Human Molecular Genetics, 2013, Vol. 22, No. 12 2539–2550
doi:10.1093/hmg/ddt089
Advance Access published on March 27, 2013

Common genetic determinants of breast-cancer risk in East Asian women: a collaborative study of 23 637 breast cancer cases and 25 579 controls

Wei Zheng[1,*], Ben Zhang[1], Qiuyin Cai[1], Hyuna Sung[3], Kyriaki Michailidou[7], Jiajun Shi[1], Ji-Yeob Choi[3,4], Jirong Long[1], Joe Dennis[7], Manjeet K. Humphreys[7], Qin Wang[7], Wei Lu[9], Yu-Tang Gao[10], Chun Li[2], Hui Cai[1], Sue K. Park[3,4,5], Keun-Young Yoo[5], Dong-Young Noh[4,6], Wonshik Han[4,6], Alison M. Dunning[8], Javier Benitez[11], Daniel Vincent[12], Francois Bacot[12], Daniel Tessier[12], Sung-Won Kim[13], Min Hyuk Lee[14], Jong Won Lee[15], Jong-Young Lee[16], Yong-Bing Xiang[10], Ying Zheng[9], Wenjin Wang[9], Bu-Tian Ji[17], Keitaro Matsuo[18], Hidemi Ito[18], Hiroji Iwata[19], Hideo Tanaka[18], Anna H. Wu[20], Chiu-chen Tseng[20], David Van Den Berg[20], Daniel O. Stram[20], Soo Hwang Teo[21,22], Cheng Har Yip[22], In Nee Kang[21], Tien Y. Wong[23], Chen-Yang Shen[28], Jyh-Cherng Yu[29], Chiun-Sheng Huang[30], Ming-Feng Hou[31], Mikael Hartman[24,25], Hui Miao[24], Soo Chin Lee[26,32], Thomas Choudary Putti[27], Kenneth Muir[33], Artitaya Lophatananon[33], Sarah Stewart-Brown[33], Pornthep Siriwanarangsan[34], Suleeporn Sangrajrang[35], Hongbing Shen[36], Kexin Chen[37], Pei-Ei Wu[38,39], Zefang Ren[40], Christopher A. Haiman[20], Aiko Sueta[18], Mi Kyung Kim[41], Ui Soon Khoo[42], Motoki Iwasaki[43], Paul D.P. Pharoah[7,8], Wanqing Wen[1], Per Hall[44], Xiao-Ou Shu[1], Douglas F. Easton[7,8], and Daehee Kang[3,4,5]

[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]Department of Biomedical Sciences, [4]Cancer Research Institute, [5]Department of Preventive Medicine and [6]Department of Surgery, Seoul National University College of Medicine, Seoul, Korea, [7]Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology and [8]Department of Oncology, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK, [9]Shanghai Center for Disease Control and Prevention, Shanghai, China, [10]Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China, [11]Human Genetics Group, Human Cancer Genetics Program, Spanish National Cancer Research Centre, Madrid, Spain, [12]McGill University and Génome Québec Innovation Centre, Montréal, Quebec, Canada, [13]Department of Surgery, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Korea, [14]Department of Surgery, Soon Chunh Yang University Hospital, Seoul, Korea, [15]Department of Surgery, University of Ulsan College of Medicine and Asan Medical Center, Seoul, Korea, [16]Center for Genome Science, Korea National Institute of Health, Osong Health Technology Administration Complex, Chungcheongbuk-do, Korea, [17]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA, [18]Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan, [19]Department of Breast Oncology, Aichi Cancer Center Central Hospital, Nagoya, Japan, [20]Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA, USA, [21]Cancer Research Initiatives Foundation, Sime Darby Medical Centre, Subang Jaya, Selangor, Malaysia, [22]Breast

*To whom correspondence should be addressed at: Vanderbilt Epidemiology Center and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, 2525 West End Avenue, 8th Floor, Nashville, TN 37203-1738, USA. Tel: +1 6159360682; Fax: +1 615936-8241; Email: wei.zheng@vanderbilt.edu

Cancer Research Unit, University Malaya Medical Centre, University Malaya Cancer Research Institute, Kuala Lumpur, Malaysia, [23]Singapore Eye Research Institute, [24]Saw Swee Hock School of Public Health, [25]Department of Surgery, Yong Loo Lin School of Medicine, [26]Cancer Science Institute of Singapore and [27]Department of Pathology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, [28]Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan, [29]Department of Surgery, Tri-Service General Hospital, Taipei, Taiwan, [30]Department of Surgery, National Taiwan University Hospital, Taipei, Taiwan, [31]Cancer Center and Department of Surgery, Kaohsiung Medical University Chung-Ho Memorial Hospital, Kaohsiung, Taiwan, [32]Department of Haematology-Oncology, National University Health System, Singapore, [33]Warwick Medical School, Warwick University, Coventry, UK, [34]Ministry of Public Health, Nonthaburi, Thailand, [35]National Cancer Institute, Bangkok, Thailand, [36]Department of Epidemiology and Biostatistics, Nanjing Medical University, Nanjing, China, [37]Department of Epidemiology and Biostatistics, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China, [38]Institute of Biomedical Sciences, Taiwan Biobank, Taipei, Taiwan, [39]College of Public Health, China Medical University, Taichung, Taiwan, [40]School of Public Health, Sun Yat-sen University, Guangzhou, China, [41]Cancer Epidemiology Division, Breast Cancer Center, National Cancer Center, Korea, [42]Department of Pathology, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong SAR, China, [43]Epidemiology and Prevention Division, Research Center for Cancer Prevention and Screening, National Cancer Center, Tokyo, Japan and [44]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

Received July 23, 2012; Revised November 30, 2012; Accepted February 18, 2013

In a consortium including 23 637 breast cancer patients and 25 579 controls of East Asian ancestry, we investigated 70 single-nucleotide polymorphisms (SNPs) in 67 independent breast cancer susceptibility loci recently identified by genome-wide association studies (GWASs) conducted primarily in European-ancestry populations. SNPs in 31 loci showed an association with breast cancer risk at $P < 0.05$ in a direction consistent with that reported previously. Twenty-one of them remained statistically significant after adjusting for multiple comparisons with the Bonferroni-corrected significance level of <0.0015. Eight of the 70 SNPs showed a significantly different association with breast cancer risk by estrogen receptor (ER) status at $P < 0.05$. With the exception of rs2046210 at 6q25.1, the seven other SNPs showed a stronger association with ER-positive than ER-negative cancer. This study replicated all five genetic risk variants initially identified in Asians and provided evidence for associations of breast cancer risk in the East Asian population with nearly half of the genetic risk variants initially reported in GWASs conducted in European descendants. Taken together, these common genetic risk variants explain ∼10% of excess familial risk of breast cancer in Asian populations.

INTRODUCTION

Breast cancer is one of the most common malignancies diagnosed among women worldwide, including East Asian women (1). Genetic factors play an important role in the etiology of both sporadic and familial breast cancer (2,3). Since 2007, common genetic risk variants in approximately 25 loci have been associated with breast cancer risk (3–18). However, with only a few exceptions (9,13,14,16,18), the vast majority of these risk variants were initially identified in studies conducted in European descendants (4–8,10–12,15,17). Several previous studies have evaluated some of these risk variants in relation to breast cancer risk in non-European populations (19–26), including East Asian women (19,24–26). Sample sizes in these studies, however, were small, and only a few variants were evaluated. Recently, as part of the international Collaborative Oncological Gene-Environment Study (COGS), 42 additional genetic susceptibility loci for breast cancer risk were identified in genome-wide association studies (GWASs) conducted in European descendants (27).

Given differences in genetic architecture, breast cancer incidence rates and environmental exposures across different ethnic populations, it is important to systematically investigate whether these genetic risk variants are associated with breast cancer risk in other ethnic populations. This investigation not only assesses the generalizability of initial GWAS findings but also provides valuable data to guide fine-mapping efforts in the search for disease variants. In this study, we combined data generated in the Asia Breast Cancer Consortium (ABCC) with Asian samples from COGS to systematically evaluate risk variants in all 67 loci identified to date.

RESULTS

This study combined data obtained from 41 586 women (19 963 cases and 21 623 controls) included in the ABCC and

Table 1. Summary of characteristics of study participants

Study abbreviation	Number of cases	Number of controls	Population ethnicity	Study design[a]	Mean age[b] (years)	Menopause[c] (%)	ER+[d] (%)	PR+[d] (%)
ABCC								
SBCGS1[e]	2918	2324	Chinese	Population-based	52/50	43/42	65	62
SBCGS2[e]	1613	1800	Chinese	Population-based	53/53	50/55	62	58
SBCGS3[e]	2601	2386	Chinese	Population-based	54/55	50/53	65	58
Taiwan	1066	1065	Chinese	Hospital-based	52/48	52/40	66	65
Hong Kong	491	642	Chinese	Hospital-based	46/46	50/42	71	60
Guangzhou	1083	1224	Chinese	Hospital-based	49/50	43/54	70	61
Tianjin	1546	1601	Chinese	Hospital-based	52/52	52/55	44	44
Nanjing	1786	1837	Chinese	Hospital-based	51/50	51/48	56	56
SeBCS1	2246	2052	Korean	Hospital-based	48/51	36/56	63	56
SeBCS2	777	1104	Korean	Hospital-based	48/48	36/37	63	53
KOHBRA/ KoGES	1397	3209	Korean	Hospital-based	40/50	23/NA	63	63
Korea-NCC	505	504	Korean	Hospital-based	49/49	50/45	65	58
Nagoya	644	644	Japanese	Hospital-based	51/51	49/49	73	100
Nagano	401	401	Japanese	Hospital-based	54/54	55/65	75	60
MEC	889	830	Japanese	Cohort	67/66	NA/NA	85	76
Subtotal	19 963	21 623						
COGS								
SBCGS	848	892	Chinese	Population-based	55/53	42/43	65	65
SeBCS	1162	1129	Korean	Hospital-based	49/52	41/NA	64	60
TWBCS	889	236	Chinese	Hospital-based	52/51	52/52	70	69
HERPACC	694	1376	Japanese	Hospital-based	52/51	49/49	74	63
ACP	423	636	Chinese/Thai	Hospital-based	46/47	42/41	63	NA
MyBrCa	770	610	Asians[f]	Hospital-based	49/49	53/61	60	52
SGBCC	533	502	Asians[f]	Hospital-based	52/50	NA	72	74
TBCS	138	253	Thai	Hospital-based	46/42	53/32	50	NA
LAABC	812	990	Chinese/Japanese	Population-based	54/51	60/46	74	69
Subtotal	6269	6624						
Total[g]	23 637	25 579						

ABCC, Asia Breast Cancer Consortium; COGS, International Collaborative Oncological Gene-Environment Study; NA, not available.
[a]Unless otherwise specified, case–control study design was used.
[b]Mean age of cases/controls with available data.
[c]Proportion of postmenopausal status of cases/controls with available data.
[d]Among cases with ER or PR data.
[e]Cases and controls from four studies conducted in Shanghai (see Materials and Methods).
[f]Including Chinese, Malays and Indians.
[g]Excluding overlapping samples included in both ABCC and COGS.

12 893 women (6269 cases and 6624 controls) included in COGS. All study participants were of Asian ancestry and recruited from studies conducted in Asian countries and the USA (Table 1 and Supplementary Material, Table S1). The ABCC consisted of two stages. Stage 1 included two GWASs, in which 5285 Chinese women and 4777 Korean women were scanned using primarily Affymetrix Genome-Wide Human SNP Array 6.0 [906 602 single-nucleotide polymorphisms (SNPs)]. After applying quality-control (QC) filters described previously (9,18,24), 5242 Chinese women (2918 cases and 2324 controls; 690 947 SNPs) and 4298 Korean women (2246 cases and 2052 controls; 555 525 SNPs) remained in the GWASs. These data were used to impute autosomal SNPs present in HapMap II release 22 using the MACH program v1.0 (28). Only SNPs with high imputation quality [R-squared (RSQ) > 0.50] were included in the analysis. To increase statistical power, we genotyped additional samples included in Stage 2 of the ABCC for 44 of the 70 SNPs identified in GWASs (called index SNPs in subsequent text). As part of a large collaboration of GWASs of multiple cancer sites (27), COGS samples were genotyped on a custom Illumina Infinium BeadChip, which included either index SNPs ($n = 67$) or SNPs ($n = 2$) in strong linkage disequilibrium (LD) ($r^2 > 0.8$), with the index SNPs selected for the data analysis of this study.

Evaluation of SNPs in 26 previously reported loci in Europeans and Asians

One index SNP per locus was selected for most of the previously reported loci, except for 10q21/*ZNF365* and 16q12/*TOX3*, for which two SNPs were selected per locus as they are not in LD in Asians. For 17 of the 28 SNPs, *de novo* genotyping was conducted for additional samples (ranging from 3348 to 27 166) included in the ABCC as part of previous studies (Stage 2) (9,13,14,18,19,23,26,29). A meta-analysis of ABCC Stages 1 and 2 data was performed under the fixed-effects model (30), and results are presented in Table 2 (left panel). Detailed stage-specific results are presented in Supplementary Material, Table S2. Table 2 also includes results obtained from Asian women included in COGS (middle panel), as well as combined results generated by a meta-analysis of data from the ABCC and COGS (right panel). Heterogeneity tests for associations of these SNPs

Table 2. Associations of breast cancer risk with SNPs located in 26 previously reported breast-cancer susceptibility loci: results from East Asian women

SNP[a]	Chr./gene[b]	Position[c] (bp)	Alleles[d]	EAF[e]	ABCC samples OR (95% CI)[f,g]	P_{trend}	COGS samples OR (95% CI)[f]	P_{trend}	Combined[h] Cases/controls	OR (95% CI)[f]	P_{meta}	$P^{i}_{heterogeneity}$
Loci initially identified in Asians												
rs9485372	6q25/TAB2	149 650 567	A/G	0.453	0.90 (0.87–0.92)	4.66×10^{-12}	0.89 (0.84–0.94)	7.64×10^{-6}	19 893/21 663	0.90 (0.87–0.92)	2.27×10^{-13}	0.599
rs2046210	6q25/ESR1	151 990 059	A/G	0.342	1.27 (1.23–1.31)	2.59×10^{-50}	1.28 (1.22–1.35)	2.81×10^{-20}	22 313/23 063	1.27 (1.24–1.31)	1.95×10^{-62}	0.702
rs10822013	10q21/ZNF365	63 921 983	T/C	0.469	1.09 (1.06–1.13)	7.98×10^{-9}	1.05 (1.00–1.10)	0.075	21 578/22 364	1.08 (1.05–1.11)	2.98×10^{-8}	0.133
rs7107217	11q24/BARX2	128 978 900	C/A	0.352	1.08 (1.05–1.12)	2.11×10^{-7}	1.09 (1.03–1.15)	0.002	21 545/23 560	1.08 (1.05–1.11)	6.69×10^{-8}	0.752
rs4784227	16q12/TOX3	51 156 689	T/C	0.247	1.27 (1.22–1.32)	4.58×10^{-31}	1.20 (1.12–1.27)	1.86×10^{-8}	18 459/16 622	1.24 (1.20–1.29)	2.83×10^{-32}	0.155
Loci initially identified in Europeans												
rs11249433	1p11/FCGR1B	120 982 136	G/A	0.030	1.13 (0.98–1.30)	0.101	1.18 (1.03–1.35)	0.018	11 937/11 667	1.16 (1.05–1.28)	0.005	0.743
rs13387042	2q35/TNP1	217 614 077	G/A	0.895	0.93 (0.86–1.01)	0.080	0.95 (0.88–1.03)	0.226	12 063/11 874	0.94 (0.88–0.99)	0.022	0.493
rs4973768	3p24/SLC4A7	27 391 017	T/C	0.192	1.12 (1.06–1.19)	8.73×10^{-5}	1.11 (1.05–1.18)	7.18×10^{-4}	13 504/12 111	1.11 (1.06–1.16)	4.93×10^{-6}	0.920
rs10069690	5p15/TERT	1 332 790	T/C	0.224	1.01 (0.92–1.11)	0.808	1.04 (0.98–1.11)	0.188	10 328/9870	1.05 (0.99–1.11)	0.103	0.865
rs10941679	5p12/MRPS30	44 742 255	G/A	0.484	1.09 (1.04–1.14)	6.74×10^{-4}	1.09 (1.04–1.15)	8.20×10^{-4}	12 066/11 875	1.08 (1.04–1.12)	2.24×10^{-5}	0.706
rs889312	5q11/MAP3K1	56 067 641	A/C	0.477	0.95 (0.90–1.00)	0.037	0.94 (0.89–0.99)	0.013	12 070/11 879	0.95 (0.92–0.99)	0.013	0.325
rs2180341	6q22/RNF146	127 642 323	A/G	0.740	1.01 (0.96–1.07)	0.607	NA	NA	7049/6364	1.01 (0.96–1.07)	0.607	NA
rs13281615	8q24/MYC	128 424 800	G/A	0.522	1.03 (0.98–1.08)	0.199	1.02 (0.97–1.08)	0.388	12 077/11 876	1.03 (0.99–1.07)	0.103	0.665
rs1011970	9p21/CDKN2A/2B	22 052 134	T/G	0.081	1.03 (0.93–1.14)	0.599	1.06 (0.97–1.16)	0.178	10 328/9870	1.06 (0.99–1.14)	0.112	0.893
rs865686	9q31/KLF4	109 928 299	T/G	0.934	0.99 (0.89–1.12)	0.927	1.06 (0.96–1.17)	0.213	10 328/9870	1.04 (0.96–1.12)	0.389	0.354
rs2380205	10p15/ANKRD16	5 926 740	T/C	0.106	0.92 (0.84–1.01)	0.093	1.00 (0.92–1.07)	0.927	10 328/9870	0.98 (0.92–1.04)	0.501	0.426
rs10995190	10q21/ZNF365	63 948 688	A/G	0.022	0.89 (0.73–1.10)	0.284	1.17 (1.00–1.38)	0.057	10 328/9870	1.06 (0.93–1.22)	0.361	0.041
rs704010	10q22/ZMIZ1	80 511 154	C/T	0.706	0.94 (0.88–1.00)	0.056	0.95 (0.90–1.00)	0.070	10 328/9870	0.95 (0.91–0.99)	0.013	0.772
rs1219648	10q26/FGFR2	123 336 180	G/A	0.367	1.16 (1.11–1.21)	1.63×10^{-12}	1.12 (1.07–1.18)	1.21×10^{-5}	15 130/14 584	1.14 (1.10–1.18)	2.90×10^{-13}	0.476
rs3817198	11p15/LSP1	1 865 582	C/T	0.128	1.03 (0.96–1.11)	0.346	1.07 (0.99–1.15)	0.080	13 655/12 083	1.07 (1.01–1.13)	0.015	0.938
rs614367	11q13/CCND1	69 037 945	T/C	NA	NA	NA	1.29 (1.06–1.58)	0.013	6269/6624	1.29 (1.06–1.58)	0.013	NA
rs10771399	12p11/PTHLH	28 046 347	G/A	0.191	0.88 (0.81–0.94)	5.30×10^{-4}	0.87 (0.81–0.93)	4.91×10^{-5}	10 328/9870	0.87 (0.83–0.92)	2.47×10^{-7}	0.920
rs1292011	12q24/MED13L	114 320 905	G/A	0.257	0.92 (0.86–0.99)	0.029	0.89 (0.84–0.95)	1.48×10^{-4}	10 328/9870	0.90 (0.86–0.94)	1.38×10^{-5}	0.668
rs999737	14q24/RAD51L1	68 104 435	T/C	0.001	1.96 (0.94–4.07)	0.072	0.95 (0.68–1.33)	0.764	10 772/10 711	1.08 (0.80–1.46)	0.628	0.075
rs3803662	16q12/TOX3	51 143 842	G/A	0.360	0.86 (0.82–0.91)	4.84×10^{-9}	0.85 (0.81–0.90)	3.97×10^{-9}	13 606/12 064	0.87 (0.84–0.90)	7.83×10^{-14}	0.444
rs6504950	17q22/STXBP4	50 411 470	A/G	0.072	1.01 (0.93–1.11)	0.757	0.98 (0.90–1.07)	0.685	13 642/12 111	0.98 (0.92–1.05)	0.642	0.926
rs8170	19p13/BABAM1	17 250 704	A/G	NA	NA	NA	1.10 (0.76–1.61)	0.605	6269/6624	1.10 (0.76–1.61)	0.605	NA
rs2823093	21q21/NRIP1	15 442 703	A/G	0.034	0.94 (0.80–1.11)	0.495	0.91 (0.80–1.03)	0.147	10 328/9870	0.93 (0.84–1.04)	0.200	0.486

Chr., chromosome; EAF, effect allele frequency; OR, odds ratio; CI, confidence interval; ABCC, Asia Breast Cancer Consortium; NA, not available; COGS, International Collaborative Oncological Gene-Environment Study.

[a]SNPs in high LD with the index SNPs were used in COGS samples: rs9485370 for rs9485372 ($r^2 = 1.0$) and rs17271951 for rs4784227 ($r^2 = 0.81$) based on Asian data from HapMap release27.

[b]The closest gene.

[c]Location based on NCBI Human Genome Build 36.3.

[d]Effect/reference alleles based on NCBI Human Genome Build 36.3, dbsnp b126 forward strand.

[e]Effect allele frequency in controls from the ABCC samples. EAFs from the COGS samples are similar to those presented in the table.

[f]Adjusted for age and study site if appropriate.

[g]Data described in studies reported in references (9,13,14,18,19,23,26,29) were included in this analysis.

[h]Combined results for all available studies after excluding samples that overlapped in both ABCC and COGS.

[i]*P*-value for heterogeneity between ABCC and COGS results derived using a Cochran's *Q*-test.

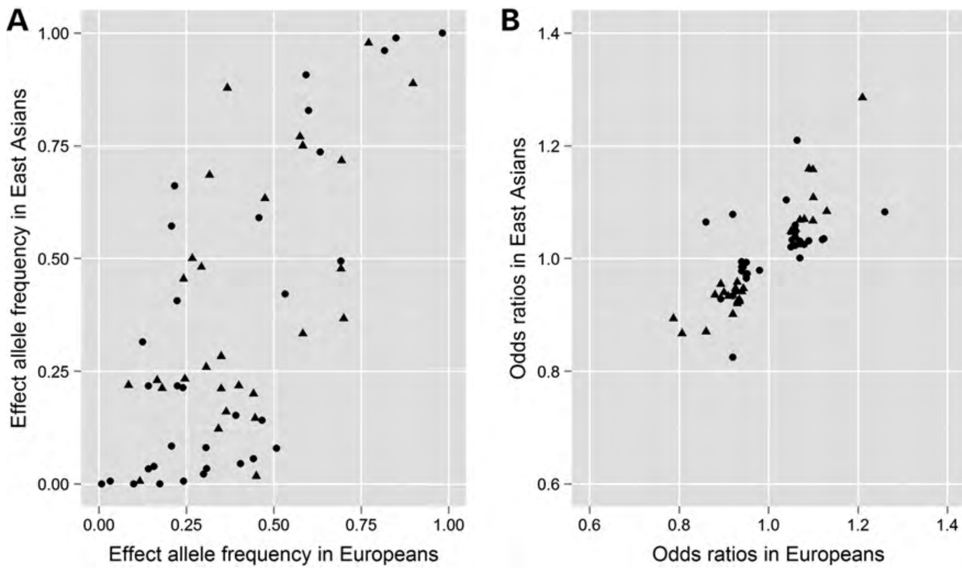

Figure 1. Correlation of effect allele frequency (**A**) and effect size measured in per-allele ORs (**B**) between East Asians and European-ancestry populations for 60 index SNPs identified initially in GWASs conducted in European descendants (filled triangle for SNPs replicated in Asians, filled circle for SNPs not replicated in Asians).

between the ABCC and COGS were statistically significant only for rs10995190/*ZNF365* ($P = 0.04$), for which the minor allele frequency (MAF) is only 0.02 in Asians. Of the 28 SNPs evaluated, 17 showed associations with breast cancer risk at $P < 0.05$ in Asian women in the same direction as previously reported, and 11 of them remained statistically significant after adjusting for multiple comparisons of 67 SNPs with the Bonferroni-corrected significance level of 0.0015 (0.10/67, one-sided test). However, the effect size—measured using per-allele odds ratios (ORs) in Asians—was smaller than that initially identified in Europeans (Fig. 1), and the difference was statistically ($P < 0.05$) or marginally ($P < 0.06$) significant for six of the replicated SNPs (Supplementary Material, Table S3). Five breast cancer-associated signals were previously discovered in Asian GWASs (9,13,14,18), and all of them were replicated in COGS samples with a combined P-value of 1.95×10^{-62} for rs2046210 (6q25.1/*ESR1*), 2.83×10^{-32} for rs4784227 (16q12/*TOX3*), 2.27×10^{-13} for rs9485372 (6q25/*TAB2*), 2.98×10^{-8} for rs10822013 (10q21/*ZNF365*) and 6.69×10^{-8} for rs7107217 (11q24/*BARX2*). In the 10q21/*ZNF365* locus, rs10995190, identified initially in a GWAS of European descendants, was not related to breast cancer risk in Asians, whereas the association with rs10822013, identified initially in the ABCC (14), was replicated in COGS samples with a borderline significant P-value. At 16q12.1/*TOX3*, associations with both rs4784227 and rs3803662 remained statistically significant after mutual adjustment (Table 3). Of the 11 SNPs not replicated in our study, 4 have a very low MAF ($\leq$0.05) in Asians, whereas the rest show very weak associations, with allelic ORs ranging from 0.93 to 1.06. The *CASP8* SNP (rs1045485) identified previously through a candidate gene study (4) has a very low MAF in Asians and was not replicated in the study (data not shown in the table). Additional analyses were performed to evaluate the association of the 17 replicated SNPs by major study populations included in this study, Chinese, Koreans and Japanese (Supplementary Material, Table S4). With the exception of rs889312 ($P = 0.002$) and rs3817198 ($P = 0.012$), the remaining 15 SNPs showed a consistent association in Chinese, Koreans and Japanese. SNPs rs889312 and rs3817198 were statistically associated with breast cancer risk only in Koreans and Chinese, respectively. Because of a small sample size, only three SNPs (rs2046210, rs1219648 and rs4784227) were found to be associated with breast cancer in Japanese at $P < 0.05$.

Evaluation of SNPs in 42 newly identified loci in Europeans

In addition to GWAS data, we genotyped 27 index SNPs in 7294 cases and 9404 controls included in the ABCC. Results for this analysis are presented in Table 4 and Supplementary Material, Table S5. Of the 42 SNPs evaluated (1 SNP per locus), 16 showed associations with breast cancer risk at $P < 0.05$ in the same direction reported in European-ancestry populations (27). Of them, 10 remained statistically significant after adjusting for multiple comparisons ($P < 0.0015$). Adjusted allelic ORs for the replicated SNPs ranged from 0.92 to 1.16, showing a smaller effect size than those presented in Table 2. SNP rs11820646 at 11q24 is correlated with rs7217217 ($r^2 = 0.386$ in Asian and $r^2 = 0.842$ in CEU) identified initially in the ABCC (18) and replicated in Asian samples included in COGS (Table 2). The association with rs7107217 remained after adjusting for rs11820646 ($P = 2.73 \times 10^{-4}$) (Table 3). On the other hand, adjusting for rs7107217 diminished the association with rs11820646, suggesting that rs7107217 may tag the disease variant better than rs11820646 in East Asians. Of the 22 non-replicated SNPs, 3 have a small MAF in both Asia and European samples (based on the HapMap data), including rs132390 (~500 kb downstream to the *CHEK2* gene), rs11571833 (*BRCA2*) and rs11814448 (10p12). The MAF for the other

Table 3. Conditional analysis for the association of breast cancer risk with SNPs in four loci for which more than one risk variant was reported: results from East Asian women

Locus	Cases/controls	SNPs[a]	EAF[b]	LD (r^2)[c] (Asian/CEU)	Per-allele association OR (95% CI)[d]	P_{trend}
10q21/ZNF365	5164/4376	rs10822013[e]	0.467	0.026/0.157	1.09 (1.04–1.15)	2.92×10^{-4}
		rs10995190	0.022		0.88 (0.72–1.08)	0.237
10q26/FGFR2	9173/6588	rs1219648	0.364	0.639/1.000	1.12 (1.02–1.23)	0.022
		rs2981582	0.706		0.97 (0.88–1.07)	0.527
11q24/BARX2	12 254/13 586	rs7107217[e]	0.357	0.386/0.842	1.10 (1.04–1.15)	2.73×10^{-4}
		rs11820646	0.555		1.00 (0.95–1.05)	0.909
16q12/TOX3	9024/6139	rs3803662	0.362	0.139/0.813	0.92 (0.88–0.98)	0.005
		rs4784227[e]	0.251		1.21 (1.14–1.28)	3.87×10^{-10}

[a]SNPs included in the same logistic regression model.
[b]Effect allele frequency in controls.
[c]Based on Asian data from HapMap release 22.
[d]Adjusted for age and study site if appropriate.
[e]SNPs identified initially in GWASs conducted in East Asians (13,14,18).

six SNPs (rs1353747, rs720475, rs7072776, rs7904519, rs2588809 and rs13329835) also was very low (<0.05) in Asian samples but not in European samples (mean MAF = 0.23) included in the HapMap project. With the exception of three SNPs (rs11242675, rs17817449 and rs3760982), heterogeneity tests were not statistically significant for SNPs between ABCC and COGS samples. Again, the effect size for most of the SNPs in Asians was smaller than those identified in Europeans (Fig. 1), and the difference was statistically significant for four SNPs (Supplementary Material, Table S3). We also performed subgroup analysis for the 16 replicated SNPs by Chinese, Korean and Japanese (Supplementary Material, Table S4). Only four SNPs (rs11242675, rs11780156, rs11199914 and rs4808801) showed some evidence of heterogeneity ($P < 0.05$) in their associations with breast cancer risk in the three populations. In Japanese, however, only two SNPs showed a statistically significant association (rs7697216 and rs4808801), likely due to a small sample size in this group.

Associations by ER status

Of the 70 SNPs in 67 loci evaluated, SNPs in 8 loci showed a significantly different association with breast cancer risk by ER status at $P < 0.05$ (Table 5 and Supplementary Material, Table S6). With the exception of rs2046210/6q25, the seven other SNPs showed a stronger association with ER-positive cancer compared with ER-negative cancer. In fact, six of the eight SNPs showed no significant association with ER-negative cancer except rs2046210/6q25 (per allele OR = 1.36, $P = 6.04 \times 10^{-49}$) and rs1219648/*FGFR2* (per allele OR = 1.08, $P = 0.0055$). Details of all evaluated SNPs by ER status are presented in Supplementary Material, Table S6.

DISCUSSION

By analyzing data from up to 23 637 breast cancer cases and 25 579 controls of Asian ancestry, we identified a significant association at $P < 0.05$ for 31 of the 67 independent breast cancer association signals reported from previous GWASs conducted mostly in European descendants. The number of SNPs identified in our study with a significant association at $P < 0.05$ was substantially greater than the 1.68 significant associations expected by chance under a null hypothesis ($67 \times 0.025 = 1.675$). Twenty-one of these associations remained statistically significant after adjusting for multiple comparisons using the Bonferroni-corrected *P*-value of 0.0015. Our study has substantially boosted the number of genetic risk variants identified to date for East Asian women. Together, these SNPs explain ~10% of the excess familial risk for breast cancer among East Asian women (Supplementary Material, Table S7).

This study represents the largest, most comprehensive effort made to date to evaluate the generalizability of GWAS findings to any non-European population. With 23 637 cases and 25 579 controls included in the study, the statistical power was very large. For some analyses, however, only about 10 200 cases and 9800 controls were included. Even in this sample size set, we still have 85% power to detect associations with ORs as low as 1.10 for an MAF of 0.10 or above (Supplementary Material, Table S8). This study was designed to directly evaluate risk variants identified in previous GWASs, and thus, SNPs showing an association at $P < 0.05$ in the same direction as previously reported were considered statistically significant. It is possible, however, that some of the significant findings ($n = 1.68$) could be due to chance because of multiple comparisons. Nevertheless, we also presented findings based on a more stringent significant level with the adjustment for multiple comparisons.

Differences in LD patterns between East Asian- and European-ancestry populations are likely to be the major reason for the lack of replication for some of the index SNPs in our study. The vast majority of non-replicated SNPs showed a substantial difference in the frequency of the effect alleles (Fig. 1 and Supplementary Material, Table S9). Fourteen SNPs which did not replicate had very small MAFs in East Asian women (mean = 0.0176) but much higher MAFs in European descendants (mean = 0.177) ($P < 0.001$) (Supplementary Material, Table S9). Most SNPs with an association identified by GWASs are tagging SNPs of the disease variant(s). Because of the differences in LD patterns between East Asian and European

Table 4. Associations of breast cancer risk with SNPs located in 42 recently reported breast-cancer susceptibility loci: results from East Asian women

SNP[a]	Chr./gene[b]	Position[c] (bp)	Alleles[d]	EAF[e]	ABCC samples OR (95% CI)[f]	P_{trend}	COGS samples OR (95% CI)[f]	P_{trend}	Combined[g] Cases/controls	OR (95% CI)[f]	P_{meta}	$P^{h}_{heterogeneity}$
rs616488	1p36/PEX14	10 488 802	G/A	0.314	0.93 (0.89–0.96)	1.01×10^{-4}	0.95 (0.90–1.01)	0.096	17 477/19 133	0.94 (0.91–0.97)	3.41×10^{-4}	0.517
rs11552449	1p13/AP4B1	114 249 912	T/C	0.608	1.03 (0.99–1.07)	0.179	1.02 (0.97–1.07)	0.523	17 447/19 114	1.03 (0.99–1.06)	0.131	0.679
rs4849887	2q14/INHBB	120 961 592	C/T	0.817	1.08 (1.03–1.13)	8.35×10^{-4}	1.04 (0.97–1.11)	0.245	17 471/19 143	1.07 (1.03–1.11)	0.001	0.282
rs2016394	2q31/DLX2	172 681 217	A/G	0.186	0.98 (0.90–1.08)	0.733	1.00 (0.94–1.07)	0.974	10 328/9870	0.99 (0.94–1.05)	0.790	0.616
rs1550623	2q31/CDCA7	173 921 140	A/G	0.992	1.47 (0.89–2.42)	0.131	1.15 (0.93–1.40)	0.190	10 328/9870	1.21 (1.00–1.47)	0.054	0.097
rs16857609	2q35/DIRC3	218 004 753	T/C	0.589	1.07 (1.03–1.10)	5.09×10^{-4}	1.06 (1.01–1.12)	0.018	17 452/19 120	1.07 (1.04––1.10)	2.31×10^{-5}	0.816
rs6762644	3p26/ITPR1	4 717 276	G/A	0.079	1.02 (0.96–1.09)	0.483	1.07 (0.98–1.17)	0.151	17 481/19 143	1.03 (0.98–1.09)	0.286	0.324
rs12493607	3p24/TGFBR2	30 657 943	C/G	0.673	1.07 (1.01–1.14)	0.024	1.04 (0.99–1.10)	0.133	10 328/9870	1.05 (1.01–1.10)	0.019	0.547
rs9790517	4q24/TET2	106 304 227	T/C	0.611	1.04 (0.98–1.10)	0.191	1.00 (0.95–1.06)	0.882	10 328/9870	1.02 (0.98–1.06)	0.342	0.321
rs7697216	4q34/ADAM29	176 064 611	C/T	0.758	1.07 (1.03–1.12)	9.99×10^{-4}	1.08 (1.02–1.14)	0.014	17 437/19 127	1.07 (1.04–1.11)	1.16×10^{-4}	0.889
rs10472076	5q11/RAB3C	58 219 818	C/T	0.270	1.02 (0.98–1.07)	0.242	0.99 (0.94–1.05)	0.745	17 476/19 184	1.02 (0.98–1.05)	0.375	0.279
rs1353747	5q11/PDE4D	58 373 238	G/T	NA	NA	NA	0.82 (0.57–1.20)	0.316	6269/6624	0.82 (0.57–1.20)	0.316	NA
rs1432679	5q33/EBF1	158 176 661	T/C	0.336	0.92 (0.88–0.96)	2.07×10^{-5}	0.92 (0.87–0.97)	0.001	17 107/19 040	0.92 (0.89–0.96)	2.79×10^{-6}	0.730
rs11242675	6p25/FOXQ1	1 263 878	T/C	0.480	1.01 (0.97–1.04)	0.751	0.93 (0.89–0.98)	0.008	17 425/19 030	0.99 (0.96–1.03)	0.703	0.003
rs204247	6p23/RANBP9	13 830 502	A/G	0.384	0.97 (0.94–1.01)	0.164	0.96 (0.91–1.01)	0.081	17 444/19 128	0.97 (0.94–1.00)	0.085	0.383
rs17529111	6q14/FAM46A	82 185 105	C/T	0.198	1.07 (0.99–1.15)	0.080	1.03 (0.97–1.10)	0.326	10 328/9870	1.04 (0.99–1.09)	0.152	0.793
rs720475	7q35/ARHGEF5	143 705 862	A/G	0.026	1.00 (0.89–1.12)	0.982	0.96 (0.84–1.10)	0.579	17 494/19 157	0.98 (0.89–1.07)	0.619	0.765
rs9693444	8p21/RPL17P33	29 565 535	C/A	0.706	0.92 (0.88–0.95)	1.83×10^{-5}	0.93 (0.88–0.98)	0.008	17 448/19 112	0.93 (0.90–0.96)	1.01×10^{-5}	0.970
rs6472903	8q21/HNF4G	76 392 856	T/G	0.957	1.16 (1.05–1.28)	0.003	1.15 (1.01–1.30)	0.040	17 480/19 190	1.16 (1.07–1.26)	4.23×10^{-4}	0.839
rs2943559	8q21/HNF4G	76 580 492	G/A	0.107	1.03 (0.93–1.13)	0.600	0.94 (0.86–1.02)	0.154	10 328/9870	0.96 (0.90–1.03)	0.299	0.325
rs11780156	8q24/MYC	129 263 823	T/C	0.216	1.01 (0.97–1.05)	0.707	0.99 (0.93–1.06)	0.834	17 472/19 143	1.00 (0.96–1.04)	0.973	0.775
rs10759243	9q31/KLF4	109 345 936	C/A	0.579	0.95 (0.90–1.01)	0.102	0.94 (0.89–0.99)	0.017	10 328/9870	0.95 (0.91–0.99)	0.008	0.653
rs7072776	10p12/MLLT10	22 072 948	G/A	0.972	0.88 (0.73–1.04)	0.136	1.01 (0.90–1.13)	0.876	10 328/9870	0.96 (0.87–1.06)	0.424	0.095
rs11814448	10p12/DNAJC1	22 355 849	C/A	0.005	1.08 (0.84–1.39)	0.548	1.08 (0.85–1.38)	0.527	17 291/19 079	1.08 (0.90–1.30)	0.389	1.000
rs7904519	10q25/TCF7L2	114 763 917	G/A	0.042	1.01 (0.88–1.17)	0.866	1.06 (0.94–1.19)	0.347	10 328/9870	1.02 (0.93–1.12)	0.638	0.342
rs11199914	10q26/FGFR2	123 083 891	T/C	0.396	0.96 (0.91–1.02)	0.193	0.98 (0.93–1.03)	0.441	10 328/9870	0.97 (0.93–1.01)	0.148	0.551
rs12575663	11q13/OVOL1	65 331 111	A/G	0.175	1.00 (0.95–1.04)	0.880	0.94 (0.88–1.00)	0.040	17 462/19 133	0.98 (0.94–1.01)	0.211	0.104
rs11820646	11q24/BARX2	128 966 381	C/T	0.555	1.06 (1.02–1.10)	0.002	1.04 (0.99–1.10)	0.129	17 479/19 136	1.05 (1.01–1.08)	0.004	0.790
rs12422552	12p13/ATF7IP	14 305 198	C/G	0.275	1.06 (0.99–1.14)	0.099	1.06 (1.00–1.12)	0.035	10 328/9870	1.05 (1.00–1.10)	0.043	0.431
rs17356907	12q22/NTN4	94 551 890	G/A	0.227	0.94 (0.90–0.98)	0.004	0.92 (0.87–0.97)	0.005	17 460/19 120	0.93 (0.90–0.97)	1.99×10^{-4}	0.513
rs11571833	13q13/BRCA2	31 870 626	T/A	NA	NA	NA	2.21 (0.70–6.97)	0.177	6269/6624	2.21 (0.70–6.97)	0.177	NA
rs2236007	14q13/PAX9	36 202 520	A/G	0.246	0.91 (0.87–0.95)	1.93×10^{-5}	0.93 (0.88–0.99)	0.014	17 499/19 186	0.92 (0.89–0.95)	5.61×10^{-6}	0.609
rs2588809	14q24/RAD51L1	67 730 181	C/T	0.978	0.92 (0.81–1.03)	0.157	0.97 (0.84–1.11)	0.646	17 490/19 146	0.94 (0.86–1.04)	0.225	0.618
rs941764	14q32/CCDC88C	90 910 822	G/A	0.138	1.06 (1.01–1.11)	0.029	1.02 (0.95–1.10)	0.507	17 467/19 134	1.05 (1.01–1.10)	0.025	0.385
rs17817449	16q12/FTO	52 370 868	G/T	0.124	0.88 (0.83–0.93)	6.79×10^{-6}	1.01 (0.94–1.08)	0.868	16 797/18 983	0.92 (0.88–0.97)	5.78×10^{-4}	0.001
rs13329835	16q23/DYL2	79 208 306	G/A	0.051	1.05 (0.97–1.13)	0.260	1.04 (0.93–1.17)	0.488	17 482/19 192	1.02 (0.96–1.10)	0.483	0.735
rs527616	18q11/AQP4	22 591 422	G/C	0.692	1.04 (1.00–1.08)	0.067	1.01 (0.95–1.07)	0.818	17 466/19 127	1.03 (1.00–1.07)	0.065	0.269
rs1436904	18q11/CHST9	22 824 665	G/T	0.453	0.99 (0.93–1.05)	0.692	1.00 (0.95–1.05)	0.931	10 328/9870	0.98 (0.95–1.02)	0.440	0.391
rs4808801	19p13/ELL	18 432 141	G/A	0.252	0.95 (0.91–0.99)	0.016	0.97 (0.91–1.03)	0.300	17 316/19 098	0.96 (0.92–0.99)	0.018	0.624
rs3760982	19q13/KCNN4	48 978 353	G/A	0.865	0.96 (0.91–1.01)	0.091	1.05 (0.98–1.12)	0.148	17 570/19 207	0.98 (0.94–1.03)	0.464	0.012
rs132390	22q12/EMID1	27 951 477	T/C	NA	NA	NA	0.93 (0.54–1.59)	0.787	6269/6624	0.93 (0.54–1.59)	0.787	NA
rs6001930	22q13/MKL1	39 206 180	C/T	0.278	1.03 (0.99–1.07)	0.192	1.05 (0.99–1.12)	0.092	17 450/19 129	1.03 (1.00–1.07)	0.063	0.463

Chr., chromosome; EAF, effect allele frequency; OR, odds ratio; CI, confidence interval; ABCC, Asia Breast Cancer Consortium; NA, not available; COGS, International Collaborative Oncological Gene-Environment Study.

[a]SNPs in complete LD ($r^2 = 1.0$, based on Asian data from HapMap release27) with index SNPs were used in this study: rs7697216 for rs6828523, and rs12575663 for rs3903072.

[b]The closest gene.

[c]Location based on NCBI Human Genome Build 36.3.

[d]Effect/reference alleles based on NCBI Human Genome Build 36.3, dbsnp b126 forward strand.

[e]Effect allele frequency in controls from the ABCC samples. EAFs from the COGS samples are similar to those presented in the table.

[f]Adjusted for age and study site if appropriate.

[g]Combined results for all available studies after excluding samples that overlapped in both ABCC and COGS.

[h]*P*-value for heterogeneity between ABCC and COGS results derived using a Cochran's *Q*-test.

Table 5. Breast cancer-associated SNPs at 67 susceptibility loci showing a different association by ER status: results from East Asian women

SNP	Chr./gene[a]	Position[b] (bp)	Alleles[c]	ER positive OR (95% CI)[d]	P_{meta}	ER negative OR (95% CI)[d]	P_{meta}	$P^{e}_{heterogeneity}$
rs10941679	5p12/MRPS30	44 742 255	G/A	1.11 (1.06–1.16)	5.44×10^{-6}	1.01 (0.95–1.07)	0.787	0.009
rs2046210	6q25/ESR1	151 990 059	A/G	1.23 (1.19–1.27)	4.60×10^{-32}	1.36 (1.31–1.42)	6.04×10^{-49}	1.10×10^{-4}
rs1219648	10q26/FGFR2	123 336 180	G/A	1.20 (1.15–1.25)	2.55×10^{-17}	1.08 (1.02–1.14)	0.0055	0.002
rs204247	6p23/RANBP9	13 830 502	A/G	0.96 (0.92–0.99)	0.021	1.03 (0.98–1.08)	0.235	0.018
rs11199914	10q26/FGFR2	123 083 891	T/C	0.94 (0.89–0.99)	0.011	1.03 (0.96–1.09)	0.399	0.026
rs12422552	12p13/ATF7IP	14 305 198	C/G	1.09 (1.03–1.15)	0.003	0.98 (0.92–1.06)	0.647	0.030
rs527616	18q11/AQP4	22 591 422	G/C	1.06 (1.02–1.10)	0.006	0.97 (0.92–1.02)	0.276	0.010
rs6001930	22q13/MKL1	39 206 180	C/T	1.07 (1.02–1.11)	0.002	0.98 (0.93–1.03)	0.398	0.011

Chr., chromosome; EAF, effect allele frequency; OR, odds ratio; CI, confidence interval.
[a]The closest gene.
[b]Location based on NCBI Human Genome Build 36.3.
[c]Effect/reference alleles based on NCBI Human Genome Build 36.3, dbsnp b126 forward strand.
[d]Odds ratios based on meta-analysis of ABCC and COGS using a fixed-effect model.
[e]*P*-value for heterogeneity between ER-positive and -negative cases was calculated using a Cochran's *Q*-test.

populations, some of the SNPs identified by GWASs may be more closely associated with disease variants in European-ancestry populations compared with East Asian populations, which may partially explain the weaker association for some index SNPs in East Asians than European-ancestry populations. Thus, fine-mapping of these regions in both Asian and European populations may be fruitful in identifying relevant disease variants.

Our findings for SNPs located at 10q21/*ZNF365* illustrate how differences in LD patterns may result in a different SNP being identified by a GWAS. One SNP at 10q21/*ZNF365* (rs10822013, OR = 1.10, $P = 5.87 \times 10^{-9}$) was reported previously to be associated with breast cancer risk in a GWAS conducted in East Asians (14). This SNP is ~26.7 kb upstream of rs10995190, which was identified independently in a GWAS conducted in European descendants (12). Interestingly, rs10995190, a common SNP in European descendants (MAF = 0.13), is rare in East Asians (MAF = 0.02). It is likely that the disease variants are tagged by different SNPs in East Asians (rs10822013) and Europeans (rs10995190) in this region. SNPs at 16q12.1/*TOX3* also are of interest. The two top breast cancer-associated SNPs in this region identified to date are rs4784227 and rs3803662. These two SNPs, however, are highly correlated in Europeans ($r^2 = 0.813$) but virtually not correlated in Asians ($r^2 = 0.139$). Although rs4784227 showed a stronger association with breast cancer risk than rs3803662 in Asians, the association with rs3803662 remained statistically significant after adjustment for rs4784227, suggesting the existence of a second disease variant in this region (18). *In vitro* experiments have provided evidence for functional significance of rs4784227 (30). The differences in LD patterns across ethnic groups may help to narrow the region of interest and/or the number of candidate SNPs for fine-mapping analyses to identify disease variants.

The lack of association with index SNPs in some loci could also be explained by allelic heterogeneity, in which different underlying disease variants exist in Asian- and European-ancestry populations, or by possible differences between these populations in genetic and environmental modifiers. In addition to differences in LD patterns between East Asian and European populations, and possible allelic heterogeneity in these populations, other factors may have contributed to weaker associations observed in Asians than in European-ancestry populations for some of the index SNPs. Some SNPs were imputed in the ABCC data set, and thus imputation accuracy may affect the risk estimate. However, we included in the analysis only SNPs that were imputed with high quality (RSQ > 0.5, mean RSQ = 0.92). Furthermore, all SNPs were directly genotyped in the COGS data set, which should facilitate direct comparison of the results obtained in the Asian and European-ancestry samples included in the COGS project. Additional studies, including fine-mapping and functional characterization of SNPs, are needed to clarify the reasons for the different associations observed between Asian- and European-ancestry populations.

In conclusion, our study replicated all five breast cancer risk variants identified previously in GWASs conducted in East Asians and found that nearly half of the variants identified initially in European-ancestry GWASs can be directly replicated in East Asians. These results show the complexity of uniformly applying GWAS findings across ancestral groups. Common genetic variants identified to date explain ~14% of familial relative risk (FRR) of breast cancer in European-ancestry populations (27) but only 10% in East Asians. The lower estimated FRR in Asian- than in European-ancestry populations is expected since most known common genetic variants for breast cancer were identified in GWASs conducted in European-ancestry populations, and most of these risk variants show a stronger association with breast cancer risk in European- than in Asian-ancestry populations. It is possible that other risk variants may exist in some of the loci that could show a stronger association with breast cancer risk in Asians than the variants analyzed in this study. It is also possible that multiple independent risk variants may exist in some of the loci we evaluated. Fine-mapping of known breast cancer susceptibility loci may identify breast cancer risk variants more relevant in Asians than those identified initially in GWASs of European-ancestry populations, which could improve risk assessment in Asians.

MATERIALS AND METHODS

All methodology, results and interpretation in this study were reported according to the STREGA guidelines (31) and STROBE statement (32). Approval was granted from the relevant institutional review boards at all study sites, and all included participants gave informed consent.

Study populations

This study is a collaborative effort between the ABCC and the international COGS. The ABCC included 19 963 cases and 21 623 controls from 16 studies (Table 1). Detailed descriptions of these participating studies and demographic characteristics of study participants have been published previously (9,13,14,18). The consortium included 25 983 Chinese women, 11 794 Korean women and 3809 Japanese women. Chinese participants came from nine studies: Shanghai [$n = 13\,642$, the Shanghai Breast Cancer Study (SBCS) and Shanghai Breast Cancer Survival Study (SBCSS), Shanghai Endometrial Cancer Study (SECS, controls only), Shanghai Women's Health Study (SWHS)] (9,33,34), Nanjing ($n = 3623$) (35), Tianjin ($n = 3147$) (36), Taiwan ($n = 2131$) (37), Guangzhou ($n = 2307$) (18,23) and Hong Kong ($n = 1133$) (38). Korean participants came from four studies [the Seoul Breast Cancer Study (SeBCS) ($n = 6179$) (24), Korea National Cancer Center ($n = 1009$), Korea Genome Epidemiology Study (KoGES; $n = 3209$) (39) and Korean Hereditary Breast Cancer (KOHBRA; $n = 1397$) (40)]. Japanese participants came from three studies conducted in Hawaii and Los Angeles [$n = 1719$; the Multiethnic Cohort Study (MEC) (41)], Nagoya ($n = 1288$) (42) and Nagano ($n = 802$) (43) (Table 1). COGS includes 6269 breast cancer patients and 6624 controls of Asian ancestry who were recruited by nine studies. Approximately 2600 cases and 2650 controls were included in both the ABCC and COGS. After excluding overlapping samples, 23 637 cases and 25 580 controls remained for this study.

Genotyping methods

The ABCC consisted of two GWASs, in which 5285 Chinese women and 4777 Korean women were scanned primarily using Affymetrix Genome-Wide Human SNP Array 6.0 (Stage 1). Genotyping protocols for Stage I have been described elsewhere (9,24). In the Chinese GWASs, the initial 300 samples were genotyped using the Affymetrix GeneChip Mapping 500K Array Set. The remaining 4985 samples were genotyped using Affymetrix Genome-Wide Human SNP Array 6.0. We included one negative control and at least three positive QC samples from the Coriell Cell Repositories (http://ccr.coriell.org/) in each of the 96-well plates for Affymetrix SNP Array 6.0 genotyping. A total of 273 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. Genetically identical and unexpected duplicate samples were excluded, as were close relatives with a pair-wise proportion of identify-by-descent estimate >0.25. All samples with a call rate <95% were excluded. SNPs were excluded if: (i) MAF <1%; (ii) call rate <95%; or (iii) genotyping concordance rate <95% in QC samples. The final data set included 2918 cases and 2324 controls for 690 947 markers. For the Korean GWASs, Affymetrix Genome-Wide Human SNP Array 6.0 was also used (24). A total of 30 QC samples were successfully genotyped, and the concordance rate was 99.8%. The sex of all samples was confirmed to be female. SNPs were excluded if: (i) genotype call rate <95%; (ii) MAF <1% in either cases or controls; (iii) evidence for deviation from Hardy–Weinberg equilibrium (HWE) at P-value $< 10^{-6}$; or (iv) poor cluster plot in either cases or controls. After these QC exclusions, the final data set included 2246 cases and 2052 controls for 555 525 markers.

Genotyping for Stage 2 in the ABCC was completed primarily using a custom Illumina Infinium BeadChip and the iPLEX Sequenom MassArray platform. To compare consistency between Stage-1 (Affymetrix) and Stage-2 genotyping, we included 43 and 45 Stage-1 samples in the assay using the Illumina BeadChip and Sequenom platforms, respectively, which yielded concordance rates of 99.9 and 99.5%, respectively, compared with results obtained from the Affymetrix 6.0 genotyping. Additional QC samples were used in the Sequenom assay, including one negative control (water), two blinded duplicates and two samples from the HapMap project in each 96-well plate. The mean concordance rate was 99.7% for the blind duplicates and 98.9% for HapMap samples. Some samples were genotyped using TaqMan assays, for which assay protocols were developed and validated at the Vanderbilt Molecular Epidemiology Laboratory, and assay reagents were provided to investigators who performed the assays. For the MEC study, SNP data needed for the study were extracted from the data generated using Illumina Human 660W.

Genotyping in COGS was conducted using a custom Illumina Infinium BeadChip, which included 211 155 SNPs, as part of a large collaboration for replication of promising associations selected from GWASs of multiple cancers (27). Individuals were excluded for any of the following reasons: genotypically not female XX; overall call rate <95%; low or high heterozygosity ($P < 10^{-6}$); individuals not concordant with previous genotyping with BCAC; individuals where genotypes for the duplicate sample appeared to be from a different individual; or 'cryptic' duplicates where phenotypic data indicated that the individuals were different. For known and cryptic duplicates, the sample with the lower call rate was excluded. We attempted to identify first-degree relative pairs using identity-by-state estimates based on approximately 37 000 uncorrelated SNPs. For apparent first-degree relative pairs, we removed the control from a case–control pair, otherwise we removed the individual with the lower call rate. Ethnic outliers were identified by multidimensional scaling, combining COGS data with the three HapMap2 populations, based on a subset of 37 000 uncorrelated markers which passed QC (including approximately 1000 selected as ancestry informative markers). Individuals with >15% minority ancestry, based on the first two components, were excluded. Although the vast majority of study participants were of East-Asian origin, women from other Asian regions also were included in studies from Singapore (SGBCC) (92 Indians and 180 Malays) and Malaysia (MyBrCa) (152 Indians and 166 Malays). Exclusion of these subjects should not change the results given the small sample size. Therefore, for these studies, no exclusions for ethnic outliers were made, but principal components analysis adequately corrected for inflation. Principal components analyses were carried out based on a subset of 37 000 uncorrelated SNPs, and two principal components were used for the studies in Asian populations. We excluded SNPs which had a call rate <95%, deviated from HWE in controls at $P < 10^{-7}$, or had genotype discrepancies in >2% of duplicate samples, across all COGS consortia. Final analyses were based on 199 961 SNPs.

Statistical analyses

PLINK version 1.07 (http://pngu.mgh.harvard.edu/~purcell/plink/) was used to analyze genome-wide data obtained in Stage 1 (44). A set of 4305 SNPs with MAF $\geq$ 35% and a distance $\geq$100 kb between two adjacent SNPs were selected to evaluate the population structure in the Chinese GWASs. Inflation factor λ was estimated to be 1.04. Similar analyses were performed for the Korean GWASs, which yielded a λ of 1.04 (24). Inflation factor λ is <1.03 in all eight studies included in COGS. These data suggest that any population substructure, if present, should not have any appreciable effect on the results. ORs associated with each SNP and 95% CIs were estimated using logistic regression implemented in Plink with adjustment for age.

We used the program MACH 1.0 (28) to impute genotypes for autosomal SNPs which were present in HapMap Phase II release 22 for samples included in the Chinese and Korean GWASs. Only SNPs with imputation quality score RSQ > 0.50 were included in analyses. Dosage data for imputed SNPs for samples in each Stage-1 study were analyzed using the program mach2dat (28). Associations between genotype dosage and breast cancer risk were assessed assuming a log-additive model. ORs associated with each SNP and 95% confidence intervals (CIs) were estimated using logistic regression adjusted for age.

Individual data were obtained from each study included in the ABCC and COGS for two separate pooled analyses. Results from these two pooled analyses were combined in meta-analyses performed under a fixed-effects model using the METAL program (45). Data quality for each study was evaluated by examining genotyping cluster plots and HWE in controls for all SNPs. SNPs which failed to meet data QC criteria were excluded from the analysis. Case–control differences for 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 the effect allele based on a log-additive model and adjusted for age and study site, when appropriate. Stratified analyses by ER status were performed.

The fraction of the FRR explained by a single locus, under a multiplicative model, can be expressed as $\ln(\lambda)/\ln(\lambda_o)$, where λ is the FRR to offspring of an affected individual due to the locus, and λ_o is the overall FRR. λ_o is assumed to be 1.8 for breast cancer (46). Note that if an individual locus fits a log-additive model, the formula of λ for a single locus is:

$$\lambda = \frac{(pr^2 + q)}{(pr + q)^2},$$

where p is the frequency of the risk allele, $q = 1 - p$ is the frequency of the reference allele and r is the per-allele relative risk. We assumed that the risks associated with each locus combine multiplicatively, and the FRRs also multiply, so that the combined contribution is using formula:

$$\ln\,(\Pi_i \lambda_i)/\ln(\lambda_O).$$

URLS

PLINK version 1.07, http://pngu.mgh.harvard.edu/~purcell/plink/
MACH 1.0, http://www.sph.umich.edu/csg/abecasis/MACH/
mach2dat, http://www.sph.umich.edu/csg/abecasis/MACH/
R version 2.13.0, http://www.r-project.org/
METAL, http://www.sph.umich.edu/csg/abecasis/Metal/
SNAP, http://www.broadinstitute.org/mpg/snap/
HapMap project, http://hapmap.ncbi.nlm.nih.gov/
BCAC, http://www.srl.cam.ac.uk/consortia/bcac/
COGS, http://cogseu.org/

AUTHORS' CONTRIBUTIONS

W.Z. conceived and directed the study and wrote the manuscript. B.Z. coordinated the study, performed data analysis and contributed to manuscript preparation. Q.C. directed the lab operation. J.S. performed the genotyping experiments. Q.C., J.L., C.L., W.W. and X.-O.S. contributed to study coordination, data analysis and/or manuscript revision. D.K. directed the Korean studies and is the corresponding author for these studies. H.S. and J.-Y.C. are key members of the Korean studies. D.F.E. directed the COGS. K.M., J.D., M.H., J.W.L., A.M.D., J.B., D.V., F.B., D.T. and P.H. contributed substantially to the COGS component of this project. All authors contributed to data and biological sample collection in the original studies and/or manuscript revision.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

The authors wish to thank all individuals who took part in these studies and all researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. We thank Regina Courtney, Jie Wu, Jing He, Mary Jo Daly and Bethanie Rammer for their help with sample preparation and genotyping, statistical analysis and technical support for the project at Vanderbilt. We are very grateful to Illumina, in particular Jen Stone, Sandy McBean, Jane Hadlington, Aziz Mustafa and Karen Cook for their help in designing and manufacturing the COGS array and technical assistance with the genotyping process. We thank Jonathan Tyrer, Ali Amin Al Olama and Karoline Kuchenbaecker for their help with COGS statistical analyses, and Shahana Ahmed, Melanie J. Maranian and Catherine S. Healey for their contributions to COGS genotyping quality control.

Conflict of Interest statement. None declared.

FUNDING

The work for this project at Vanderbilt was supported primarily by US NIH grants R01CA124558, R01CA148667, R01CA64277, R37CA070867 and U19 CA148065, as well as Ingram Professorship funds. The BCAC is funded by Cancer Research UK (C1287/A10118, C1287/A12014) and by the European Community's Seventh Framework Programme under grant agreement 223175 (HEALTH-F2-2009-223175) (COGS). BCAC meetings have been funded by the European Union COST program (BM0606). Genotyping of the COGS array was funded by the European Union (HEALTH-F2-2009-223175), Cancer Research UK (C1287/A10710), the Canadian Institutes of Health Research for the CIHR Team in Familial Risks of Breast Cancer program (J.S. and D.F.E.) and the Ministry of Economic Development, Innovation and Export Trade of Quebec (grant PSR-SIIRI-701) (J.S., D.F.E. and P.H). 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 CA68485). Funding for the constituent studies was provided by US NIH grants (R01CA124558, R01CA148667, R01CA64277, R37CA070867, R01CA118229, R01CA0 92585, R01CA122756, R01 CA137013, R01CA63464, R01CA54281 and CA132839), US Department of Defense Breast Cancer Research Program (BC011118 and BC050791), National Research Foundation, Ministry of Education, Science and Technology, National Biobank of Korea and National R&D Program of the Republic of Korea (2011-0001564, 1020350, 0620410-1, 2012-0000347 and KOBB-2011-03), National Natural Science Foundation of China (09KJA330001, 30771844 and 81072383), Taiwan Biobank (DOH97-01), Hong Kong Research Grant Council (7520/05M and 76730M) and the Japanese Ministry of Health, Labor and Welfare and Ministry of Education (17015049, 221S0001, 17015052 and H20-002). The pKARMA study was supported by Märit and Hans Rausings Initiative Against Breast Cancer and Cancer Risk Prediction Center, a Linneus Centre (contract 70867902) financed by the Swedish Research Council. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies. The sponsors of this study had no role in study design, data collection, data analyses, data interpretation or writing the final report. The corresponding author had full access to all the data used in this study and had final responsibility for the decision to submit for publication.

REFERENCES

1. Jemal, A., Bray, F., Center, M.M., Ferlay, J., Ward, E. and Forman, D. (2011) Global cancer statistics. *CA Cancer J. Clin.*, **61**, 69–90.
2. 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.
3. 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.
4. Cox, A., Dunning, A.M., Garcia-Closas, M., Balasubramanian, S., Reed, M.W., Pooley, K.A., Scollen, S., Baynes, C., Ponder, B.A., Chanock, S. *et al.* (2007) A common coding variant in CASP8 is associated with breast cancer risk. *Nat. Genet.*, **39**, 352–358.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
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. 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.
14. Cai, Q., Long, J., Lu, W., Qu, S., Wen, W., Kang, D., Lee, J.Y., Chen, K., Shen, H., Shen, C.Y. *et al.* (2011) Genome-wide association study identifies breast cancer risk variant at 10q21.2: results from the Asia Breast Cancer Consortium. *Hum. Mol. Genet.*, **20**, 4991–4999.
15. 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.
16. Haiman, C.A., Chen, G.K., Vachon, C.M., Canzian, F., Dunning, A., Millikan, R.C., Wang, X., Ademuyiwa, F., Ahmed, S., Ambrosone, C.B. *et al.* (2011) A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. *Nat. Genet.*, **43**, 1210–1214.
17. Ghoussaini, M., Fletcher, O., Michailidou, K., Turnbull, C., Schmidt, M.K., Dicks, E., Dennis, J., Wang, Q., Humphreys, M.K., Luccarini, C. *et al.* (2012) Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nat. Genet.*, **44**, 312–318.
18. Long, J., Cai, Q., Sung, H., Shi, J., Zhang, B., Choi, J.Y., Wen, W., Delahanty, R.J., Lu, W., Gao, Y.T. *et al.* (2012) Genome-wide association study in East Asians identifies novel susceptibility loci for breast cancer. *PLoS Genet.*, **8**, e1002532.
19. 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.
20. Zheng, W., Cai, Q., Signorello, L.B., Long, J., Hargreaves, M.K., Deming, S.L., Li, G., Li, C., Cui, Y. and Blot, W.J. (2009) Evaluation of 11 breast cancer susceptibility loci in African-American women. *Cancer Epidemiol. Biomarkers Prev.*, **18**, 2761–2764.
21. Chen, F., Chen, G.K., Millikan, R.C., John, E.M., Ambrosone, C.B., Bernstein, L., Zheng, W., Hu, J.J., Ziegler, R.G., Deming, S.L. *et al.* (2011) Fine-mapping of breast cancer susceptibility loci characterizes genetic risk in African Americans. *Hum. Mol. Genet.*, **20**, 4491–4503.

22. Huo, D., Zheng, Y., Ogundiran, T.O., Adebamowo, C., Nathanson, K.L., Domchek, S.M., Rebbeck, T.R., Simon, M.S., John, E.M., Hennis, A. *et al.* (2012) Evaluation of 19 susceptibility loci of breast cancer in women of African ancestry. *Carcinogenesis*, **33**, 835–840.
23. Cai, Q., Wen, W., Qu, S., Li, G., Egan, K.M., Chen, K., Deming, S.L., Shen, H., Shen, C.Y., Gammon, M.D. *et al.* (2011) Replication and functional genomic analyses of the breast cancer susceptibility locus at 6q25.1 generalize its importance in women of Chinese, Japanese, and European ancestry. *Cancer Res.*, **71**, 1344–1355.
24. Kim, H.C., Lee, J.Y., Sung, H., Choi, J.Y., Park, S.K., Lee, K.M., Kim, Y.J., Go, M.J., Li, L., Cho, Y.S. *et al.* (2012) A genome-wide association study identifies a breast cancer risk variant in ERBB4 at 2q34: results from the Seoul Breast Cancer Study. *Breast Cancer Res.*, **14**, R56.
25. Sueta, A., Ito, H., Kawase, T., Hirose, K., Hosono, S., Yatabe, Y., Tajima, K., Tanaka, H., Iwata, H., Iwase, H. and Matsuo, K. (2012) A genetic risk predictor for breast cancer using a combination of low-penetrance polymorphisms in a Japanese population. *Breast Cancer Res. Treat.*, **132**, 711–721.
26. Li, H., Beeghly-Fadiel, A., Wen, W., Lu, W., Gao, Y.T., Xiang, Y.B., Cai, Q., Long, J., Shi, J., Chen, K. *et al.* (2013) Gene-environment interactions for breast cancer risk among Chinese women: A report from the Shanghai Breast Cancer Genetics Study. *Am. J. Epidemiol.*, **177**, 161–170.
27. Michailidou, K., Hall, K., Gonzalez-Niera, A., Ghoussaini, M., Dennis, J., Milne, R.L., Schmidt, M.K., Chang-Claude, J., Bojesen, S.E., Bolla, M.K. *et al.* (2013) Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.*, in press.
28. Li, Y., Willer, C.J., Ding, J., Scheet, P. and Abecasis, G.R. (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.*, **34**, 816–834.
29. 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.
30. Cowper-Sal, L.R., Zhang, X., Wright, J.B., Bailey, S.D., Cole, M.D., Eeckhoute, J., Moore, J.H. and Lupien, M. (2012) Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat. Genet.*, **44**, 1191–1198.
31. Little, J., Higgins, J.P., Ioannidis, J.P., Moher, D., Gagnon, F., von Elm, E., Khoury, M.J., Cohen, B., Davey-Smith, G., Grimshaw, J. *et al.* (2009) STrengthening the REporting of Genetic Association studies (STREGA): an extension of the STROBE statement. *Ann. Intern. Med.*, **150**, 206–215.
32. von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gotzsche, P.C. and Vandenbroucke, J.P. (2007) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*, **370**, 1453–1457.
33. Zheng, W., Chow, W.H., Yang, G., Jin, F., Rothman, N., Blair, A., Li, H.L., Wen, W., Ji, B.T., Li, Q. *et al.* (2005) The Shanghai Women's Health Study: rationale, study design, and baseline characteristics. *Am. J. Epidemiol.*, **162**, 1123–1131.
34. Shu, X.O., Zheng, Y., Cai, H., Gu, K., Chen, Z., Zheng, W. and Lu, W. (2009) Soy food intake and breast cancer survival. *JAMA*, **302**, 2437–2443.
35. 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.
36. 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.
37. 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.
38. 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.
39. 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.
40. Han, S.A., Park, S.K., Ahn, S.H., Lee, M.H., Noh, D.Y., Kim, L.S., Noh, W.C., Jung, Y., Kim, K.S. and Kim, S.W. (2011) The Korean Hereditary Breast Cancer (KOHBRA) study: protocols and interim report. *Clin. Oncol. (R. Coll. Radiol.)*, **23**, 434–441.
41. 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.
42. 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.
43. 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.
44. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. and Sham, P.C. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
45. Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190–2191.
46. Collaborative Group of Hormonal factors in Breast Cancer (2001) 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*, **358**, 1389–1399.