNPs with a case-control study consisting of 878 cases and 900 controls. We found that the two risk SNPs reported in the European population were neither associated with breast cancer risk in our Chinese population, nor did the fine-mapping SNPs after controlling multiple comparison.**

*Journal of Human Genetics* (2012) **57**, 638–641; doi:10.1038/jhg.2012.83; published online 26 July 2012

**Keywords:** breast cancer; molecular epidemiology; polymorphisms; 5p12

## INTRODUCTION

Breast cancer is the most common cancer among women worldwide.<sup>1</sup> In the past decade, China's urban cancer registries have documented increased incidence rates of between 20 and 30% for breast cancer.<sup>2,3</sup> Epidemiological investigation of breast cancer has identified a number of environmental and lifestyle risk factors, such as age at menarche, menopause and first birth and exogenous hormone use.<sup>4–6</sup> Furthermore, accumulative studies also indicate that a substantial proportion of breast cancer is due to inherited susceptibility. Not only the known high-penetrance genes (e.g., *BRCA1*, *BRCA2*, *ATM*, etc.) but also the other lower penetrance genes contribute to breast cancer risk.<sup>7</sup>

Recently, a genome-wide association study conducted in European ancestry (EA) population by Stacey *et al.*<sup>8</sup> identified two genetic susceptibility loci, rs4415084 and rs10941679, at chromosome 5p12 were associated with breast cancer risk. Later, another study directed by Thomas *et al.*<sup>9</sup> confirmed the single-nucleotide polymorphism (SNP) rs4415084 might affect breast cancer risk in the EA population. By examining the HapMap data, we found that, in HapMap CEU (Utah residents with ancestry from northern and western Europe) samples, the two SNPs rs4415084 and rs10941679 resided in a linkage disequilibrium block (chr5:44666047...45078551) containing the *MRPS30* gene, which is also known as *PDCD9* (programmed cell death protein 9) and encodes a component of

the mitochondrial ribosome. A few studies reported that it was previously implicated in mammalian cells apoptosis and estrogen receptor (ER)-positive breast tumors.<sup>10–12</sup> Therefore, it seems plausible that genetic variations, such as SNPs on 5p12, that affect *MRPS30* gene expression or protein function would be associated with the risk of breast cancer.

The exciting results encouraged us to test the hypothesis that if SNPs on chromosome 5p12 also have a role in susceptibility to breast cancer in the Chinese population. Through the HapMap database, we found that the risk SNPs implicated in the EA population have different allele frequencies and linkage patterns in Chinese. Thus, we applied a fine-mapping study in the susceptible region of 5p12 identified by Stacey *et al.*<sup>8</sup> in their genome-wide association study of breast cancer. Finally, we identified 10 SNPs with minor allele frequency  $\geq 0.05$  in the Chinese population and genotyped these SNPs in a case-control study with 878 cases and 900 controls.

## MATERIALS AND METHODS

### Study subjects

The case-control study was approved by the Institutional Review Board of Nanjing Medical University. A total of 878 breast cancer cases and 900 cancer-free controls were included. The cases were incident breast cancer patients and were consecutively recruited from the First Affiliated Hospital of Nanjing Medical University, the Cancer Hospital of Jiangsu Province and the

<sup>1</sup>Department of General Surgery, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China; <sup>2</sup>State Key Laboratory of Reproductive Medicine, Institute of Toxicology, Nanjing Medical University, Nanjing, China and <sup>3</sup>Department of Epidemiology and Biostatistics, Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, School of Public Health, Nanjing Medical University, Nanjing, China

<sup>4</sup>These authors contributed equally to this work.

Correspondence: Dr H Shen, Department of Epidemiology and Biostatistics, MOE Key Laboratory of Modern Toxicology, School of Public Health, Nanjing Medical University, 140 Hanzhong Road, Nanjing 210029, Jiangsu, China.

E-mail: hbshen@njmu.edu.cn

or Dr S Wang, Department of Breast Surgery, The First Affiliated Hospital of Nanjing Medical University, 300 Guangzhou Road, 210029 Nanjing, China.

E-mail: ws0801@hotmail.com

Received 17 March 2012; revised 30 May 2012; accepted 30 May 2012; published online 26 July 2012

Drum-tower Hospital, Nanjing, China, from January 2004 to April 2010. All the breast cancer cases were histopathologically confirmed without restrictions of age or histological type. Exclusion criteria included self-reported previous cancer history, metastasized cancer from other organs, and previous radiotherapy or chemotherapy. Cancer-free women controls, frequency matched to the cases on age ( $\pm 5$  years) and residential area (urban or rural), were randomly selected from a cohort of >30 000 participants in a community-based screening program for noninfectious diseases conducted in Jiangsu Province, China. All the subjects were genetically unrelated, ethnic Han Chinese women. After providing informed consent, each woman was personally interviewed face-to-face by trained interviewers using a pre-tested standard questionnaire to obtain information on demographic data, menstrual and reproductive history, environmental exposure history and family history of cancer in first-degree relatives (parents, siblings and children). After the interview, each subject provided 5 ml of venous blood. The ER and progesterone receptor status of breast cancer were determined from the results of immunohistochemistry examinations according to the medical records of the patients.

### SNPs selection and genotyping

By using information from the HapMap database (<http://www.hapmap.org/>, phaseII Nov 08) and the HaploView 4.2 software (Broad Institute, Cambridge, MA, USA), we conducted a block-based tagging strategy to find tagging SNPs (minor allele frequency  $\geq 0.05$  in the Chinese population, Hardy–Weinberg equilibrium  $P \geq 0.05$ ) on the basis of pair-wise linkage disequilibrium  $r^2$  threshold of 0.8 in this region (chr5:44666047...45078551). The two index SNPs from the genome-wide association study, rs4415084 and rs10941679, were forced into the set. Together, nine tagging SNPs (rs13357090, rs4415084, rs10941678, rs3747479, rs11951760, rs9790896, rs7703618, rs2067980, rs11742346 and rs4533894) were chosen. In addition, we also added a common nonsynonymous SNP rs3747479 in exon 1 of *MRPS30*. As a result, a total of 10 SNPs (rs13357090, rs4415084, rs10941678, rs3747479, rs11951760, rs9790896, rs7703618, rs2067980, rs11742346 and rs4533894) were included in our study.

Genotypes were determined using the middle-throughput TaqMan Open-Array Genotyping Platform (Applied Biosystems Inc., Foster City, CA, USA) and genotyping was performed without knowing the case or control status. Two blank controls in each plate were used for quality control. Samples were analyzed with AutoCaller Software (Applied Biosystems Inc.). 96 samples were also randomly selected from the Openarray platform and re-genotyped by using TaqMan allelic assays for the 10 SNPs and the results were all consistent.

### Statistical analysis

Differences between the cases and controls on the demographic characteristics, selected variables and frequencies of the genotypes of 10 SNPs were calculated by using the Student's *t*-test (for continuous variables) or  $\chi^2$ -test (for categorical variables). The Hardy–Weinberg equilibrium was tested by a goodness-of-fit  $\chi^2$ -test to compare the observed genotype frequencies to the expected ones among the control subjects. The associations between the 10 SNPs genotypes and the risk of breast cancer were estimated by computing the odds ratios (ORs) and their 95% confidence intervals (CIs) from logistic regression analyses, with and without the adjustment for age, age at menarche and menopausal status (natural menopause as one status while unnatural menopause and premenopausal status were merged to the other). The potential gene–environment interaction was also evaluated by logistic regression analyses and tested by comparing the changes in deviance ( $-2 \log$  likelihood) between the models of main effects with or without the interaction term. We used the PHASE 2.0 program (Manchester University, Manchester, UK) to infer haplotype frequencies based on the observed 5p12 genotypes. All of the statistical analyses were performed using the Statistical Analysis System software (9.1.3; SAS Institute, Cary, NC, USA). All tests were two-sided and the significance level was set at  $P < 0.05$ .

### RESULTS

The characteristics of 878 cases and 900 controls were summarized in Table 1. There were statistically significant differences between the

**Table 1 Distributions of select variables in breast cancer cases and cancer-free controls**

Variables	Cases N = 878	Controls N = 900	P-value
Age, years (means $\pm$ s.d.)	51.29 $\pm$ 11.38	51.47 $\pm$ 11.67	0.75
Age at menarche, years (means $\pm$ s.d.) <sup>a</sup>	15.25 $\pm$ 1.85	15.85 $\pm$ 1.90	<0.0001
Age at first live birth, years (means $\pm$ s.d.) <sup>b</sup>	25.65 $\pm$ 3.48	24.90 $\pm$ 3.40	<0.0001
Age at natural menopause, years (means $\pm$ s.d.) <sup>c</sup>	49.85 $\pm$ 3.18	49.15 $\pm$ 4.09	0.0075
Menopausal status <sup>d</sup>			<0.0001
Premenopausal	416 (48.37%)	437 (49.38%)	
Natural menopausal	386 (44.88%)	432 (48.81%)	
Unnatural menopausal	58 (6.74%)	16 (1.81%)	
ER <sup>e</sup>			
Positive	369 (53.48%)		
Negative	321 (46.52%)		
PR <sup>e</sup>			
Positive	396 (57.39%)		
Negative	294 (42.61%)		

Abbreviations: ER, estrogen receptor; PR, progesterone receptor.

<sup>a</sup>Age at menarche information was available in 862 breast cancer cases and 896 controls.

<sup>b</sup>Age at first live birth information was available in 830 breast cancer cases and 872 controls.

<sup>c</sup>Age at natural menopause information was available in 380 breast cancer cases and 411 controls.

<sup>d</sup>Menopausal status information was available in 860 breast cancer cases and 885 controls.

<sup>e</sup>Information of ER and PR was available in 690 breast cancer cases, respectively.

cases and controls in the distribution of age at menarche, age of first live birth, menopausal status and age at natural menopause ( $P < 0.01$ ).

The genotype distributions of the 10 SNPs in cases and controls were shown in Table 2. The observed genotype frequencies of 10 SNPs were all in Hardy–Weinberg equilibrium in the controls ( $P > 0.05$  for all SNPs). In single-locus analyses, the genotype distribution of SNP rs4533894 was significantly different between cases and controls ( $P = 0.04$ ). Multivariate logistic regression analyses revealed that the variant GG genotype of rs4533894 was associated with a significantly increased risk of breast cancer (adjusted OR = 3.13, 95% CI = 1.31–7.48), compared with the wild-type AA. However, no overall significant associations were observed between the other nine SNPs, including the two risk SNPs (rs4415084 and rs10941679) previously identified in the EA population and breast cancer risk in our study population. Furthermore, the association between rs4533894 and breast cancer did not remain significant after controlling multiple comparisons.

We further evaluated the effect of rs4533894 on breast cancer risk stratified by age, menopausal status (premenopausal and natural menopausal), age at menarche, age at first live birth and ER/progesterone receptor status on the basis of the dominant model. The increased risk of rs4533894 AG/GG was more pronounced only among women with ER $-$  (OR = 1.51, 95% CI = 1.06–2.16; data not shown). Interestingly, stratified analyses of the other nine SNPs showed that the variant genotypes of rs4415084 (TC/CC) and rs11742346 (CT/TT) were significantly associated with breast cancer risk among subjects with younger menarche age (rs4415084 OR = 0.67, 95% CI = 0.46–0.97,  $P$  for heterogeneity = 0.04; rs11742346 OR = 1.52, 95% CI = 1.08–2.15,  $P$  for heterogeneity = 0.02; data not shown). We conducted interaction analyses

**Table 2** Logistic regression analyses on associations between the 10 SNPs in 5p12 and risk of breast cancer

Genotype	Case N = 878 (%)	Control N = 900 (%)	Adjusted OR <sup>a</sup> (95% CI)	
rs13357090 C>T	851	883		
CC	712 (83.67)	739 (83.69)	0.55 <sup>b</sup>	1.00
CT	137 (16.10)	139 (15.74)		1.01 (0.77–1.31)
TT	2 (0.24)	5 (0.57)		0.44 (0.08–2.32)
CC	712 (83.67)	739 (83.69)		1.00
TT + CT	139 (16.33)	144 (16.31)	0.93 <sup>b</sup>	0.99 (0.76–1.28)
T allele frequency	8.28	8.44	0.79 <sup>c</sup>	0.97 (0.75–1.24)
rs4415084 T>C	806	849		
TT	283 (35.11)	285 (33.57)	0.80 <sup>b</sup>	1.00
TC	357 (44.29)	387 (45.58)		0.94 (0.75–1.18)
CC	166 (20.60)	177 (20.85)		0.95 (0.72–1.25)
TT	283 (35.11)	285 (33.57)		1.00
CC + TC	523 (64.89)	564 (66.43)	0.58 <sup>b</sup>	0.94 (0.77–1.16)
C allele frequency	42.74	43.64	0.67 <sup>c</sup>	0.97 (0.85–1.11)
rs10941679 A>G	846	882		
AA	224 (26.48)	220 (24.94)	0.77 <sup>b</sup>	1.00
AG	417 (49.29)	443 (50.23)		0.91 (0.72–1.16)
GG	205 (24.23)	219 (24.83)		0.92 (0.70–1.21)
AA	224 (26.48)	220 (24.94)		1.00
GG + AG	622 (73.52)	662 (75.06)	0.45 <sup>b</sup>	0.92 (0.74–1.15)
G allele frequency	48.88	49.6	0.54 <sup>c</sup>	0.96 (0.84–1.10)
rs3747479 G>C	850	867		
GG	269 (31.65)	289 (33.33)	0.44 <sup>b</sup>	1.00
GC	424 (49.88)	406 (46.83)		1.11 (0.89–1.38)
CC	157 (18.47)	172 (19.84)		0.99 (0.74–1.30)
GG	269 (31.65)	289 (33.33)		1.00
CC + GC	581 (68.35)	578 (66.67)	0.51 <sup>b</sup>	1.07 (0.87–1.32)
C allele frequency	43.41	43.25	0.92 <sup>c</sup>	1.01 (0.88–1.16)
rs11951760 A>G	850	881		
AA	260 (30.59)	279 (31.67)	0.64 <sup>b</sup>	1.00
AG	428 (50.35)	424 (48.13)		1.06 (0.85–1.33)
GG	162 (19.06)	178 (20.20)		0.98 (0.74–1.29)
AA	260 (30.59)	279 (31.67)		1.00
GG + AG	590 (69.14)	602 (68.38)	0.71 <sup>b</sup>	1.04 (0.84–1.28)
G allele frequency	44.24	44.27	0.96 <sup>c</sup>	1.00 (0.87–1.14)
rs9790896 G>A	859	876		
GG	281 (32.71)	272 (31.05)	0.51 <sup>b</sup>	1.00
GA	422 (49.13)	427 (48.74)		0.95 (0.76–1.18)
AA	156 (18.16)	177 (20.21)		0.87 (0.66–1.15)
AA	260 (30.59)	279 (31.67)		1.00
AA + GA	578 (67.29)	604 (68.95)	0.45 <sup>b</sup>	0.92 (0.75–1.14)
A allele frequency	42.72	44.58	0.32 <sup>c</sup>	0.93 (0.81–1.07)
rs7703618 G>A	850	876		
GG	441 (51.88)	470 (53.65)	0.37 <sup>b</sup>	1.00
GA	352 (41.41)	337 (38.47)		1.11 (0.91–1.37)
AA	57 (6.71)	69 (7.88)		0.84 (0.57–1.23)
GG	441 (51.88)	470 (53.65)		1.00
AA + GA	409 (48.12)	406 (46.35)	0.53 <sup>b</sup>	1.06 (0.88–1.29)
A allele frequency	27.41	27.11	0.99 <sup>c</sup>	1.00 (0.86–1.17)
rs2067980 A>G	838	881		
AA	224 (26.73)	221 (25.09)	0.47 <sup>b</sup>	1.00

**Table 2 (Continued)**

Genotype	Case N = 878 (%)	Control N = 900 (%)	P	Adjusted OR <sup>a</sup> (95% CI)
AG	406 (48.45)	453 (51.42)		0.85 (0.67–1.08)
GG	208 (24.82)	207 (23.50)		0.98 (0.75–1.30)
AA	224 (26.73)	221 (25.09)		1.00
GG + AG	614 (73.27)	660 (74.91)	0.32 <sup>b</sup>	0.89 (0.72–1.12)
G allele frequency	49.05	49.21	0.88 <sup>c</sup>	0.99 (0.86–1.14)
rs11742346 C>T	848	886		
CC	438 (51.65)	466 (52.60)	0.72 <sup>b</sup>	1.00
CT	336 (39.62)	352 (39.73)		1.03 (0.84–1.27)
TT	74 (8.73)	68 (7.67)		1.12 (0.78–1.61)
CC	438 (51.65)	466 (52.60)		1.00
TT + CT	410 (48.35)	420 (47.40)	0.63 <sup>b</sup>	1.05 (0.86–1.27)
T allele frequency	28.54	27.54	0.54 <sup>c</sup>	1.05 (0.90–1.22)
rs4533894 A>G	839	860		
AA	709 (84.51)	747 (86.86)	0.04 <sup>b</sup>	1.00
AG	109 (12.99)	105 (12.21)		1.05 (0.78–1.41)
GG	21 (2.50)	8 (0.93)		3.13 (1.31–7.48)
AA	709 (84.51)	747 (86.86)		1.00
GG + AG	130 (15.49)	113 (13.14)	0.25 <sup>b</sup>	1.18 (0.89–1.56)
G allele frequency	9.00	7.03	0.07 <sup>c</sup>	1.25 (0.98–1.59)

Abbreviations: CI, confidence interval; OR, odds ratio; SNP, single-nucleotide polymorphism.  
<sup>a</sup>Adjusted for age, age at menarche and menopausal status (natural menopause as one status while unnatural menopause and premenopausal status were merged to the other).

<sup>b</sup>Two-sided  $\chi^2$ -test for difference in frequency distribution of genotypes between cases and controls.

<sup>c</sup>P trend for genotypes between cases and controls.

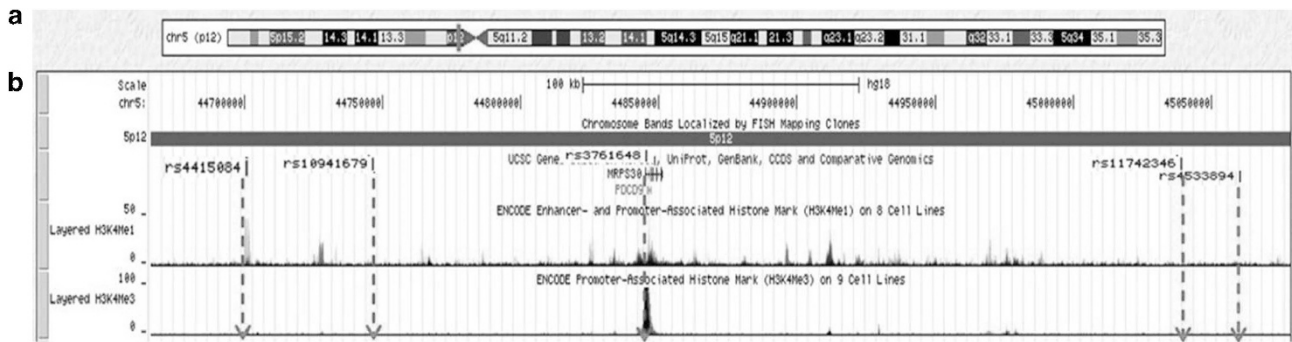
between all these 10 SNPs and four nongenetic factors (age, menopausal status, age at menarche and age at first live birth), however, no significant association was observed after controlling multiple comparisons (data not shown).

We also conducted haplotype-based risk assessment for these 10 SNPs. Linkage disequilibrium among the 10 SNPs were shown in Supplement Table 1. However, no association was observed between different haplotypes and risk of breast cancer (Supplement Table 2).

## DISCUSSION

In our present case-control study consisted of 878 breast cancer cases and 900 controls, we examined the effect of 10 SNPs (rs13357090, rs4415084, rs10941678, rs3747479, rs11951760, rs9790896, rs7703618, rs2067980, rs11742346 and rs4533894) at 5p12 on breast cancer risk in the Chinese population. No significant results were found between these 10 SNPs and breast cancer risk after the multiple test adjustment.

Recently, Stacey *et al.*<sup>8</sup> have identified two low-penetrance loci rs4415084 (T>C) and rs10941679 (A>G) on 5p12 conferring susceptibility to ER-positive breast cancer in the EA population. Subsequently, the effect of SNP rs4415084 on breast cancer risk in the EA population was confirmed by Thomas *et al.*,<sup>9</sup> but not SNP rs10941679. In addition, the associations between these two SNPs and breast cancer risk were also investigated in African Americans.<sup>13,14</sup> For example, the SNP rs10941679 was evaluated in a recent study of African Americans and had no association with breast cancer risk in a combined group of 810 cases and 1784 controls from two separate studies conducted in the Southern United States.<sup>13</sup> An independent case-control study with 886 cases and 1089 controls conducted in African Americans women<sup>14</sup> reported a borderline significant association of rs4415084 with overall risk of breast cancer (Per-allele OR = 1.13, 95% CI = 0.99–1.28). Furthermore, a fine-mapping



**Figure 1** Overview of the linkage disequilibrium (LD) block and the four breast cancer susceptibility SNPs(rs4415084, rs10941679, rs11742346 and rs4533894)and rs3761648 that was highly correlated with the SNP rs4415084. (a) A view of the genomic region of 5p12 (44 666 047–45 078 551) from the UCSC browser build 36 assembly (hg18). (b)The analyzed LD block, its flanking region, related genes, ENCODE enhancer and promoter associated histone mark (H3K4Me1) on eight cell lines (Gm12878, H1ES, HMEC, HSMM, HUVEC, K562, NHEK and NHLF) and ENCODE promoter associated histone mark (H3K4Me3) on nine cell lines (Gm12878, H1ES, HepG2, HMEC, HSMM, HUVEC, K562, NHEK and NHLF). A full color version of this figure is available at the *Journal of Human Genetics* journal online.

analysis of 5p12 from the Shanghai Breast Cancer Study conducted by Long *et al.*<sup>15</sup> reported that rs10941679 showed null association with breast cancer risk, which was consistent with our finding.

The biological mechanism through which genetic variations in 5p12 influences breast cancer risk remains unclear. We examined the four SNPs (rs4415084, rs10941679, rs11742346 and rs4533894) using the UCSC genome browser database (build 36 assembly, hg18; Figure 1). From Figure 1, we can find that the only gene in this region is *MRPS30* (also known as *PDCD9*, programmed cell death protein 9), which encodes a component of the small subunit of the mitochondrial ribosome and has been implicated in apoptosis.<sup>10,11</sup> Some studies have showed that *MRPS30* is not expressed in normal breast luminal epithelial cells, but it is upregulated in infiltrating ductal carcinomas.<sup>16</sup> Moreover, it was also a part of a gene expression profile that differentiated ER-positive from ER-negative breast tumors.<sup>11</sup> Interestingly, a SNP rs3761648 in 5' near of *MRPS30* is highly correlated with the SNP rs4415084 ( $r^2 = 0.83$  in HapMap CHB,  $r^2 = 0.72$  in HapMap CEU) and located at a site of H3K4Me3 histone modification marks. Therefore, we speculated that rs4415084 might be a proxy for some potentially functional SNPs, such as rs3761648, that influence the expression of *MRPS30* and in consequence the cancer risk.

In conclusion, as a fine-mapping study on the associations of SNPs at 5p12 with breast cancer risk in China, none of genetic variants on 5p12 was independently associated with risk of breast cancer. Larger well-designed epidemiological studies as well as functional evaluations are warranted to confirm our findings.

#### ACKNOWLEDGEMENTS

We thank the study participants and research staff for their contributions and commitment to this study. Funding support for this project was provided by the National Natural Science Foundation of China (grant numbers 81071715 and 81102179), Key Grant of Natural Science Foundation of Jiangsu Higher Education Institutions (09KJA330001) and the Program for Changjiang Scholars and Innovative Research Team in University (IRT0631).

- Parkin, D. M., Bray, F., Ferlay, J. & Pisani, P. Global cancer statistics, 2002. *CA. Cancer J. Clin.* **55**, 74–108 (2005).
- Anderson, B. O., Yip, C. H., Smith, R. A., Shyyan, R., Sener, S. F., Eniu, A. *et al.* Guideline implementation for breast healthcare in low-income and middle-income countries: overview of the Breast Health Global Initiative Global Summit 2007. *Cancer* **113**, 2221–2243 (2008).
- Porter, P. 'Westernizing' women's risks? Breast cancer in lower-income countries. *New Engl. J. Med.* **358**, 213–216 (2008).
- Kelsey, J. L. & Gammon, M. D. The epidemiology of breast cancer. *CA. Cancer J. Clin.* **41**, 146–165 (1991).
- McPherson, K., Steel, C. M. & Dixon, J. M. ABC of breast diseases. Breast cancer epidemiology, risk factors, and genetics. *BMJ.* **321**, 624–628 (2000).
- Bray, F., McCarron, P. & Parkin, D. M. The changing global patterns of female breast cancer incidence and mortality. *Breast Cancer Res.* **6**, 229–239 (2004).
- Chen, Y. C. & Hunter, D. J. Molecular epidemiology of cancer. *CA. Cancer J. Clin.* **55**, 45–54, quiz 57 (2005).
- Stacey, S. N., Manolescu, A., Sulem, P., Thorlacius, S., Gudjonsson, S. A., Jonsson, G. F. *et al.* Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat. Genet.* **40**, 703–706 (2008).
- Thomas, G., Jacobs, K. B., Kraft, P., Yeager, M., Wacholder, S., Cox, D. G. *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.* **41**, 579–584 (2009).
- Cavdar Koc, E., Ranasinghe, A., Burkhart, W., Blackburn, K., Koc, H., Moseley, A. *et al.* A new face on apoptosis: death-associated protein 3 and PDCD9 are mitochondrial ribosomal proteins. *FEBS. Lett.* **492**, 166–170 (2001).
- Cavdar Koc, E., Burkhart, W., Blackburn, K., Moseley, A. & Spremulli, L. L. The small subunit of the mammalian mitochondrial ribosome. Identification of the full complement of ribosomal proteins present. *J. Biol. Chem.* **276**, 19363–19374 (2001).
- van't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536 (2002).
- Zheng, W., Cai, Q., Signorello, L. B., Long, J., Hargreaves, M. K., Deming, S. L. *et al.* Evaluation of 11 breast cancer susceptibility loci in African-American women. *Cancer Epidemiol Biomarkers Prev.* **18**, 2761–2764 (2009).
- Ruiz-Narvaez, E. A., Rosenberg, L., Rotimi, C. N., Cupples, L. A., Boggs, D. A., Adeyemo, A. *et al.* Genetic variants on chromosome 5p12 are associated with risk of breast cancer in African American women: the Black Women's Health Study. *Breast Cancer Res. Treat.* **123**, 525–530 (2010).
- Long, J., Shu, X. O., Cai, Q., Gao, Y. T., Zheng, Y., Li, G. *et al.* Evaluation of breast cancer susceptibility loci in Chinese women. *Cancer Epidemiol Biomarkers Prev.* **19**, 2357–2365 (2010).
- Grigoriadis, A., Mackay, A., Reis-Filho, J. S., Steele, D., Iseli, C., Stevenson, B. J. *et al.* Establishment of the epithelial-specific transcriptome of normal and malignant human breast cells based on MPSS and array expression data. *Breast Cancer Res.* **8**, R56 (2006).

Supplementary Information accompanies the paper on *Journal of Human Genetics* website (<http://www.nature.com/jhg>)