ARTICLE

Comparison of genetic variation of breast cancer susceptibility genes in Chinese and German populations

David Barzan[1,4], Marlon R Veldwijk[1], Carsten Herskind[1], Yang Li[2,3], Bo Zhang[1], Elena Sperk[1], Wei-Dong Du[2,3], Xue-Jun Zhang[2,3] and Frederik Wenz*,[1]

Genome-wide association studies (GWAS) identified several genetic risk factors for breast cancer, however, most of them were validated among women of European ancestry. This study examined single-nucleotide polymorphisms (SNPs) contributing to breast cancer in Chinese (984 cases and 2206 controls) and German (311 cases and 960 controls) populations. Eighteen SNPs significantly associated with breast cancer, previously identified in GWAS were genotyped. Twelve SNPs passed quality control and were subjected to statistical analysis. Seven SNPs were confirmed to be significantly associated with breast cancer in the Chinese population, reflecting three independent loci (ESR1, FGFR2, TOX3) and five of these were also confirmed in the German population. The strongest association was identified for rs2046210 in the Chinese (odds ratio (OR) = 1.42, 95% confidence interval (CI) = 1.28–1.59, $P = 1.9 \times 10^{-10}$) and rs3803662 in the German population (OR = 1.43, 95% CI = 1.17–1.74, $P = 4.01 \times 10^{-4}$), located upstream of the ESR1 and TOX3 gene, respectively. For the first time, rs3757318 at 6q25.1, located next to the gene encoding estrogen receptor α (ESR1) was found to be strongly associated with breast cancer (OR = 1.33, 95% CI = 1.18–1.49, $P = 1.94 \times 10^{-6}$) in the Chinese population. The frequency of this variant was markedly lower in the German population and the association was not significant. Despite the genetic differences, essentially the same risk loci were identified in the Chinese and the German populations. Our study suggested the existence of common genetic factors as well as disease susceptibility differences for breast cancer in both populations and highlighted the importance of performing comparison analyses for disease susceptibility within ethnic populations.

INTRODUCTION

Breast cancer is the most common cancer in women and a common cause of cancer-related death worldwide.[1] The risk of breast cancer is determined by both genetic and lifestyle factors.[2] The variation in breast cancer incidence between populations can be explained in part by differences in lifestyle factors, such as reproductive patterns or diet. However, there is substantial variation within a population that seems to be determined by inherited genetic risk factors, possibly modified by external factors.[3] The overall contribution of inherited genes to the development of a disease can be quantified by the familial aggregation of the disease. Epidemiological studies have shown that breast cancer is about twice as common in first-degree relatives of women with the disease than in the general population, reflecting the inheritable component of the disease.[4] A considerably higher risk in monozygotic twins of affected relatives than in dizygotic twins has been demonstrated in twin studies, suggesting that the familial aggregation is largely determined by inheritable rather than environmental risk factors.[5,6]

GWAS plus previous linkage and candidate gene association studies have identified many susceptibility genes, the most prominent being the high-penetrance genes breast cancer-related BRCA1 and BRCA2, which contribute up to 20% of hereditary breast cancer.[7] Further, high-penetrance (TP53 and PTEN) and moderate-penetrance genes (CHEK2, ATM, BRIP1 and PALB2) were shown to predispose to breast cancer susceptibility.[8] However, these genes account only for ~8% of the genetic risk of breast cancer[9] and it is unlikely that any single variant will have a major impact on risk prediction. The residual genetic risk is therefore likely to be due to a large number of common variants. The risk conferred by each of these alleles may be small but may combine in an additive or synergistic fashion to affect breast cancer susceptibility.[10,11]

In the past, breast cancer incidence in China was low but a substantial increase in new cases of breast cancer is expected due to rapidly changing reproductive and lifestyle risk factors among Chinese women. While the overall cancer rate in urban Shanghai decreased by ~0.5% per year between 1972 and 1994, the breast cancer incidence increased by about 50% over the same 23 year period.[12] A recent study on reproductive and demographic changes stated that by the year 2021, the incidence of breast cancer is expected to increase from current rates of 10–60 cases per 100 000 women to more than 100

new cases per 100 000 women aged between 55–69.[13] Incorporation of this disease into the Chinese public health-care system is an urgent need for efficient future health-care planning. Recently, a risk assessment model, which integrated genetic and demographic factors evaluated the importance of genetic risk factors to breast cancer, for screening and prevention programs.[14]

Currently, little is known about the differential role of SNPs contributing to breast cancer in the Chinese population compared with the European population. However, the use of studies from multiple populations with different patterns of linkage disequilibrium (LD) can substantially reduce the number of variants that need to be subjected in post-GWAS functional analysis.[15,16]

In this study, we evaluated the association of candidate SNPs and the risk of breast cancer in Chinese and German cohorts. Eighteen candidate SNPs were selected from previous GWAS on Caucasian and Chinese populations. The objective was to validate the SNPs in an independent German cohort (311 cases *vs* 960 controls) and to identify SNPs not previously associated with breast cancer risk in the Chinese population (984 cases *vs* 2206 controls) and vice versa.

MATERIALS AND METHODS

Study population

Two independent study groups were evaluated in this work. All Chinese samples used for genotyping were Chinese Han. These included 984 breast cancer cases and 2206 healthy controls obtained by doctors through collaborations with multiple hospitals from provinces in the central area of China. All Chinese controls were clinically assessed to be without breast cancer, other neoplastic diseases, systemic disorders or family history of neoplastic diseases (including first-, second- and third-degree relatives). All breast cancer patients were diagnosed and categorized according to the TNM breast cancer classification. Clinical information was collected from the affected individuals through a full clinical checkup by breast cancer specialists. Additional demographic information was collected from cases through a structured questionnaire. The German cases were obtained from the University Medical Center Mannheim and other hospitals in the German Rhein-Neckar region and included 311 cases and 960 controls. German controls were derived from healthy blood donors, obtained by the German Red Cross and have partly been used as control groups in previous studies on breast cancer.[17–19] The characteristics of the two study populations are presented in Table 1. All of the cases and controls were females. Cases with one or more first-degree relatives having breast or ovarian cancer were considered to have a family history of breast cancer. All participants provided written informed consent. The study was approved by the Institutional Ethical Committee of Anhui Medical University and the Medical Ethics Commission II of the Medical Faculty of Mannheim, University Heidelberg and was conducted according to Declaration of Helsinki principles.

SNP selection

In recent years, several GWAS have identified more than 27 SNPs within 20 different loci to be associated with the risk of breast cancer. Most of these studies were conducted on women of European descent. In this study, 18 SNPs from previously published GWAS in the European[20–27] and Chinese[28,29] populations were selected (Table 2). Only SNPs with a minor allele frequency (MAF) higher than 5% in the HapMap CHB or CEU data were selected. These 18 SNPs represent 13 independent loci, either present in genes or in intergenic regions.

Genotyping and quality controls

Genotyping analyses were conducted by using the Sequenom MassArray system at the State Key Laboratory Incubation Base of Dermatology, Ministry of National Science and Technology, Hefei, Anhui, China. Genomic DNA was extracted from whole-blood or buffy coat using FlexiGene DNA kits (QIAGEN, Hilden, Germany). All samples were surveyed for DNA quality

Table 1 Baseline patient characteristics of breast cancer cases and controls

Characteristics	Chinese	German
Cases		
No.	984	311
Mean age (years)	47.9 ± 10.3	60.6 ± 10.0
TNM staging (%)		
Tis	3.5%	3.9%
T1	21.2%	68.4%
T2	67.2%	26.1%
T3	6.7%	0.6%
T4	1.4%	1.0%
N0	48.8%	81.6%
N1	38.3%	14.8%
N2	10.8%	2.3%
N3	2.1%	1.3%
M0	85.7%	99.7%
M1	14.3%	0.3%
Familial history of breast cancer (%)		
Familial	7.7%	12.6%
Sporadic	92.2%	87.4%
Controls		
No.	2206	960
Mean age (years)	34.4 ± 9.1	58.5 ± 5.6

Table 2 Information about the 18 candidate SNPs selected

Chromosome/position[a]	SNP	Allele[b]	Associated gene	Source
1p11.2/128457034	rs1562430	T/C	*Intergenic*	Turnbull[27]
2q35/217614077	rs13387042	G/A	*Intergenic*	Stacey[30]
5p12/44742255	rs10941679	G/A	MRPS30	Stacey[25]
5q11.2/56067641	rs889312	C/A	MAP3K1	Easton[21]
5q11.2/56058840	rs16886165	T/G	MAP3K1	Thomas[26]
6q22.33/127642323	rs2180341	G/A	ECHDC1	Gold[22]
6q25.1/151990059	rs2046210	A/G	ESR1	Zheng[29]
6q25.1/151955806	rs3757318	G/A	ESR1	Turnbull[27]
8q24.21/128424800	rs13281615	G/A	*Intergenic*	Easton[21]
9p21.3/22052134	rs1011970	G/T	CDKN2A	Turnbull[27]
10q22.3/80511154	rs704010	T/C	ZMIZ1	Turnbull[27]
10q26.13/123336180	rs1219648	G/A	FGFR2	Hunter[23]
10q26.13/123342307	rs2981582	A/G	FGFR2	Easton[21]
12q21.1/72276104	rs1154865	C/G	*Intergenic*	Murabito[24]
16q12.1/51156689	rs4784227	C/T	TOX3	Long[28]
16q12.1/51143842	rs3803662	A/G	TOX3	Easton[21]
16q12.1/51091668	rs8051542	T/C	TOX3	Easton[21]
17q22/50416621	rs9902718	C/T	COX11	Ahmed[20]

Abbreviation: SNP, single-nucleotide polymorphisms.
[a]Chromosome base position according to NCBI Genome build 36.3.
[b]Risk allele/reference allele, as initially reported.

using a Nanodrop Spectrophotometer ND-2000 (Thermo Scientific, Wilmington, DE, USA) and by agarose gel electrophoresis assay to ensure genomic integrity. Approximately 15 ng of genomic DNA was used to genotype each sample. Locus-specific PCR and detection primers were designed using the MassARRAY Assay Design 3.0 software (Sequenom, San Diego, CA, USA). Following the manufacturer's instructions, the DNA samples were amplified by multiplex PCR reactions, and the PCR products were then used for locus-specific single-base extension reactions. The resulting products were desalted

and transferred to a 384-element SpectroCHIP array. Allele detection was performed using MALDI-TOF MS. The mass spectrograms were analyzed by the MassARRAY Typer software (Sequenom). Exclusion criteria for genotyped SNPs were a call rate <95%, MAF <0.05 and deviation from Hardy–Weinberg equilibrium (HWE, $P<0.05$) in the controls. 12 SNPs passed the quality control and were subjected to statistical analysis.

Statistical analysis

The association between the SNPs and disease susceptibility were assessed using the χ^2 test or Fisher's exact test. The strength of association was estimated by calculating odds ratio (OR) with 95% confidence interval (CI). As a quality control of the genotyping, genotype data were analyzed for deviations from Hardy–Weinberg equilibrium using χ^2 statistics. All alleles were observed to be in normal Hardy–Weinberg equilibrium. Association with subphenotype (age of onset, clinical stage of breast cancer) were analyzed by comparing cases with a certain subphenotype with controls. All statistics were analyzed with SPSS 13.0 (SPSS Inc. Chicago, IL, USA) and Plink 1.07 (http://pngu.mgh. harvard.edu/~purcell/plink/) software packages. Haplotype analyses were performed using Haploview v4.2 to generate haplotype frequencies and calculate the significance of the associations.

RESULTS

Clinical evaluation of patients and controls

Two populations were evaluated in this study, participants recruited from the Anhui province in China and German patients treated at the University Medical Center Mannheim were included. The age of German controls was well matched to German breast cancer patients with a mean age of 60.6 years. As shown in Table 1, the percentage of patients with carcinoma *in situ* (Tis) was similar in the Chinese and German population, but tumor size (T1–T4) was larger and infiltration of lymph nodes (N0–N3) and metastatic spread (M0, M1) were increased in the Chinese cohort. These may be in part due to the breast cancer screening program established in Germany in the last decade, which results in the diagnosis of breast cancer at an early, non-invasive stage.

Association analyses

Six of the 18 SNPs (Table 2) were excluded for further analyses, because they did not pass quality control due to a significant deviation from HWE (rs1562430, rs889312, rs1011970 and rs2180341) or showed a low call rate (rs9902718 and rs13281615). All of the remaining SNPs genotyped in this study showed allele frequencies similar to the HapMap data (Supplementary Table 1)

thereby confirming that the recruitment of our study cohort is representative for the Chinese and the German population, respectively. Of the 12 SNPs analyzed in this study, seven SNPs (rs2046210, rs3757318, rs4784227, rs1219648, rs3803662, rs8051542 and rs2981582) in the Chinese population and five SNPs (rs2046210, rs4784227, rs1219648, rs3803662 and rs8051542) in the German population were significantly associated with breast cancer (Table 3).

Three common susceptibility loci between the Chinese and the European population have been confirmed in this study. The five breast cancer susceptibility SNPs identified in the German population represent three independent genetic loci, ESR1 (6q25.1; rs2046210), FGFR2 (10q26.13; rs1219648) and TOX3 (16q12.1; rs3803662, rs8051542, rs4784227) and were also significant in the Chinese population plus two additional SNPs (rs3757318 and rs2981582) in the same loci (ESR1 and FGFR2, respectively).

ESR1 locus

The present study identified a significant association for rs2046210 on chromosome 6q25.1 in both populations (Table 3). In the Chinese population rs2046210 at ESR1[29] resulted in our study in a highly significant $P=1.9 \times 10^{-10}$ and an OR (95% CI) of 1.42 (1.28–1.59) as well as in the German population $P=3.62 \times 10^{-2}$ and OR (95% CI) = 1.23 (1.01–1.48) (Table 3). This SNP showed the highest significant level and the strongest association with breast cancer risk in the Chinese population in our study sample. However, in the German population this SNP had the lowest association within the significant SNPs and was only marginally significant. Another SNP, rs3757318, which was located 200 kb upstream of the transcription start site of the gene ESR1, showed a significant association in the Chinese population ($P=1.94 \times 10^{-6}$) with an OR (95% CI) of 1.33 (1.18–1.49), but was not significantly associated with breast cancer in the German study group ($P=1.74 \times 10^{-1}$). However, the allele frequency of rs3757318 was much lower in the German (MAF = 0.08 in the controls) than in the Chinese population (MAF = 0.26 in the controls).

The two SNPs in this locus, rs2046210 and rs3757318, are in weak linkage disequilibrium in the Asian population ($r^2 = 0.50$) and in the Caucasian population ($r^2 = 0.09$). The pattern of linkage disequilibrium of our study populations is consistent with the HapMap data (Supplementary Table 2) and indicates that the two SNPs in this region are independent of each other. Haplotype analyses of two SNPs

Table 3 Result of genotyped SNPs in Chinese and German populations

| SNP | Chromo-some | Gene | Chinese | | | | | | German | | | | | |
			Minor allele	MAF (%) cases	MAF (%) controls	Major allele	OR (95% CI)	P-value*	Minor allele	MAF (%) cases	MAF (%) controls	Major allele	OR (95% CI)	P-value*
rs2046210	6q25.1	ESR1	T	0.44	0.36	C	1.42 (1.28–1.59)	*1.9 × 10⁻¹⁰*	T	0.37	0.33	C	1.23 (1.01–1.48)	*3.6 × 10⁻²*
rs4784227	16q12.1	TOX3	T	0.29	0.24	C	1.31 (1.16–1.47)	*9.3 × 10⁻⁶*	T	0.29	0.25	C	1.25 (1.02–1.54)	*2.9 × 10⁻²*
rs1219648	10q26.13	FGFR2	G	0.45	0.40	A	1.23 (1.11–1.37)	*1.4 × 10⁻⁴*	G	0.46	0.40	A	1.26 (1.05–1.51)	*1.5 × 10⁻²*
rs3803662	16q12.1	TOX3	C	0.30	0.34	T	0.83 (0.74–0.93)	*1.3 × 10⁻³*	T	0.34	0.26	C	1.43 (1.17–1.74)	*4.0 × 10⁻⁴*
rs8051542	16q12.1	TOX3	T	0.21	0.18	C	1.19 (1.04–1.36)	*1.0 × 10⁻²*	T	0.51	0.44	C	1.30 (1.08–1.56)	*5.2 × 10⁻³*
rs3757318	6q25.1	ESR1	A	0.31	0.26	G	1.33 (1.18–1.49)	*1.9 × 10⁻⁶*	A	0.10	0.08	G	1.24 (0.91–1.70)	1.7 × 10⁻¹
rs2981582	10q26.13	FGFR2	T	0.37	0.33	C	1.20 (1.07–1.34)	*1.4 × 10⁻³*	T	0.43	0.40	C	1.11 (0.92–1.34)	2.6 × 10⁻¹
rs10941679	5p12	MRPS30	A	0.47	0.49	G	0.92 (0.83–1.03)	1.5 × 10⁻¹	G	0.29	0.26	A	1.20 (0.98–1.47)	7.2 × 10⁻²
rs16886165	5q11.2	MAP3K1	G	0.32	0.33	T	0.96 (0.86–1.08)	5.4 × 10⁻¹	G	0.17	0.15	T	1.18 (0.92–1.50)	1.9 × 10⁻¹
rs704010	10q22.3	ZMIZ1	A	0.30	0.29	G	1.03 (0.91–1.15)	6.6 × 10⁻¹	A	0.36	0.38	G	0.92 (0.76–1.11)	3.7 × 10⁻¹
rs13387042	2q35	Intergenic	A	0.12	0.11	G	1.03 (0.87–1.22)	7.2 × 10⁻¹	G	0.44	0.47	A	0.89 (0.74–1.07)	2.2 × 10⁻¹
rs1154865	12q21.1	Intergenic	G	0.14	0.15	C	0.98 (0.84–1.14)	7.8 × 10⁻¹	G	0.20	0.23	C	0.83 (0.66 –1.04)	1.1 × 10⁻¹

Abbreviations: CI, confidence interval; MAF, minor allele frequency; OR, odds ratio; SNP, single-nucleotide polymorphism.
*$P<0.05$ are shown in italics.

Table 4 Haplotype analysis

| Locus | SNPs | Haplotype | Chinese | | | | German | | | |
			Frequency cases	Frequency controls	P-value	OR	Frequency cases	Frequency controls	P-value	OR
ESR1	rs3757318 | rs2046210	GC	0.53	0.62	8.3×10^{-11}	0.69	0.63	0.67	2.5×10^{-2}	0.80
		AT	0.29	0.23	1.7×10^{-6}	1.35	0.09	0.08	3.2×10^{-1}	1.17
		GT	0.15	0.12	1.6×10^{-3}	1.29	0.28	0.25	7.0×10^{-2}	1.21
		AC	0.03	0.02	5.7×10^{-1}	1.11	—	—		—
FGFR2	rs1219648 | rs2981582	AC	0.55	0.60	1.2×10^{-4}	0.81	0.55	0.60	2.1×10^{-2}	0.82
		GT	0.36	0.32	1.9×10^{-3}	1.19	0.43	0.39	1.7×10^{-1}	1.15
		GC	0.09	0.08	1.4×10^{-1}	1.16	0.03	0.01	5.4×10^{-5}	3.40
TOX3	rs8051542 | rs3803662 | rs4784227	TTT	0.18	0.14	2.1×10^{-4}	1.32	0.25	0.20	6.8×10^{-3}	1.36
		CCC	0.28	0.33	7.3×10^{-4}	0.81	0.43	0.51	9.4×10^{-4}	0.73
		CTT	0.11	0.09	6.0×10^{-2}	1.19	0.04	0.04	4.8×10^{-1}	0.80
		TTC	0.02	0.02	9.8×10^{-2}	0.69	0.02	0.01	1.2×10^{-1}	1.84
		TCC	0.01	0.01	4.7×10^{-1}	0.80	0.23	0.23	8.0×10^{-1}	1.03
		CTC	0.40	0.40	7.8×10^{-1}	0.98	0.02	0.01	2.9×10^{-4}	4.51

Abbreviations: OR, odds ratio; SNPs, single-nucleotide polymorphism.

in the Chinese samples found one protective haplotype GC, which had a higher frequency in cases (53%) and in controls (62%). This haplotype showed a stronger association evidence for breast cancer in the Chinese population ($P = 8.3 \times 10^{-11}$, OR = 0.69; Table 4).

FGFR2 locus
Two SNPs in the FGFR2 locus, rs2981582 and rs1219648, were identified in this study to be associated in cases with a strong family history of breast cancer. SNP rs2981582, which lies within intron 2 of FGFR2, showed a significant association in the Chinese population ($P = 1.39 \times 10^{-3}$) with an OR (95% CI) of 1.20 (1.07–1.34), but did not reach significance in the German population ($P = 2.6 \times 10^{-1}$). The other SNP in this locus, rs1219648, displayed a significant association in the Chinese ($P = 1.41 \times 10^{-4}$) and the German population ($P = 1.5 \times 10^{-2}$) with a similar OR for each SNP (1.23 (1.11–1.37) and 1.26 (1.05–1.51), respectively). Haplotype analyses of two SNPs in the German samples found one risk haplotype GC, which had a lower frequency in cases (3%) and in controls (1%). This haplotype showed a stronger association evidence for breast cancer in the German population ($P = 5.4 \times 10^{-5}$, OR = 3.4) (Table 4).

TOX3 locus
All three SNPs within the TOX3 locus (rs8051542, rs3803662 and rs4784227) showed a significant association for breast cancer in both populations. In the Chinese population, the strongest effect was found for rs4784227 with an OR (95% CI) of 1.31 (1.16–1.47), being significant at $P = 9.3 \times 10^{-6}$. Whereas the significance level in the German population was lower ($P = 2.9 \times 10^{-2}$), mainly due to the smaller sample size. In the German population, the overall strongest association ($P = 4.01 \times 10^{-4}$) in this study was found for rs3803662 with an OR (95% CI) of 1.43 (1.17–1.74). The third SNP in the TOX3 locus, rs8051542, had a higher effect size (OR (95% CI) = 1.30 (1.08–1.56)) and significance ($P = 5.2 \times 10^{-3}$) in the German population compared with the Chinese population.

A computational analysis of haplotypes found the most common haplotype in the German population (CCC) was present in 43% of all

cases and 51% of all controls, which conferred significant protection against breast cancer (OR = 0.73; $P = 9.4 \times 10^{-4}$) (Table 4).

Stratification
Stratification for age of onset was performed by dividing Chinese and German patients in groups younger and older than 50 years (Table 5). Analysis of the Chinese samples resulted in a higher effect size of rs4784227 ($P = 1.37 \times 10^{-7}$, OR = 1.49) in patients < 50 years than in total samples ($P = 9.3 \times 10^{-6}$, OR = 1.31) and no association in cases ≥ 50 years ($P = 4.31 \times 10^{-1}$, OR = 1.08), suggesting that this SNP confers a higher risk for the development of breast cancer for younger patients. By contrast, rs3757318 displayed a higher risk (OR = 1.44) in cases older than 50 years compared with total samples (OR = 1.33). In German patients younger than 50 years, rs2046210 and rs1219648 showed significant P-values and an increased effect size for breast cancer whereas risk alleles for rs3803662 and rs8051542 were enriched in patients older than 50 years. In addition, stratification of patients was performed according to clinical stage (Table 6). SNP rs2046210 and rs3757318 were significant and had a higher effect size for breast cancer in Stage 0 and Stage IV breast cancer cases compared with the combined analysis, suggesting that risk variants associated with the ESR1 locus were overrepresented in Stage 0 and Stage IV in the Chinese population. SNP rs1219648 and rs2981582 had a higher OR in Stage 0, suggesting a role for FGFR2 in early stages of breast cancer. The association of all significant SNPs in Stage II did not differ from all samples and stratification of Stage III samples resulted in no significant SNPs, which was due to the too small number (107 cases) belonging to this stage. Additionally, three SNPs (rs4784227, rs3803662 and rs8051542) were significant in Stage I with an increased OR and were also significant in the total sample analysis.

DISCUSSION
The association of 18 SNPs previously shown to be associated with breast cancer were evaluated, of which 12 SNPs passed quality control. Out of these, seven SNPs could be significantly confirmed in the Chinese population and five SNPs in the German population. To our

Table 5 Stratification—age of onset

	Chinese			
	< 50 years old (601 cases/2206 controls)		≥ 50 years old (383 cases/2206 controls)	
SNP	P-value	OR (95% CI)	P-value	OR (95% CI)
rs2046210	4.27×10^{-8}	1.48 (1.28–1.70)	2.17×10^{-4}	1.37 (1.16–1.62)
rs4784227	1.37×10^{-7}	1.49 (1.29–1.73)	4.31×10^{-1}	1.08 (0.89–1.31)
rs1219648	1.08×10^{-2}	1.20 (1.04–1.38)	3.38×10^{-3}	1.28 (1.09–1.51)
rs3803662	4.72×10^{-3}	0.80 (0.69–0.93)	5.31×10^{-2}	0.84 (0.70–1.00)
rs8051542	4.86×10^{-2}	1.19 (1.00–1.41)	2.40×10^{-1}	1.13 (0.92–1.39)
rs3757318	1.10×10^{-3}	1.29 (1.11–1.49)	5.68×10^{-5}	1.44 (1.20–1.72)
rs2981582	1.88×10^{-2}	1.19 (1.03–1.37)	2.01×10^{-2}	1.23 (1.03–1.46)

	German			
	< 50 years old (46 cases/960 controls)		≥ 50 years old (265 cases/960 controls)	
SNP	P-value	OR (95% CI)	P-value	OR (95% CI)
rs2046210	3.26×10^{-2}	1.59 (1.04–2.44)	1.17×10^{-1}	1.18 (0.96–1.44)
rs4784227	3.20×10^{-1}	0.77 (0.45–1.30)	6.36×10^{-3}	1.35 (1.09–1.67)
rs1219648	6.65×10^{-3}	1.79 (1.17–2.73)	8.04×10^{-2}	1.19 (0.98–1.45)
rs3803662	5.50×10^{-1}	0.86 (0.51–1.43)	4.66×10^{-5}	1.54 (1.25–1.90)
rs8051542	1.00×10^{-1}	0.69 (0.45–1.08)	2.99×10^{-4}	1.43 (1.18–1.74)

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

Table 6 Stratification – Clinical stage of breast cancer (A) Chinese; (B) German

(A)

	Stage 0 (48 cases/2206 controls)		Stage I (156 cases/2206 controls)		Stage II (550 cases/2206 controls)		Stage III (107 cases/2206 controls)		Stage IV (123 cases/2206 controls)	
SNP	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)
rs2046210	9.96×10^{-5}	2.24 (1.48–3.39)	1.14×10^{-1}	1.22 (0.95–1.55)	7.84×10^{-6}	1.38 (1.19–1.58)	4.32×10^{-1}	1.13 (0.83–1.54)	3.57×10^{-5}	1.74 (1.33–2.26)
rs4784227	3.68×10^{-1}	1.23 (0.78–1.95)	5.95×10^{-3}	1.43 (1.11–1.85)	3.35×10^{-4}	1.32 (1.13–1.54)	7.31×10^{-2}	1.35 (0.97–1.87)	3.06×10^{-1}	1.17 (0.86–1.57)
rs1219648	3.06×10^{-2}	1.59 (1.04–2.42)	1.72×10^{-1}	1.18 (0.92–1.50)	4.15×10^{-3}	1.23 (1.06–1.41)	4.45×10^{-1}	1.13 (0.83–1.52)	9.59×10^{-1}	0.99 (0.75–1.30)
rs3803662	9.63×10^{-1}	0.99 (0.63–1.54)	5.58×10^{-1}	0.77 (0.59–1.01)	2.84×10^{-2}	0.85 (0.72–0.98)	3.13×10^{-1}	0.84 (0.60–1.17)	3.70×10^{-1}	0.88 (0.66–1.16)
rs8051542	7.09×10^{-1}	1.10 (0.66–1.86)	2.03×10^{-1}	1.21 (0.90–1.62)	2.56×10^{-2}	1.22 (1.02–1.44)	3.17×10^{-1}	1.21 (0.83–1.75)	4.94×10^{-1}	1.12 (0.80–1.56)
rs3757318	1.49×10^{-2}	1.69 (1.10–2.60)	2.21×10^{-1}	1.18 (0.90–1.53)	3.01×10^{-4}	1.32 (1.13–1.53)	4.70×10^{-1}	1.13 (0.80–1.58)	4.86×10^{-3}	1.49 (1.12–1.97)
rs2981582	3.53×10^{-2}	1.57 (1.03–2.39)	1.19×10^{-1}	1.22 (0.95–1.55)	1.23×10^{-2}	1.20 (1.04–1.38)	5.01×10^{-1}	1.11 (0.81–1.52)	2.15×10^{-1}	0.83 (0.62–1.11)

(B)

	Stage I (165 cases/960 controls)		Stage II (78 cases/960 controls)	
SNP	P-value	OR (95% CI)	P-value	OR (95% CI)
rs2046210	1.16×10^{-1}	1.22 (0.95–1.55)	3.84×10^{-1}	1.16 (0.82–1.63)
rs4784227	8.33×10^{-3}	1.41 (1.09–1.82)	6.39×10^{-1}	1.09 (0.75–1.59)
rs1219648	8.50×10^{-2}	1.23 (0.97–1.55)	3.88×10^{-1}	1.16 (0.83–1.60)
rs3803662	3.45×10^{-4}	1.58 (1.22–2.02)	1.19×10^{-1}	1.33 (0.92–1.90)
rs8051542	2.52×10^{-3}	1.44 (1.13–1.82)	2.19×10^{-2}	1.46 (1.05–2.03)

Abbreviations: CI, confidence interval; OR, odds ratio; SNPs, single-nucleotide polymorphisms.

knowledge, we have shown for the first time that rs3757318, reported to be associated with breast cancer risk in the European population, is also correlated with breast cancer in the Chinese population. These results extend the findings from previous GWAS in European descends to the Chinese population and may help to capture the causal variants in fine-mapping approaches.

In 2010 Turnbull *et al*[27] identified for the first time SNP rs3757318 in the Caucasian population with an OR (95% CI) of 1.30 (1.17–1.46). Interestingly, no significant association with breast cancer risk could be replicated for this SNP in the pooled analysis of the German samples, albeit after stratification, a significant association was observed in Stage II breast cancer cases. However, a

similar association (OR = 1.33) was obtained for our Chinese study cohort. This SNP is located at chromosome 6q25.1, 200 kb upstream of the transcription start site of the gene encoding estrogen receptor 1 (ESR1). In 2009, Zheng *et al*[29] conducted the first GWAS of breast cancer with an Asian population-based background in which a highly significant association ($P = 2.0 \times 10^{-15}$) was identified for rs2046210, a SNP located 180 kb upstream of ESR1. In our Chinese samples the strongest association was identified for rs2046210 with a slightly lower *P*-value ($P = 1.9 \times 10^{-10}$), potentially caused by the smaller study sample size. The minor allele frequency of rs3757318 was much lower in the German (MAF = 0.08 in the controls) than in the Chinese population (MAF = 0.26 in the controls). Taking into account the smaller German sample size, the power to identify a significant association for rs3757318 may not have been high enough for the German study group.

The two SNPs in this locus, rs2046210 and rs3757318, are in weak LD in the German and Chinese populations ($r^2 = 0.15$ and 0.44, respectively). The pattern of LD of our study populations is consistent with the HapMap data (Supplementary Table 1) and indicates that the two SNPS in this region are independent of each other. Interestingly, the ESR1 locus has recently been confirmed in East Asians by Long *et al*.[31] However, in this study SNP rs9383951 located in Intron 5 of the ESR1 gene showed the strongest association with breast cancer. ESR1 is a hormone receptor, which functions as a transcription factor upon ligand binding. The estrogen receptors have a central role in the pathology of breast cancer. Several studies have focused on the role of genetic variants in ESR1 and an increased breast cancer risk in the Asian[32–34] and the Caucasian population.[35,36] The close location of rs3757318 to ESR1 suggests that correlated variants change expression patterns of ESR1 and thereby increase breast cancer risk.

The first GWAS on breast cancer in the Caucasian population identified SNP rs2981582 to be associated with breast cancer in cases with a strong family history of breast cancer.[21] This SNP lies in an LD block within intron 2 of FGFR2. At the same time, Hunter *et al*[23] identified rs1219648, located 6 kb upstream of rs2981582, as being significantly associated with sporadic breast cancer in postmenopausal women. SNP rs2981582 and rs1219648 were highly significant in the pooled analysis ($P = 2 \times 10^{-76}$ and $P = 4.2 \times 10^{-10}$) with a heterogeneous OR of 1.23 each, being similar in our Chinese and German cases (OR = 1.23 and 1.26, respectively). FGFR2 is a tyrosine receptor kinase with several alternatively spliced variants, resulting in differential ligand binding and signal transduction.[37] Amplification and overexpression of FGFR2 have been observed in primary tumors of the breast[38] and breast cancer cell lines.[39,40] Interestingly, expression of FGFR2 has been implicated in the development of estrogen receptor-positive breast cancer,[20,41,42] suggesting an interaction with genetic breast cancer risk variants involved in the metabolism of sex hormones.

The third locus identified in this work involves three SNPs located close to or in the open reading frame of TOX3 on chromosome 16q12.1. The pattern of LD in this region is very different in Asians from that in the Caucasian population (Supplementary Table 2). In the Chinese samples, SNP rs4784227 is in low LD with rs3803662 and rs8051542, with r^2 being 0.14 and 0.37, respectively. In the German samples, however, SNP rs4784227 is in strong LD with rs3803662, having an r^2 value of 0.86, whereas linkage with rs8051542 is low ($r^2 = 0.20$). Loss of chromosomal material on the long arm of chromosome 16 was observed frequently in several cancers, including breast, prostate, ovarian and fallopian tubes cancer.[43,44] The function of TOX3 is unknown, but a putative high-mobility group motif suggests that it might act as a transcription factor or is involved in the alteration of the chromatin structure.[45]

In conclusion, using a Chinese and a German population-based case-control study, SNPs in the Chinese and German population associated with breast cancer were identified. These results provide additional support that the association of SNPs detected in Europeans is also found in the Chinese population. In spite of the difference in segregation of complex traits between the Chinese and German population, in principle the same loci were identified for the association of breast cancer. Although common variants confer only a minor increase of risk, a combination of genetic risk variants might be useful as a powerful predictor for the development of breast cancer. Fine-mapping is an essential step in the identification of a functional variant, which may help to improve clinical therapies for breast cancer.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

1 Key TJ, Verkasalo PK, Banks E: Epidemiology of breast cancer. *Lancet Oncol* 2001; **2**: 133–140.
2 Fletcher O, Houlston RS: Architecture of inherited susceptibility to common cancer. *Nat Rev Cancer* 2010; **10**: 353–361.
3 Narod SA: Modifiers of risk of hereditary breast and ovarian cancer. *Nat Rev Cancer* 2002; **2**: 113–123.
4 Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58 209 women with breast cancer and 101 986 women without the disease. *Lancet* 2001; **358**: 1389–1399.
5 Peto J, Mack TM: High constant incidence in twins and other relatives of women with breast cancer. *Nat Genet* 2000; **26**: 411–414.
6 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**: 78–85.
7 Kirova YM, Savignoni A, Sigal-Zafrani B *et al*: Is the breast-conserving treatment with radiotherapy appropriate in BRCA1/2 mutation carriers? Long-term results and review of the literature. *Breast Cancer Res Treat* 2010; **120**: 119–126.
8 Zhang B, Beeghly-Fadiel A, Long J, Zheng W: Genetic variants associated with breast-cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. *Lancet Oncol* 2011; **12**: 477–488.
9 Stratton MR, Rahman N: The emerging landscape of breast cancer susceptibility. *Nat Genet* 2008; **40**: 17–22.
10 Pharoah PD, Antoniou AC, Easton DF, Ponder BA: Polygenes, risk prediction, and targeted prevention of breast cancer. *N Engl J Med* 2008; **358**: 2796–2803.
11 McCarthy MI, Abecasis GR, Cardon LR *et al*: Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008; **9**: 356–369.
12 Jin F, Devesa SS, Chow WH *et al*: Cancer incidence trends in urban shanghai, 1972-1994: an update. *Int J Cancer* 1999; **83**: 435–440.
13 Linos E, Spanos D, Rosner BA *et al*: Effects of reproductive and demographic changes on breast cancer incidence in China: a modeling analysis. *J Natl Cancer Inst* 2008; **100**: 1352–1360.
14 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**: 972–981.
15 Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M: Genome-wide association studies in diverse populations. *Nat Rev Genet* 2010; **11**: 356–366.
16 Freedman ML, Monteiro AN, Gayther SA *et al*: Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet* 2011; **43**: 513–518.
17 Tchatchou S, Jung A, Hemminki K *et al*: A variant affecting a putative miRNA target site in estrogen receptor (ESR) 1 is associated with breast cancer risk in premenopausal women. *Carcinogenesis* 2009; **30**: 59–64.
18 Yang R, Dick M, Marme F *et al*: Genetic variants within miR-126 and miR-335 are not associated with breast cancer risk. *Breast Cancer Res Treat* 2011; **127**: 549–554.
19 Frank B, Wiestler M, Kropp S *et al*: Association of a common AKAP9 variant with breast cancer risk: a collaborative analysis. *J Natl Cancer Inst* 2008; **100**: 437–442.
20 Ahmed S, Thomas G, Ghoussaini M *et al*: Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet* 2009; **41**: 585–590.

21 Easton DF, Pooley KA, Dunning AM *et al*: Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007; **447**: 1087–1093.

22 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 USA* 2008; **105**: 4340–4345.

23 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**: 870–874.

24 Murabito JM, Rosenberg CL, Finger D *et al*: A genome-wide association study of breast and prostate cancer in the NHLBI's Framingham Heart Study. *BMC Med Genet* 2007; **8**:Suppl 1 S6.

25 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**: 703–706.

26 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**: 579–584.

27 Turnbull C, Ahmed S, Morrison J *et al*: Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet* 2010; **42**: 504–507.

28 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**: e1001002.

29 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**: 324–328.

30 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**: 865–869.

31 Long J, Cai Q, Sung H *et al*: Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer. *PLoS Genet* 2012; **8**: e1002532.

32 Cai Q, Shu XO, Jin F *et al*: Genetic polymorphisms in the estrogen receptor alpha gene and risk of breast cancer: results from the Shanghai Breast Cancer Study. *Cancer Epidemiol Biomarkers Prev* 2003; **12**: 853–859.

33 Cai Q, Gao YT, Wen W *et al*: Association of breast cancer risk with a GT dinucleotide repeat polymorphism upstream of the estrogen receptor-alpha gene. *Cancer Res* 2003; **63**: 5727–5730.

34 Shin A, Kang D, Nishio H *et al*: Estrogen receptor alpha gene polymorphisms and breast cancer risk. *Breast Cancer Res Treat* 2003; **80**: 127–131.

35 Andersen TI, Heimdal KR, Skrede M, Tveit K, Berg K, Borresen AL: Oestrogen receptor (ESR) polymorphisms and breast cancer susceptibility. *Hum Genet* 1994; **94**: 665–670.

36 Gold B, Kalush F, Bergeron J *et al*: Estrogen receptor genotypes and haplotypes associated with breast cancer risk. *Cancer Res* 2004; **64**: 8891–8900.

37 Gilbert E, Del Gatto F, Champion-Arnaud P, Gesnel MC, Breathnach R: Control of BEK and K-SAM splice sites in alternative splicing of the fibroblast growth factor receptor 2 pre-mRNA. *Mol Cell Biol* 1993; **13**: 5461–5468.

38 Adnane J, Gaudray P, Dionne CA *et al*: BEK and FLG, two receptors to members of the FGF family, are amplified in subsets of human breast cancers. *Oncogene* 1991; **6**: 659–663.

39 Moffa AB, Tannheimer SL, Ethier SP: Transforming potential of alternatively spliced variants of fibroblast growth factor receptor 2 in human mammary epithelial cells. *Mol Cancer Res* 2004; **2**: 643–652.

40 Tannheimer SL, Rehemtulla A, Ethier SP: Characterization of fibroblast growth factor receptor 2 overexpression in the human breast cancer cell line SUM-52PE. *Breast Cancer Res* 2000; **2**: 311–320.

41 Zhang Y, Gorry MC, Post JC, Ehrlich GD: Genomic organization of the human fibroblast growth factor receptor 2 (FGFR2) gene and comparative analysis of the human FGFR gene family. *Gene* 1999; **230**: 69–79.

42 Tamaru N, Hishikawa Y, Ejima K *et al*: Estrogen receptor-associated expression of keratinocyte growth factor and its possible role in the inhibition of apoptosis in human breast cancer. *Lab Invest* 2004; **84**: 1460–1471.

43 Cleton-Jansen AM, Buerger H, Haar N *et al*: Different mechanisms of chromosome 16 loss of heterozygosity in well- versus poorly differentiated ductal breast cancer. *Genes Chromosomes Cancer* 2004; **41**: 109–116.

44 Rakha EA, Green AR, Powe DG, Roylance R, Ellis IO: Chromosome 16 tumor-suppressor genes in breast cancer. *Genes Chromosomes Cancer* 2006; **45**: 527–535.

45 O'Flaherty E, Kaye J: TOX defines a conserved subfamily of HMG-box proteins. *BMC Genomics* 2003; **4**: 13.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)