Introduction

Breast cancer is the most common malignancy among women in the United States and many other parts of the world. Genetic factors play an important role in the etiology of breast cancer. Only a very small fraction of cases in the general population, however, can be explained by high-penetrance breast cancer susceptibility genes, such as *BRCA1* and *BRCA2*. Recent genome-wide association (GWA) studies [1–8], including our own study among Chinese women in Shanghai [6], have identified multiple common genetic susceptibility loci for breast cancer. Each of the common genetic factors identified thus far confer only a small to moderate risk for breast cancer. With the exception of our study, all other reported GWA studies have been conducted among women of European ancestry. GWA studies conducted in other populations could identify not only additional novel genetic variants for breast cancer but also help to fine map causal variants for regions reported from previous GWA studies.

In early 2009, we reported a novel genetic susceptibility locus at 6q25.1 for breast cancer risk in a fast-track replication of promising SNPs selected from a GWA scan of 1,505 cases and 1,522 controls recruited in the Shanghai Breast Cancer Study (SBCS) [6]. We have since increased the sample size for the initial GWA scan to 2,073 cases and 2,084 controls to increase the statistical power to identify novel genetic risk variants for breast cancer. We have recently completed the second fast-track replication using data and biological samples collected from 13,395 cases and 10,917 controls recruited in 12 studies of Asian and European ancestry. SNP rs4784227, located at 16q12.1, a region identified from a previous GWA study conducted in Europeans [1,3], was found to be a risk variant for breast cancer in Asian women independent of SNPs reported from the previous study [1,3]. *In vitro* experimental results provide strong support for the functional significance of this SNP and suggest that this SNP may explain the association observed for breast cancer in this locus. Herein, we report findings from this large genetic study of breast cancer.

Methods

Ethics statement

Approval was granted from relevant review boards in all study sites; all included subjects gave informed consent.

Study population

Included in this consortium project were 15,468 cases and 13,001 controls from 12 studies (Table 1). Detailed descriptions of these participating studies and demographic characteristics of study participants are provided in the supplement Text S1 and Table S1. Briefly, the consortium included 19,796 Chinese women from seven studies conducted in Shanghai [6,9] (three studies, n = 10,497), Tianjin [10] (n = 3,115), Nanjing [11,12] (n = 2,885), Taiwan [13] (n = 2,131), and Hong Kong [14] (n = 1,168); 3,214 Japanese women from three studies conducted in Hawaii [15] (n = 1,120), Nagoya [16] (n = 1,288), and Nagano (n = 806) [17]; and 5,459 European Americans from the Nashville Breast Health Study (NBHS, n = 3,172) and the Nurses' Health Study (NHS, n = 2,287, included as part of the Cancer Genetic Markers of Susceptibility Project - CGEMS). All cases and controls recruited in the Shanghai studies were included in Stages I and II, and subjects from the remaining Asian studies were included in Stage III. Data from CGEMS were used for help to select SNPs for Stage II. Cases and controls recruited in NBHS and the NHS (CGEMS) were included in the final stage to evaluate the generalizability of the findings.

Table 1. Characteristics of study participants and number of SNPs analyzed in each stage.

	Cases		Controls	
Study population	**N**	**Age[c]**	**N**	**Age**
Stage I (684,457 SNPs)[a]				
Shanghai – SBCS (Chinese)	2,073	49.3±8.3	2,084	49.4±8.5
Stage II (53 SNPs)				
Stage II total	4,425	53.9±10.2	1,915	52.8±9.2
Shanghai -SBCS (Chinese)	972	50.4±8.3	1,001	50.9±9.3
Shanghai - SBCSS (Chinese)	3,453	54.9±10.5		
Shanghai - SECS (Chinese)			914	54.9±8.5
Stage III (4 SNPs)[b]				
Stage III total	6,173	52.6±11.5	6,340	51.2±10.9
Tianjin (Chinese)	1,532	51.7±11.4	1,583	51.9±10.5
Nanjing (Chinese)	1,446	51.5±11.4	1,439	51.3±11.2
Taiwan (Chinese)	1,066	51.5±10.7	1,065	47.5±10.1
Hong Kong (Chinese)	517	45.8±9.5	651	45.6±10.3
Nagoya - Japan (Japanese)	644	51.4±11.0	644	51.1±10.9
Nagano - Japan (Japanese)	403	53.7±10.5	403	53.9±10.2
Hawaii - MEC (Japanese)	565	65.2±8.4	555	60.4±8.4
European Americans [b]				
NBHS	1,652	54.9±10.2	1,520	52.2±11.0
CGEMS	1,145		1,142	
Total	15,468	52.5±10.7	13,001	51.2±10.2
Chinese	11,059	51.8±10.3	8,737	50.4±10.1
Japanese	1,612	56.8±11.8	1,602	55.0±10.7
European Americans	2,797		2,662	

a Selected from SNPs included in the Affymetrix 6.0 SNP array with MAF≥1%, call rate≥95%, and QC consistency.
b With the exception of studies in Nanjing (2 SNPs), Tianjin (3 SNPs) and Nagoya Japan (1 SNP), NBHS (3 SNPs), CGEMS (2 SNPs genotyped and 2 SNPs imputed), four SNPs were analyzed in all other studies.
c Mean ± SD.
doi:10.1371/journal.pgen.1001002.t001

Genotyping and quality control procedures

Genotyping for Stage I has been described previously [6]. Briefly, the initial 300 subjects were genotyped using the Affymetrix GeneChip Mapping 500 K Array Set and the remaining 3,918 subjects were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0. We included one negative control and three positive quality control (QC) samples from 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 127 positive QC samples were successfully genotyped and the average concordance rate was 99.9% with a median value of 100%. The sex of all study samples was confirmed to be female. The identity-by-descent analysis based on identity-by-state was conducted to detect first-degree cryptic relationships using PLINK, version 1.06. All samples with a call rate <95% were excluded. The SNPs were excluded if: (i) minor allele frequency (MAF) <1%, (ii) call rate <95%, or (iii) genotyping concordance rate <95% in quality control samples. The final dataset included 2,073 cases and 2,084 controls for 684,457 markers.

Genotyping for Stage II was completed using the iPLEX Sequenom MassArray platform. Included in each 96-well plate as QC samples were two negative controls (water), two blinded

Author Summary

Breast cancer is one of the most common malignancies among women worldwide. Genetic factors play an important role in the etiology of breast cancer. To identify genetic susceptibility loci for breast cancer, we performed a genome-wide association study in 15,468 breast cancer cases and 13,001 controls. A single nucleotide polymorphism (SNP) rs4784227 located on chromosome 16q12.1, a previously-reported region for breast cancer risk, was found to be associated with breast cancer risk. The association of this SNP with breast cancer risk remained highly significant in Asians after adjusting all previously-reported SNPs in this region. *In vitro* biochemical experiments using both luciferase reporter and electrophoretic mobility shift assays confirmed the functional importance of this SNP. Our results demonstrate the importance of conducting genetic association studies in populations with different genetic backgrounds to identify functional variants.

duplicates, and two samples from the HapMap project. To compare the consistency between the Affymetrix and Sequenom platforms, we also included 124 samples from Stage I that were genotyped by Affymetrix SNP 6.0. The mean concordance rate was 99.7% for the blind duplicates, 98.8% for HapMap samples, and 98.6% between Sequenom and Affymetrix 6.0 genotyping.

Genotyping for Stage III and NBHS was performed using TaqMan at five different centers. The genotyping assay protocol was developed and validated at the Vanderbilt Molecular Epidemiology Laboratory, and TaqMan genotyping assay reagents were provided to investigators from the Tianjin study (Tianjin Cancer Institute and Hospital), Nanjing study (Nanjing Medical University), Multiethnic Cohort Study (MEC, University of Southern California), and Nagano Breast Cancer study (Japan National Cancer Center), who conducted the genotyping assays at their own laboratories. Samples from the four other studies (Hong Kong, Taiwan, Nagoya, and Nashville) were genotyped at the Vanderbilt Molecular Epidemiology Laboratory. During the genotyping, two negative controls were included in each 96-well plate, along with 30 unrelated European and 45 Chinese samples from the HapMap project genotyped together with each study for QC purposes. The consistency rate was 100.0% for the HapMap samples comparing genotyping data obtained from the current study with data obtained in the HapMap project. Each of the non-Vanderbilt laboratories was asked to genotype a trial plate containing DNA from 70 Chinese samples before genotyping study samples. The consistency rate across all centers for these trial samples was 100% compared with genotypes previously determined at Vanderbilt. In addition, replicate samples comparing 3–7% of all study samples were dispersed among the genotyping plates at all centers.

Plasmid constructs and luciferase reporter assays

A 3.0 kb DNA fragment containing major allele (C) of rs4784227 was PCR amplified by using forward primer 5′-GATCAGCTAGCCATAGTGTGGTAGCTAGTTG-3′ and backward primer 5′-GATCA CTCGAGCTGCTGGGCTTAGCTACAAG-3′. This fragment was subcloned into luciferase reporter vector, pGL3 basic, pGL3 promoter, and pGL3 enhancer (Promega, WI) between *Nhe*I and *Xho*I restriction sites, respectively. The minor allele (T) sequence was generated by using QuickChange Site-Directed Mutagenesis Kit (Strategene, La Jolla, CA) with the following pair of oligonucleotides, 5′- GAGTATTTACATCACAATAATCAGCAAACACTACAAATTGGGAC-3′ and 5′- GTCCCAATTTGTAGTGTTTGCTGATTATTGTGATGTAAATACTC-3′. All DNA constructs were verified by sequencing analyses. Transfection was performed with the use of FuGene 6 Transfection Reagent (Roche Diagnostics, Indianapolis, IN) in triplicate for each of the constructs. Briefly, $1-2\times10^5$ cells of HEK 293, MCF-7, MCF10A, and MDA-231 cells were seeded in 24-well plates and co-transfected with pGL4.73, a Renilla expressing vector which served as a reference for transfection efficiency. Thirty-six to 48 hours later the cells were lysed with Passive Lysis Buffer, and luminescence (relative light units) was measured using the Dual-Luciferase Assay System (Promega, WI). The rs4784227 regulatory activity was measured as a ratio of firefly luciferase activity to *renilla* luciferase activity, and the mean from four independent experiments are presented.

Electrophoretic mobility shift assay

Biotin-labeled, double stranded oligonucleotide probes 5′-ATTTGTAGTGTTTGCCGATTATTGTGATGT-3′ and 5′-ACATCACAATAATCGGCAAACACTACAAAT-3′, and 5′-ATTTGTAGTGTTTGCTGATTATTGTGATGT-3′and 5′- ACATCACAATAATCAGCAAACACTACAAAT-3′ containing the major and minor allele sequence of rs4784227 were synthesized. The probes were incubated with nuclear protein extracts from MCF10A, MCF7, and MDA-MB-231 cells, in the presence or absence of competitors, i.e. unlabelled probes. Protein-DNA complexes were resolved by polyacrylamide gel electrophoresis and detected using a LightShift Chemiluminescent EMSA kit (Pierce Biotechnology, Rockford, IL).

Statistical analyses

In Stage I, PLINK version 1.06 was used to analyze genome-wide data. Population structure was investigated by using the principal component analysis implemented in EIGENSTRAT [18] (http://genepath.med.harvard.edu/~reich/Software.htm). A set of 12,533 SNPs with MAF≥5% in Chinese and a distance ≥25 kb between two adjacent SNPs was selected to evaluate the population structure. The first two principal components were included in logistic regression models for adjustment of population structures. Odds ratios (OR) and 95% confidence intervals (CIs) were estimated by logistic regression analysis. ORs were also estimated for the variant allele based on a log-additive model. Age was adjusted for in the analyses of Stages I and II data. In Stage III, individual data were obtained from each study for a pooled analysis. ORs from multiple studies were adjusted for age and study site. Heterogeneity across studies and between ethnicities was assessed with likelihood ratio tests. Stratified analyses by ethnicity, menopausal status, and estrogen receptor (ER) status were carried out. P-values based on 2-tailed tests are presented.

Individual genotyping data from the Cancer Genetic Markers of Susceptibility (CGEMS, http://cgems.cancer.gov/data/) study were obtained through an approved data request application in order to perform meta-analyses of GWA scan data from both the Shanghai studies and the CGEMS project. Program MACH (http://www.sph.umich.edu/csg/abecasis/MACH/) was used for genotype imputation that determines the probability distribution of missing genotypes conditional on a set of known haplotypes while simultaneously estimating the fine-scale recombination map. For the Shanghai studies, imputation was based on 660,118 autosomal SNPs genotyped in Stage I that had a MAF>1% and passed the QC procedure, using the phased Asian data from HapMap Phase II (release 22) as the reference. A total of 2,272,352 SNPs showed an imputation quality score ≥0.90. The CGEMS GWA scan data were genotyped using Illumina

HumanHap550 for 1,142 breast cancer cases and 1,145 controls nested within the Nurses' Health Study cohort. For CGEMS, genotypes were imputed based on 513,602 autosomal SNPs with MAF>1%, using phased CEU data from HapMap Phase II (release 22) as the reference. A total of 2,168,847 SNPs showed an imputation quality score ≥0.90. To evaluate associations between imputed SNP data and breast cancer risk, logistic regression (additive model) was used, in which SNPs were represented by the expected allele count, an approach that takes into account the degree of uncertainty in the genotype imputation.

Meta-analyses of GWA scan data for SBCS and CGEMS were conducted for 1,968,549 SNPs with a MAF ≥1% in both populations and imputation quality scores ≥0.90. Meta-analyses were performed using a weighted z-statistics method, where weights were proportional to the square root of the number of individuals in each sample and standardized such that the weights added up to one. The z-statistic summarizes the magnitude and direction of the effect relative to the reference allele. An overall z-statistic and p value were then calculated from the weighted average of the individual statistics. Calculations were implemented in the METAL package (http://www.sph.umich.edu/csg/abecasis/Metal).

SNP selection for validation in Stage II. In the present study, we selected 59 promising SNPs for fast-track replication in Stage II, including 4,425 cases and 1,915 controls recruited in the Shanghai studies. Only SNPs included in the Affymetrix 6.0 SNP arrays were selected for Stage II evaluation. Selection criteria for these SNPs were: (i) MAF≥5%, (ii) very clear genotyping clusters, (iii) not in strong LD ($r^2 \leq 0.5$) with any of the previously confirmed breast cancer genetic risk variants in CHB/JPT, (iv) consistent with HWE with P>0.01 in controls, (v) $P<0.001$ in the meta-analyses of CGEMS and SBCS GWA scan data for SNPs on Affymetrix 6.0 array, having the same direction of association in both studies, and $P \leq 0.01$ for SBCS GWA scan data.

Results

Of the 53 successfully genotyped SNPs in Stage II (Table S2), highly significant associations with breast cancer risk were found for rs4784227 (16q12.1) with OR (95% CI) of 1.23 (1.16–1.31) per T allele (P for trend, 1.3×10^{-8}) (Table 2). Three other SNPs also showed a significant or marginally significant association with breast cancer risk. These four SNPs were selected for further validation in Stage III, which included 6,173 cases and 6,340 controls of Asian ancestry from seven studies in the Asia Breast Cancer Consortium (Methods; Table S1). SNP rs4784227 was consistently associated with breast cancer risk in all studies (Figure 1), with an OR of 1.25 (95% CI: 1.20–1.31, $P=3.2\times10^{-25}$) in the pooled analysis of Asian samples from all three stages. No heterogeneity in the association of this SNP with breast cancer was observed across the studies included in the consortium. The association of rs4784227 with breast cancer risk was observed in both pre-menopausal (OR: 1.24 (1.17–1.32) and $P=6.5\times10^{-12}$), and post-menopausal women (OR: 1.27 (1.19–1.35) and $P=3.0\times10^{-14}$) (data not shown in tables). The positive association was stronger in ER(+) breast cancer (per allele OR = 1.29, 95% CI = 1.23–1.36, $P=3.0\times10^{-23}$) than in ER(−) breast cancer (per allele OR = 1.19, 95% CI = 1.12–1.26, $P=1.3\times10^{-8}$). In case-only analyses, when compared with cases with ER(−) cancer, ORs associated with ER(+) breast cancer

Table 2. Results from Stages I to III for rs4784227 and Stages I and II for other three SNPs that showed promising associations with breast cancer risk in Stage II.

						OR(95%CI)[c]			
SNP (chr)	Allele[a]	Stage	No of cases	No of controls	Frequency[b]	Heter	Homo	Per allele	P for trend
rs10479046 (5q31.1)	C/T	I	2,071	2,083	0.73/0.70	1.22(0.96–1.55)	1.42(1.12–1.80)	1.18(1.07–1.30)	7.2×10^{-4}
		II	4,320	1,854	0.71/0.70	1.07(0.87–1.30)	1.15(0.94–1.40)	1.07(0.99–1.17)	0.10
		I & II	6,391	3,937	0.72/0.70	1.09(0.94–1.27)	1.22(1.06–1.42)	1.11(1.05–1.18)	7.5×10^{-4}
rs3829849 (9q33.3)	T/C	I	2,067	2,081	0.11/0.09	1.31(1.12–1.54)	1.15(0.66–1.99)	1.25(1.08–1.44)	0.002
		II	4,331	1,871	0.10/0.09	1.19(1.03–1.38)	2.10(1.06–4.18)	1.23(1.08–1.41)	0.003
		I & II	6,398	3,952	0.11/0.09	1.24(1.11–1.38)	1.34(0.89–2.01)	1.22(1.11–1.34)	4.4×10^{-5}
rs7966820 (12q24.1)	T/C	I	2,063	2,074	0.15/0.13	1.28(1.10–1.48)	1.33(0.86–2.05)	1.24(1.09–1.40)	7.8×10^{-4}
		II	4,351	1,856	0.15/0.14	1.00(0.88–1.14)	2.00(1.28–3.12)	1.10(0.98–1.23)	0.10
		I & II	6,414	3,930	0.15/0.13	1.12(1.02–1.23)	1.60(1.19–2.16)	1.16(1.07–1.26)	4.0×10^{-4}
rs4784227 (16q12.1)	T/C	I	2,066	2,078	0.29/0.25	1.18(1.04–1.34)	1.43(1.13–1.82)	1.19(1.08–1.31)	4.2×10^{-4}
		II	4,280	1,843	0.28/0.23	1.32(1.18–1.48)	1.62(1.29–2.05)	1.30(1.19–1.42)	1.3×10^{-8}
		I & II	6,346	3,921	0.28/0.24	1.24(1.14–1.35)	1.50(1.28–1.77)	1.23(1.16–1.31)	2.1×10^{-10}
		I, II & III[d]	12,336	10,140	0.29/0.24	1.28(1.21–1.35)	1.52(1.37–1.69)	1.25(1.20–1.31)	3.2×10^{-25}

a Risk allele/reference allele.
b Risk allele frequency in cases/controls.
c In Stage I, adjusted for age and the first two principal components for population structure, in Stage II adjusted for age.
d SNP rs4784227 was replicated in Stage III, and combined results from all three stages are presented.
doi:10.1371/journal.pgen.1001002.t002

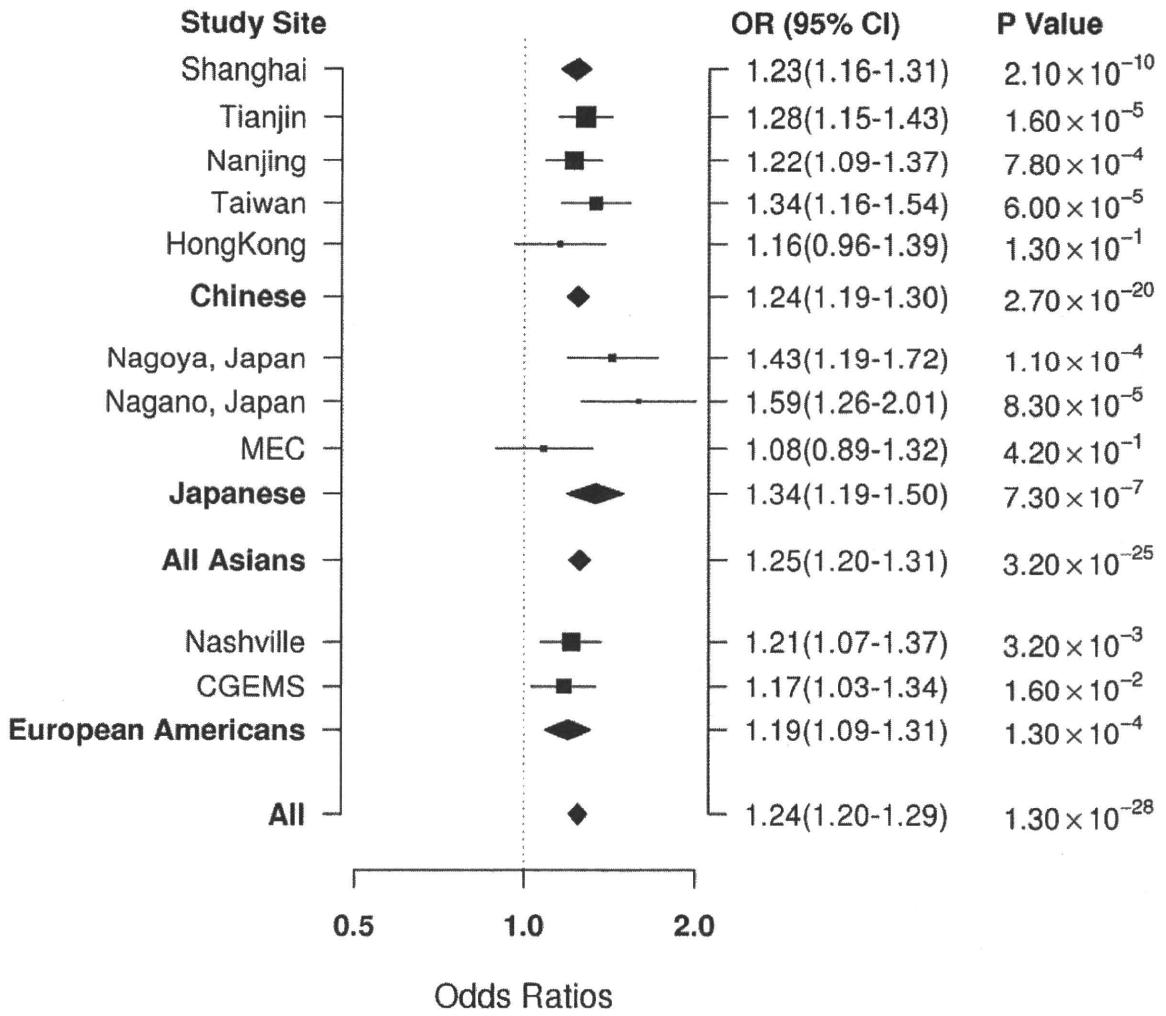

Figure 1. Per-allele OR for rs4784227 in association with breast cancer risk by study and ethnic groups.
doi:10.1371/journal.pgen.1001002.g001

were found to be 1.09 (95% CI: 1.03–1.16; P for trend, 5.8×10^{-3}). None of the other three SNPs that showed a significant association in Stage II, however, were replicated in Stage III (Table S3).

SNP rs4784227 is located in 16q12.1, a region where three genetic risk variants for breast cancer (rs8051542, rs12443621, and rs3803662) were reported previously in a study conducted among women of European ancestry [1]. Of these three previously reported SNPs, the closest (rs3803662) is approximately 12.8 Kb away from rs4784227. The linkage disequilibrium (LD) pattern of this region in Asians is very different from the pattern found in European descendents (Figure 2 and Table 3). In Stage I and II samples, SNP rs4784227 is in low LD with previously-reported SNPs, with r^2 being 0.07, 0.14, and 0.37 for rs12443621, rs3803662, and rs8051542, respectively (Table 3). In European Americans included in the HapMap project, however, SNP rs4784227 is in strong LD with SNP rs3803662 ($r^2=0.86$) but weakly correlates with the other two SNPs. SNPs rs8051542 and rs3803662 each showed a significant association with breast cancer risk ($P=2.0\times10^{-3}$ and $P=1.7\times10^{-4}$) in a combined analysis of data from Stages I and II (Table 4). However, after adjusting for rs4784227 the association with rs8051542 disappeared. Although the positive association with rs3803662 remained, it was of only borderline significance ($P=0.12$). SNP rs12443621, however, showed no association with breast cancer risk, which is consistent with results reported by the initial study of Asian women [1]. The association of rs4784227 with breast cancer risk remained highly significant after adjusting for these three previously-reported SNPs, individually or in combination (Table 4). Haplotype analyses of these four SNPs showed that all haplotypes containing the T allele of rs4784227 were associated with an increased risk of breast cancer, although not all point estimates were statistically significant at $P<0.05$ due to a small sample size for several haplotypes (Table S4).

In studies conducted in European Americans, SNP rs4784227 also showed a significant association with a per allele OR (95% CI) of 1.17 (1.03–1.34) in CGEMS and 1.21 (1.07–1.37) in NBHS (Table 5). A significant (NBHS) or marginally significant (CGEMS) association was observed for rs3803662, a previously reported SNP, but not for two other previously-reported SNPs. After adjusting for rs4784227, no association with rs3803662 was seen. On the other hand, the positive association with rs4784227

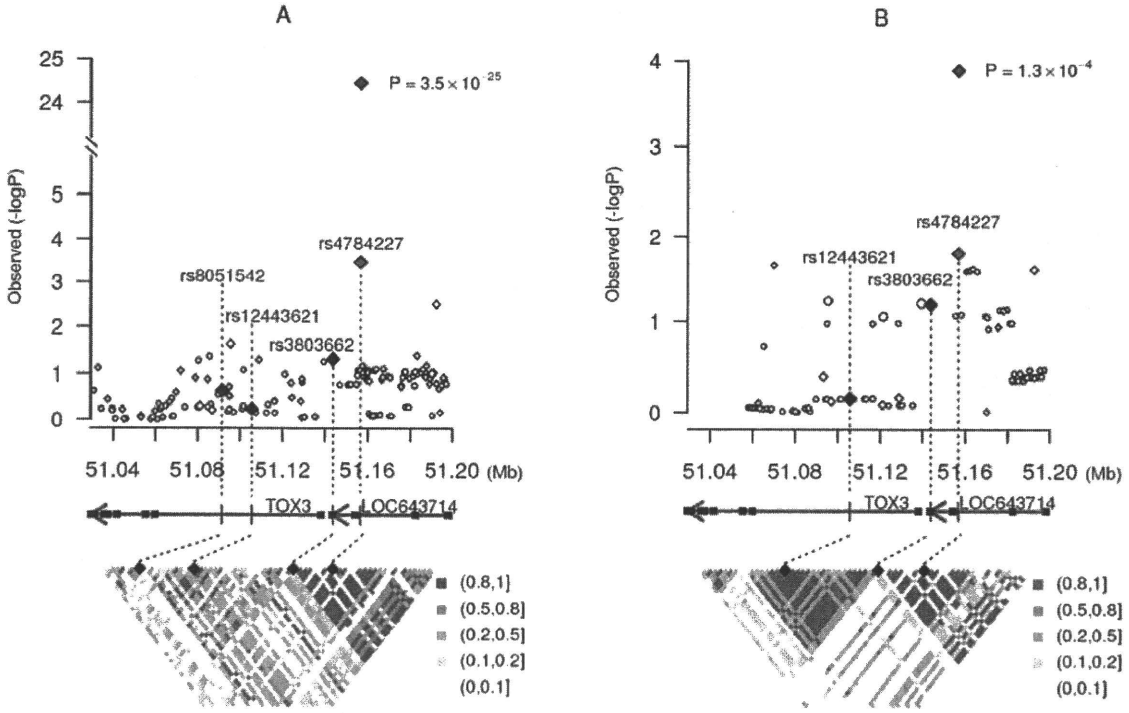

Figure 2. SNP association at 16q12 with breast cancer risk. In SBCS (A) and CGEMS (B). Top panel: Results ($-\log_{10}$P) are shown for the associations of breast cancer risk with directly genotyped (diamonds) and imputed (circles) SNPs located in the *TOX3* and *LOC643714* genes. P-values for rs4784227 are shown in red for Stage I data and in blue for the combined data. The three previously-reported SNPs (rs8051542, rs12443621, and rs3803662) are highlighted in black. Middle panel: Genomic view at 16q12.1. Gene locations are from the March 2006 UCSC genome browser assembly. Bottom panel: Estimates of pairwise LD (r^2) for common SNPs (with MAF≥5%) from HapMap release 23a in the region from 10 kb downstream of rs8051542 to 10 kb upstream of rs4784227.
doi:10.1371/journal.pgen.1001002.g002

remained after adjusting for rs3803662 or the other two SNPs, although the association was no longer statistically significant at $P<0.05$.

To evaluate whether SNP rs4784227 has any intrinsic regulatory function, we conducted an *in vitro* luciferase assay in four cell lines including metastatic breast cancer cell MDA231, non-metastatic breast cancer cell MCF-7, breast epithelial cell MCF10A, and HEK293. Luciferase reporter constructs containing a 3 kb DNA fragment with the reference allele C and the risk allele T of rs4784227, respectively, were generated and transiently transfected into these cells. By comparing to the respective empty vectors, no luciferase activity change was observed in pGL3 basic and pGL3 enhancer vectors that harbor rs4784227 fragments, which indicate that rs4784227 fragments do not have intrinsic promoter activity (data not shown). In contrast, in the pGL3 promoter vector, fragments containing rs4784227 reduced luciferase activity, and the reduction was more apparent in fragments containing risk allele T than the reference allele C (Figure 3A). With the exception of the MCF7 cells, the difference between the T and C allele was statistically significant at P≤0.05.

To investigate whether the DNA sequence containing rs4784227 may interact with nuclear proteins and if so, whether a single nucleotide change in the rs4784227 site may alter the protein-DNA interactions, we performed electrophoretic mobility shift assays. In these assays, oligonucleotide probes corresponding to the reference allele C or the risk allele T were incubated with nuclear protein extracts from MCF10A, MCF-7, and MDA-231 cells. Compared with reference allele C, risk allele T in rs4784227 resulted in an altered DNA-protein complex intensity in these cells (B and II) (Figure 3B). In contrast, risk allele T did not alter the intensity of the nonspecific DNA-protein complex band (I). These results were unaffected by the presence of large amounts of unlabeled competitors.

Table 3. Linkage disequilibrium patterns among rs4784227 and the three previously-reported SNPs in 16q12.1.

	SNPs	rs8051542	rs12443621	rs3803662	rs4784227
Chinese	rs8051542		0.1	0.08	0.37
	rs12443621	0.77		0.04	0.07
	rs3803662	0.81	0.25		0.14
	rs4784227	0.76	0.52	0.87	
Caucasians					
	rs8051542		0.01	0.13	0.12
	rs12443621	0.13		0.3	0.3
	rs3803662	0.52	0.92		0.86
	rs4784227	0.54	1	1	

D′: Lower left triangle.
r2: Upper right triangle.
Caucasians: European samples included in the HapMap project.
Chinese: subjects in Stages I and II.
doi:10.1371/journal.pgen.1001002.t003

Table 4. Association of rs4784227 and the three previously-reported SNPs at 16q12with breast cancer risk among Chinese women in Stages I and II.

					OR(95%CI)			
Tested SNP	Adjusted SNPs	No of cases	No of controls	Frequency[a]	Heter	Homo	Per allele	P for trend
rs4784227	None	6,346	3,921	0.28/0.24	1.24(1.14–1.35)	1.50(1.28–1.77)	1.23(1.16–1.31)	2.1×10^{-10}
rs8051542	None	6,158	3,658	0.20/0.18	1.10(1.01–1.21)	1.36(1.09–1.69)	1.13(1.05–1.21)	0.002
rs12443621	None	2,954	2,997	0.57/0.57	0.99(0.86–1.14)	1.02(0.88–1.18)	1.01(0.94–1.09)	0.78
rs3803662	None	6,345	3,795	0.68/0.65	1.26(1.10–1.44)	1.34(1.17–1.53)	1.12(1.06–1.19)	1.7×10^{-4}
rs4784227	rs8051542	6,082	3,614	0.28/0.24	1.27(1.14–1.40)	1.57(1.29–1.91)	1.26(1.16–1.37)	9.5×10^{-8}
rs4784227	rs12443621	2,920	2,961	0.29/0.24	1.32(1.18–1.47)	1.63(1.32–2.01)	1.29(1.19–1.41)	3.1×10^{-9}
rs4784227	rs3803662	6,221	3,749	0.28/0.25	1.20(1.10–1.32)	1.41(1.19–1.68)	1.20(1.11–1.28)	7.0×10^{-7}
rs4784227	all 3 SNPs	2,702	2,646	0.29/0.24	1.31(1.14–1.50)	1.60(1.23–2.08)	1.28(1.15–1.44)	1.6×10^{-5}
rs8051542	rs4784227	6,082	3,614	0.20/0.18	0.93(0.84–1.04)	0.97(0.76–1.25)	0.95(0.87–1.05)	0.33
rs12443621	rs4784227	2,920	2,961	0.57/0.57	0.92(0.80–1.07)	0.89(0.77–1.04)	0.95(0.88–1.02)	0.16
rs3803662	rs4784227	6,221	3,749	0.68/0.65	1.17(1.03–1.35)	1.17(1.01–1.35)	1.05(0.99–1.13)	0.12

a Risk allele frequency in cases/controls. Risk/reference alleles (based on forward strand) are T/C for rs8051542 and rs4784227 and A/G for rs12443621 and rs3803662.
doi:10.1371/journal.pgen.1001002.t004

Discussion

In this multi-stage GWA study of over 15,486 cases and 13,001 controls, we identified SNP rs4784227 as highly significantly associated with breast cancer in both Asians (per allele OR = 1.25, 95% CI = 1.20–1.31, $P=3.2\times10^{-25}$) and European Americans (per allele OR = 1.19, 95% CI = 1.09–1.31, $P=1.3\times10^{-4}$). SNP rs4784227 is located at 16q12.1, a region reported previously to harbor breast cancer genetic risk variants among European descendents [1,3]. In Asians, however, this SNP is either not in LD or only in weak LD with any of the three previously-reported SNPs in these regions, and adjusting for these SNPs did not alter the association of breast cancer with this newly-identified SNP. Although in European Americans rs4784227 is in strong LD with one of the previously-reported SNPs, rs3803602, the positive association of rs4784227 with breast cancer remained after adjusting for previously-reported SNPs. *In vitro* experiments showed that risk allele *T* reduced luciferase activity and altered DNA-protein binding patterns. These results implicate rs4784227 as a functional genetic risk variant for breast cancer, and this SNP may explain, at least partially, the association of breast cancer with other SNPs identified in 16q12.1.

SNP rs4784227 is located 18.4 kb upstream of the *TOX3* gene and in the evolutionarily-conserved region of intron of the *LOC643714* gene. Several transcription factors are predicted to bind to this SNP (http://www.cbrc.jp/research/db/TFSEARCH.html). This SNP, however, has not been shown to be in the coding region for any non-coding RNA or miRNA/snoRNA/scaRNA based on UCSC Genome Browser. Our luciferase reporter assays showed that the intronic region harboring rs4784227 may have

Table 5. Association of rs4784227 and the three previously-reported SNPs at 16q12with breast cancer risk among European American women.

Tested SNP [a]	Study	SNPs Adjusted	No of cases	No of controls	Frequency[b]	OR (95% CI)	P for trend
rs4784227	CGEMS	None	1145	1142	0.29/0.25	1.17(1.03–1.34)	0.016
	NBHS	None	1357	1148	0.28/0.25	1.21(1.07–1.37)	0.003
rs3803662	CGEMS	None	1145	1142	0.3/0.27	1.13(0.99–1.28)	0.06
	NBHS	None	1615	1467	0.31/0.27	1.22(1.09–1.36)	4.0×10^{-4}
rs8051542	NBHS	None	1587	1439	0.46/0.44	1.07(0.97–1.19)	0.18
rs12443621	CGEMS	None	1145	1142	0.50/0.49	1.02(0.91–1.15)	0.70
rs4784227	NBHS	rs8051542	1263	1080	0.28/0.25	1.22(1.06–1.40)	0.007
rs4784227	CGEMS	rs12443621	1145	1142	0.29/0.25	1.25(1.06–1.47)	0.007
rs4784227	Both	rs3803662	2431	2243	0.29/0.25	1.22(0.94–1.59)	0.13
rs8051542	NBHS	rs4784227	1263	1080	0.45/0.44	0.98(0.86–1.11)	0.74
rs12443621	CGEMS	rs4784227	1145	1142	0.50/0.49	0.91(0.79–1.05)	0.19
rs3803662	Both	rs4784227	2431	2243	0.31/0.27	0.97(0.75–1.25)	0.82

a In CGEMS, data on rs8051542 was not available, and In NBHS data on for rs12443621 was not available.
b Risk allele frequency in cases/controls. Risk/reference alleles (based on forward strand) are T/C for rs8051542 and rs4784227 and A/G for rs12443621 and rs3803662.
doi:10.1371/journal.pgen.1001002.t005

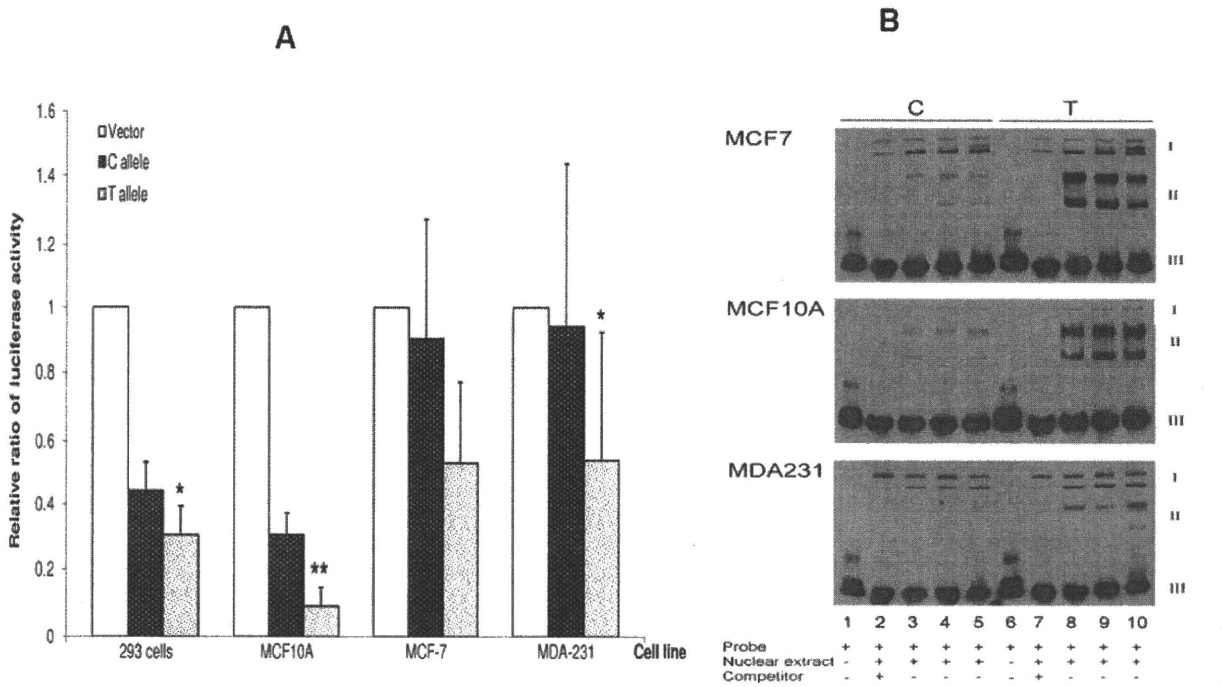

Figure 3. *In vitro* functional characterization of SNP rs4784227. (A) Effect of rs4784227 on luciferase reporter activity. HEK 293, MCF10A, MCF-7, and MDA-231 cells were transiently transfected with pGL3 promoter vector and the constructs carrying the reference allele (C) and risk allele (T) of rs4784227, respectively. Relative luciferase activities are shown as mean ± SD of four experiments (relative to empty vector). Statistical analysis was done using Student's t-test comparing C and T alleles (* $P<0.05$, ** $P<0.01$, n = 4). (B) Electrophoretic mobility shift assays. Nuclear protein extracts from MCF-7 (top panel), MCF10A (middle panel), and MDA-231 (bottom panel) cells were incubated with biotin-labeled probes corresponding to reference allele C (lanes 1–5) or the risk allele T (lanes 6–10) of rs4784227 in the absence or presence of competitors. Lanes 1 and 6, no nuclear extracts; lanes 2 and 7, unlabeled competitor in 200-fold molar excess; lanes 3 and 8 (5 mM MgCl2), lanes 4 and 9 (2.5 mM MgCl2), and lanes 5 and 10 (1.25 mM MgCl2), no competitor. I: nonspecific DNA-protein complex bands from MCF-7, MCF10A, and MDA-231 cells; II: specific DNA-protein complex bands; III: free biotin-labeled probes.
doi:10.1371/journal.pgen.1001002.g003

intrinsic repressor activities, suggesting that rs4784227 may affect its underlying gene *LOC643714* or its neighborhood gene expression and thus affect breast cancer risk. The rs4784227-associated repressing activity could be the result of differential binding affinity of transcription machinery to the rs4784227-containing DNA sequences. We examined this hypothesis by conducting electrophoretic mobility shift assays and confirmed that the risk *T* allele of rs4784227 significantly alter DNA-nuclear protein(s) interactions. Thus, it is possible that inhibitory nuclear protein(s) selectively bind to the risk allele *T* to repress transcription. A database search (http://www.cbrc.jp/research/db/TFSEARCH.html) for transcription factor binding sites showed that the sequence at the rs4784227 site has a high degree of similarity with several consensus elements recognized by transcription factors, of which HNF-3b and C/EBP prefer to bind DNA fragments with the risk T allele of this SNP site. However, these putative transcription factors or their associated proteins have not been confirmed to be involved in the regulation of *LOC643714* or its nearby genes.

In summary, through a GWA study we have identified and confirmed rs4784227 as a genetic risk variant for breast cancer. *In vitro* experiments showed a functional significance of this SNP that may explain the association of breast cancer with other SNPs identified at locus 16q12.1. This study demonstrates the importance of conducting genetic association studies in populations with different LD structures to identify causal genetic variants for breast cancer and other complex diseases.

Supporting Information

Table S1 Characteristics of participating studies in the Asian Breast Cancer Consortium.
Found at: doi:10.1371/journal.pgen.1001002.s001 (0.06 MB DOC)

Table S2 Associations of 49 SNPs evaluated in Stage II but not in Stage III.
Found at: doi:10.1371/journal.pgen.1001002.s002 (0.27 MB DOC)

Table S3 Results for four SNPs selected for Stage III evaluation by study.
Found at: doi:10.1371/journal.pgen.1001002.s003 (0.09 MB DOC)

Table S4 Haplotype analyses of the four SNPs at 16q12 with breast cancer risk in Stages I and II.
Found at: doi:10.1371/journal.pgen.1001002.s004 (0.05 MB DOC)

Text S1 Study participants.
Found at: doi:10.1371/journal.pgen.1001002.s005 (0.11 MB DOC)

Acknowledgments

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

wish to thank the study participants and research staff for their contributions and commitment to this project, Regina Courtney and the late Qing Wang for DNA preparation, and Brandy Venuti for clerical support in the preparation of this manuscript. Sample preparation and genotyping assays at Vanderbilt were conducted at the Survey and Biospecimen Shared Resources and Vanderbilt Microarray Shared Resource.

Author Contributions

Conceived and designed the experiments: W Zheng. Performed the experiments: Q Cai, S Qul, B Huang. Wrote the paper: J Long, Q Cai, XO Shu, S Qul, W Zheng. Designed and directed the study: W Zheng. Drafted the manuscript: J Long, W Zheng, Q Cai, XO Shu, S Qul. Directed the studies that contributed data to consortium or contributed to data and biological collection of these original studies: XO Shu, Y Zheng, K Gu, W Wang, YB Xiang, J Cheng, K Chen, L Zhang, H Zheng, CY Shen, CS Huang, MF Hou, H Shen, Z Hu, F Wang, SL Deming, MC Kelley, MJ Shrubsole, US Khoo, KYK Chan, SY Chan, CA Haiman, BE Henderson, L Le Marchand, M Iwasaki, Y Kasuga, S Tsugane, K Matsuo, K Tajima, H Iwata, YT Gao, W Lu, W Zheng. Coordinated the genetic study: J Long. Managed the genotyping data: J Long. Performed statistical analyses: J Long, C Li, W Wen. Directed lab operations: Q Cai. Conducted in vitro functional experiments: Q Cai, S Qul, B Huang. Performed genotyping assays: J Shi, G Li.

References

1. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, et al. (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447: 1087–1093.
2. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, 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.
3. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, et al. (2007) Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 39: 865–869.
4. Gold B, Kirchhoff T, Stefanov S, Lautenberger J, Viale A, et al. (2008) Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. Proc Natl Acad Sci U S A 105: 4340–4345.
5. Stacey SN, Manolescu A, Sulem P, Thorlacius S, Gudjonsson SA, et al. (2008) Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 40: 703–706.
6. Zheng W, Long J, Gao YT, Li C, Zheng Y, et al. (2009) Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nat Genet 41: 324–328.
7. Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, 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.
8. Ahmed S, Thomas G, Ghoussaini M, Healey CS, Humphreys MK, et al. (2009) Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. Nat Genet 41: 585–590.
9. Gao YT, Shu XO, Dai Q, Potter JD, Brinton LA, et al. (2000) Association of menstrual and reproductive factors with breast cancer risk: results from the Shanghai Breast Cancer Study. Int J Cancer 87: 295–300.
10. Zhang L, Gu L, Qian B, Hao X, Zhang W, et al. (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.
11. Liang J, Chen P, Hu Z, Zhou X, Chen L, et al. (2008) Genetic variants in fibroblast growth factor receptor 2 (FGFR2) contribute to susceptibility of breast cancer in Chinese women. Carcinogenesis 29: 2341–2346.
12. Wang Y, Hu Z, Liang J, Wang Z, Tang J, et al. (2008) A tandem repeat of human telomerase reverse transcriptase (hTERT) and risk of breast cancer development and metastasis in Chinese women. Carcinogenesis 29: 1197–1201.
13. Ding SL, Yu JC, Chen ST, Hsu GC, Kuo SJ, et al. (2009) Genetic variants of BLM interact with RAD51 to increase breast cancer susceptibility. Carcinogenesis 30: 43–49.
14. Chan KY, Liu W, Long JR, Yip SP, Chan SY, 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.
15. Kolonel LN, Henderson BE, Hankin JH, Nomura AM, Wilkens LR, et al. (2000) A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. Am J Epidemiol 151: 346–357.
16. Hamajima N, Matsuo K, Saito T, Hirose K, Inoue M, et al. (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.
17. Itoh H, Iwasaki M, Hanaoka T, Kasuga Y, Yokoyama S, et al. (2009) Serum organochlorines and breast cancer risk in Japanese women: a case-control study. Cancer Causes Control 20: 567–580.
18. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38: 904–909.

Iwasaki et al. Breast Cancer Research 2010, 12:R88
http://breast-cancer-research.com/content/12/5/R88

Breast Cancer RESEARCH

RESEARCH ARTICLE Open Access

Green tea drinking and subsequent risk of breast cancer in a population to based cohort of Japanese women

Motoki Iwasaki[1*], Manami Inoue[1], Shizuka Sasazuki[1], Norie Sawada[1], Taiki Yamaji[1], Taichi Shimazu[1], Walter C Willett[2], Shoichiro Tsugane[1], for Japan Public Health Center-based Prospective Study Group[1]

Abstract

Introduction: Although many *in vitro* and animal studies have demonstrated a protective effect of green tea against breast cancer, findings from epidemiological studies have been inconsistent, and whether high green tea intake reduces the risk of breast cancer remains unclear.

Methods: In this Japan Public Health Center-based Prospective Study, 581 cases of breast cancer were newly diagnosed in 53,793 women during 13.6 years' follow-up from the baseline survey in 1990 to 1994. After the five-year follow-up survey in 1995 to 1998, 350 cases were newly diagnosed in 43,639 women during 9.5 years' follow-up. The baseline questionnaire assessed the frequency of total green tea drinking while the five-year follow-up questionnaire assessed that of two types of green tea, *Sencha* and *Bancha/Genmaicha*, separately.

Results: Compared with women who drank less than one cup of green tea per week, the adjusted hazard ratio (HR) for women who drank five or more cups per day was 1.12 (95% confidence interval (CI) 0.81 to 1.56; *P* for trend = 0.60) in the baseline data. Similarly, compared with women who drank less than one cup of *Sencha* or *Bancha/Genmaicha* per week, adjusted HRs for women who drank 10 or more cups per day were 1.02 (95% CI 0.55 to 1.89; *P* for trend = 0.48) for *Sencha* and 0.86 (0.34 to 2.17; *P* for trend = 0.66) for *Bancha/Genmaicha*. No inverse association was found regardless of hormone receptor-defined subtype or menopausal status.

Conclusions: In this population-based prospective cohort study in Japan we found no association between green tea drinking and risk of breast cancer.

Introduction

Green tea is regularly consumed in Japan and China as a traditional habit and cultural characteristic. Although produced from the same plant, *Camellia sinensis*, differences in the manufacturing process mean that green tea has a higher catechin content than black tea [1-3], which might contribute to its beneficial effects on cancer as well as cardiovascular diseases, and other conditions [3,4]. Specifically, (-)-epigallocatechin-3-gallate (EGCG), the most abundant and biologically active catechin in green tea, might play an important role in cancer prevention [1-3,5,6]. Because breast cancer risk is substantially lower in Asian than Western countries [7], a contribution of high green tea intake to low breast cancer risk has been hypothesized. This hypothesis has been supported by *in vitro* and animal studies, which have demonstrated various protective effects of green tea and tea polyphenols acting via strong antioxidant activity, inhibition of cell proliferation and angiogenesis, induction of apoptosis, and antiestrogenic properties [1-3,5,6,8].

In contrast to *in vitro* and animal studies, few epidemiological studies have examined the association between green tea intake and risk of breast cancer, and their findings have been inconsistent [9-16]. An inverse association was found in three case-control studies among Asian-American and Chinese populations [9-11], whereas no association was observed in two cohort

* Correspondence: moiwasak@ncc.go.jp
[1]Epidemiology and Prevention Division, Research Center for Cancer Prevention and Screening, National Cancer Center, 5 - 1 - 1 Tsukiji, Chuo - ku, Tokyo, 104 - 0045, Japan
Full list of author information is available at the end of the article

studies in Japan [12,13] or one nested case-control study in Singapore [14]. This inconsistency might be in part explained by differences in study design and exposure variation. In the three case-control studies which reported an inverse association, the reference groups were non-green tea drinkers, which included approximately 36 to 68% of control subjects [9-11], whereas in the two Japanese cohort studies the reference groups were women who drank less than one cup per day or one or fewer cups per day [12,13]. Inclusion of green tea drinkers in the reference group might attenuate the difference in risk between the reference and the higher drinking groups. Moreover, the highest consumption group in the two Japanese cohort studies were women who drank five or more cups per day, which included a relatively large portion of subjects (approximately 26 to 43%) [12,13]. Several studies have suggested a possible protective effect of very large amounts of green tea intake: one case-control study in Japan showed a decreased risk of gastric cancer among those who drank 10 or more cups per day [17], for example, while a cohort study in Japan observed a decreased risk of cancer in all sites among those who drank 10 or more cups per day [18]. However, a better understanding of the role of green tea in the etiology of breast cancer would be obtained from studies which included non- to heavy green tea drinkers. In addition, given that *in vitro* studies have suggested the inhibition of aromatase by green tea extracts [8] and of estrogen binding with its receptor by EGCG [6], the effect of green tea intake on the risk of breast cancer may differ according to hormone receptor to defined subtype. Nevertheless, no study has evaluated these associations.

To address these issues, we conducted a large-scale population-based prospective cohort study in Japan on the association between green tea intake and the risk of breast cancer.

Materials and methods

Study population

The Japan Public Health Center-based Prospective Study (JPHC Study), which began in 1990 for Cohort I and in 1993 for Cohort II, included 140,420 inhabitants (68,722 men and 71,698 women) in the municipalities supervised by 11 public health centers (PHC). Details of the study design have been described elsewhere [19]. Study participants were informed about the objectives and methods of the study in written form and those who responded to the survey questionnaire were regarded as consenting to participate in the study. In addition, participants were also notified in writing that they could withdraw from the study. Given no relevant ethical guideline and committee at the time of the survey, the institutional review board of the National Cancer Center, Tokyo, Japan considered return of the questionnaire to be classified as informed consent and approved the study protocol.

The study population comprised registered Japanese inhabitants living in each PHC area, aged 40 to 59 years in Cohort I and 40 to 69 years in Cohort II. In the present analysis, one PHC area was excluded because data on cancer incidence were not available. Thus, after exclusion of ineligible women (n = 98), we defined a population-based cohort of 67,422 women.

Baseline and five-year follow-up survey

Surveys of the cohort were conducted twice by self-administered questionnaire, the baseline in 1990 to 1994 and the five-year follow-up in 1995 to 1998. Of 67,422 women, a total of 55,886 women (83%) returned the baseline questionnaire. Of these, 1,510 who reported a history of cancer in the baseline survey were excluded, leaving 54,376 women as the baseline questionnaire respondents. Of the 67,422 women, we identified 62,788 as eligible for the five-year follow-up survey after exclusion of those who had died, moved out of the study area, or been lost to follow-up before the survey, of whom 52,485 (84%) returned the questionnaire. We also excluded 1,860 women who had a history of cancer at the five-year follow-up survey and 5,813 women who did not respond to the baseline questionnaire, leaving 44,812 women as the five-year follow-up questionnaire respondents.

Exposure measurements

Information on beverages, including green tea, oolong tea, black tea, and coffee, was obtained in the baseline questionnaire in terms of frequency and amount using six precoded categories: less than one cup per week, one to two cups per week, three to four cups per week, and almost daily (further divided into one to two cups per day, three to four cups per day, and five or more cups per day). In the five-year follow-up questionnaire, the consumption of beverages, including two items of green tea, namely *Sencha* (first or second flush of green tea; that is, the first seasonal picking) and *Bancha* (third or fourth flush of green tea; that is, the late seasonal picking)/*Genmaicha* (blend of *bancha* and roasted brown rice), oolong tea, black tea, coffee and canned coffee was assessed in terms of frequency and amount using nine precoded categories: less than one cup per week, one to two cups per week, three to four cups per week, five to six cups per week, one cup per day, two to three cups per day, four to six cups per day, seven to nine cups per day, and ten or more cups per day. *Sencha* and *Bancha* are the two main types of green tea consumed in Japan, and are usually prepared by steeping the tea leaves in hot water. The amounts of *Sencha* or *Bancha*/

Genmaicha consumed (ml per day) were computed by multiplying the frequency by the portion size for each beverage (120 ml per cup). Total green tea intake was defined as the sum of *Sencha* and *Bancha/Genmaicha* intake. The validity of green tea intake reported by the cohort was assessed using dietary records for 28 days (seven-day dietary records in four seasons) or 14 days. Spearman's correlation coefficients for green tea intake between the dietary record data and baseline questionnaire were 0.63 for Cohort I [20] and 0.43 for Cohort II (unpublished data).

Follow-up

All registered women were followed from the start of the study period to 31 December 2006. Data on residential relocation were obtained from residential registries. Among the baseline respondents (n = 54,376), 2,962 women (5.4%) moved out of the study area and 201 (0.4%) were lost to follow-up. The occurrence of cancer was identified by active patient notification from major local hospitals in the study area and data linkage with population-based cancer registries, with permission from the local governments responsible for the registries. Death certificates were used to supplement information on cancer incidence. Site of origin and histologic type were coded by members of our Study Group using the International Classification of Diseases for Oncology, Third Edition (ICD-O-3), code C500-509. Information on estrogen receptor (ER) and progesterone receptor (PR) status was collected from medical records or pathology reports. Up to the end of the study period, 586 new breast cancer cases were identified among the baseline questionnaire respondents (n = 54,376) and 362 cases among the five-year follow-up questionnaire respondents (n = 44,812). Diagnosis was microscopically verified in 95% of 586 cases, and based on death certificates only in 1.0%. Information on ER and PR status was available for 278 (47%) and 262 (45%) of 586 cases, respectively.

Statistical analysis

Since the assessment of green tea intake differed between the baseline and five-year follow-up questionnaires, these were analyzed separately. We excluded 583 women with incomplete information for green tea from 54,376 women, leaving a total of 53,793 women, including 581 breast cancer cases, for inclusion in the baseline data analyses. We also excluded 1,173 women with incomplete information for *Sencha* or *Bancha/Genmaicha* from 44,812 women, leaving a total of 43,639 women, including 350 breast cancer cases, for inclusion in the five-year follow-up data analyses.

In the baseline data analyses, person-years of follow-up were calculated from the baseline survey (1990 to 1994) until the date of diagnosis of breast cancer, date of relocation from the study area, date of death, or end of the study period (31 December 2006), whichever occurred first. Similarly, person-years of follow-up were calculated from the five-year follow-up survey (1995 to 1998) until the date of diagnosis of breast cancer, date of relocation from the study area, date of death, or end of the study period (31 December 2006), whichever occurred first for the five-year follow-up data analyses.

The Cox proportional hazards model was used to estimate hazard ratio (HR) and 95% confidence intervals (CI) of breast cancer by green tea intake using the SAS program (PROC PHREG) (SAS Institute Inc., Cary, NC, USA). The following variables were used for adjustment as potential confounders: age, study area, age at menarche, menopausal status at baseline and age at menopause, number of births, age at first birth, height, body mass index, alcohol intake, smoking status, physical activity, exogenous hormone use, family history of breast cancer, oolong tea intake, black tea intake, coffee intake and canned coffee intake. Linear trends for HRs were tested in the Cox proportional hazards models using the exposure categories as ordinal variables. All P-values reported are two-sided, and significance level was set at $P < 0.05$.

Results

During 733,667 person-years of follow-up (average follow-up, 13.6 years) for 53,793 women between 1990 to 1994 and 2006, a total of 581 cases of breast cancer were newly diagnosed and included in the baseline data analyses; while during 412,801 person-years of follow-up (average follow-up, 9.5 years) for 43,639 women between 1995 to 1998 and 2006, a total of 350 cases of breast cancer were newly diagnosed and included in the five-year follow-up data analyses.

Baseline characteristics of the study participants according to green tea intake are shown in Table 1. Approximately 12% of women drank green tea less than one cup per week while 27% drank five or more cups per day in the baseline data; and 22% and 30% of women did not drink *Sencha* and *Bancha/Genmaicha*, while 5.2% and 2.5% drank 10 or more cups per day, respectively, in the five-year follow-up data. Women who drank five or more cups per day in the baseline data tended to be older, live in towns or villages, be taller, have an earlier onset of first menstruation and fewer births, have a lower use of exogenous female hormones, drink less oolong tea and coffee, and consume more fruits and isoflavones. Similar characteristics was observed for women who drank 10 or more cups of *Sencha* per day in the five-year follow-up data, while women who drank 10 or more cups of *Bancha/Genmaicha* per day tended to be older, have later onset of first

Table 1 Baseline characteristics according to green tea intake*

	Green tea assessed by baseline questionnaire			Green tea assessed by five-year follow-up questionnaire, *Sencha*			Green tea assessed by five-year follow-up questionnaire, *Bancha/Genmaicha*		
	Less than one cup per week	One to two cups per day	Five or more cups per day	Less than one cup per week	Two to three cups per day	Ten or more cups per day	Less than one cup per week	Two to three cups per day	Ten or more cups per day
No. of subjects	6,202	11,322	14,308	9,638	9,052	2,281	13,053	8,148	1,099
Age (year), mean	50.4	50.8	53.9	58.2	57.0	58.9	57.5	57.8	59.7
Residential area (town or village), %	40.9	57.3	68.5	58.7	66.4	73.2	66.9	64.0	60.2
Family history of breast cancer, %	0.7	1.2	1.1	0.9	1.5	0.8	1.3	1.1	0.9
Premenopausal women, %	46	47	31	21	24	15	23	22	14
Age at menopause (year), mean†	49.2	49.3	49.4	-	-	-	-	-	-
Age at menopause (50 to 54 years old), %	-	-	-	46	50	47	47	49	46
Age at menarche (year), mean†	15.0	14.6	14.7	15.0	14.6	14.7	14.7	14.7	14.9
Number of births, mean†	2.9	2.7	2.7	3.0	2.6	2.7	2.7	2.7	2.9
Age at first birth (year), mean†	24.8	25.0	24.9	24.6	25.0	25.0	25.0	24.8	24.6
Use of exogenous female hormones (current use), %	1.4	1.0	1.0	2.9	2.4	2.1	2.4	2.3	2.3
Height (cm), mean†	150.9	152.0	152.2	151.1	152.0	152.3	151.8	151.7	151.9
Body mass index (kg/m2), mean†	23.6	23.3	23.4	23.7	23.3	23.4	23.4	23.4	23.6
Smoking (current smoker), %	7.9	6.5	7.0	5.6	4.4	8.0	6.1	4.1	8.4
Alcohol drinking (regular drinker), %	11	14	11	9.7	13	13	13	11	12
Leisure-time physical activity (≥ once per week), %	15	18	18	-	-	-	-	-	-
Physical activity (metabolic equivalent to hours per day), mean†	-	-	-	32.0	31.9	32.2	31.8	32.1	32.4
Vitamin supplement user, %	20	19	20	12	16	15	15	15	13
Bancha/Genmaicha intake (≥ four cups per day), %	-	-	-	30	6.2	25	-	-	-
Sencha intake (≥ four cups per day), %	-	-	-	-	-	-	43	14	44
Oolong tea intake (≥ one cup per day), %	20	17	10	11	12	8.8	11	13	11
Black tea intake (≥ one cup per day), %	2.8	5.0	3.2	2.6	5.2	3.8	3.3	5.0	4.4
Coffee intake (≥ one cup per day), %	41	49	28	36	40	22	34	40	27
Canned coffee intake (≥ one cup per day), %	-	-	-	4.3	2.7	2.2	2.8	3.2	2.8
Total energy intake (kcal/day), mean†‡	1,770.0	1,764.0	1,805.2	1,770.1	1,923.5	2,194.0	1,833.9	1,934.6	2,185.8
Fish and shellfish intake (g/day), mean†‡	105.5	102.8	105.4	79.6	97.5	114.3	88.2	97.0	105.3
Meats intake (g/day), mean†‡	67.2	65.2	66.1	59.7	60.4	67.7	56.6	61.6	71.6

Table 1 Baseline characteristics according to green tea intake* *(Continued)*

Vegetable intake (g/day), mean†‡	293.8	293.7	295.7	217.0	247.7	319.9	234.9	245.8	327.0
Fruit intake (g/day), mean†‡	175.9	177.3	180.8	208.5	269.9	359.3	243.1	270.1	323.7
Isoflavone intake (mg/day), mean†‡	31.6	32.2	33.5	37.5	43.3	54.7	41.5	42.0	54.8

* Three categories were chosen from six or nine precoded categories, respectively: less than one cup per week, one to two cups per week, three to four cups per week, one to two cups per day, three to four cups per day, and five or more cups per day for baseline questionnaire, and less than one cup per week, one to two cups per week, three to four cups per week, five to six cups per week, one cup per day, two to three cups per day, four to six cups per day, seven to nine cups per day, and ten or more cups per day for five to year follow-up questionnaire.

†Adjusted for age.

‡ Intake for each subject was estimated from the food frequency questionnaires based on a regression function derived from the validation study data (baseline questionnaire only).

menstruation, earlier age at first birth and a larger number of births.

We found no inverse association between green tea intake and the risk of breast cancer regardless of green tea intake as assessed by the baseline or five-year follow-up questionnaire (Table 2). Compared with women who drank less than one cup of green tea per week, the adjusted HR for women who drank five or more cups per day was 1.12 (95% CI 0.81 to 1.56; *P* for trend = 0.60) in the baseline data analyses. Similarly, compared with women who drank less than one cup of *Sencha* or *Bancha/Genmaicha* per week, adjusted HRs for women who drank 10 or more cups per day in the five-year data analyses were 1.02 (95% CI 0.55 to 1.89; *P* for trend = 0.48) for *Sencha* and 0.86 (0.34 to 2.17; *P* for trend = 0.66) for *Bancha/Genmaicha*. Moreover, compared with women who drank neither *Sencha* nor *Bancha/Genmaicha*, the adjusted HR for women who drank more than 1,320 ml per day was 1.29 (0.60 to 2.79; *P* for trend = 0.70). No substantial change was seen after further adjustment for other potential confounders such as residential area or dietary intake of meat, fish, vegetables, fruit, energy and isoflavones (data not shown). Further, no substantial change was seen in either baseline or five-year follow-up data after the exclusion of women within the first five years of follow-up to minimize the influence of existing preclinical conditions (data not shown), or after the exclusion of women who drank more than one cup of oolong tea or black tea per week to prevent the inclusion of other tea drinkers into the green tea intake reference category (data not shown).

To assess the potential influence of changes in green tea intake during the follow-up period on our findings, we first categorized subjects into three consumption groups for each survey: non-drinkers (corresponding to 'less than one cup per week'), 1 to 719 ml per day, and more than 720 ml per day (corresponding to 'more than five cup per day' in the baseline questionnaire). We next categorized subjects into three groups based on the combination of green tea intake calculated from the both the baseline and five-year follow-up questionnaires: non-drinkers for both questionnaires, more than 720 ml per day for both questionnaires, and other combinations. Compared with women who did not drink green tea for both questionnaires, adjusted HR was 1.12 (95% CI 0.45 to 2.83) for women who drank more than 720 ml per day for both questionnaires.

For analysis of cases by hormone receptor status and among women grouped by menopausal status, we re-categorized women into four groups for the baseline data and three for the five-year follow-up data. In the baseline data analyses, we found no inverse association between green tea intake and the risk of breast cancer regardless of hormone receptor-defined subtype (Table 3). Stratified analyses according to baseline menopausal status showed no remarkable difference between strata. Similar results were obtained when we analyzed the five-year follow-up data (data not shown). Additional stratified analyses according to dietary intake observed no remarkable difference between subgroups defined by dietary isoflavone and folate intake for either the baseline or five-year follow-up data (data not shown).

In additional analyses to investigate the associations of oolong tea, black tea, and coffee intake with breast cancer risk, we found no inverse associations with the baseline (Table 4) or five-year follow-up data (data not shown).

Discussion

In this population-based prospective cohort study, we found no overall association between green tea intake and the risk of breast cancer among Japanese women regardless of menopausal status. Our findings are in general agreement with those of three prospective studies, including two Japanese cohort studies, which found no association between green tea intake and breast cancer risk [12-14]. One noteworthy strength of the present over previous studies is the our remarkably wide variation in green tea intake, from women who drank green

Table 2 Hazard ratio and 95% confidence interval of breast cancer according to green tea intake

	Green tea intake assessed by baseline questionnaire												
	Less than one cup per week	One to two cups per week		Three to four cups per week		One to two cup per day		Three to four cups per day		Five or more cups per day			*P* for trend
No. of cases	68	50		36		117		160		150			
Person-years	87,841	56,200		40,912		152,896		201,005		194,813			
Age- and area-adjusted HR (95% CI)	1.00	1.18	(0.82 to 1.71)	1.17	(0.78 to 1.75)	1.11	(0.82 to 1.50)	1.18	(0.88 to 1.59)	1.12	(0.83 to 1.51)		0.58
Multivariate HR (95% CI)*	1.00	1.19	(0.80 to 1.76)	1.13	(0.72 to 1.75)	1.13	(0.81 to 1.58)	1.17	(0.85 to 1.62)	1.12	(0.81 to 1.56)		0.60
	Green tea intake assessed by five-year follow-up questionnaire												
	Less than one cup per week	One to six cups per week		One cup per day		Two to three cups per day		Four to six cups per day		Seven to nine cups per day		Ten or more cups per day	P for trend
Sencha													
No. of cases	82	56		35		63		78		18		18	
Person-years	91,228	75,237		38,328		85,491		73,160		27,641		21,715	
Age- and area-adjusted HR (95% CI)	1.00	0.79	(0.56 to 1.12)	1.02	(0.68 to 1.52)	0.83	(0.59 to 1.16)	1.22	(0.88 to 1.69)	0.74	(0.44 to 1.26)	0.92 (0.55 to 1.55)	0.77
Multivariate HR (95% CI)†	1.00	0.71	(0.46 to 1.08)	0.97	(0.59 to 1.58)	0.85	(0.57 to 1.27)	1.25	(0.84 to 1.86)	0.68	(0.36 to 1.31)	1.02 (0.55 to 1.89)	0.48
Bancha/Genmaicha													
No. of cases	107	90		38		69		30		7		9	
Person-years	123,189	99,201		42,254		76,532		46,563		14,733		10,329	
Age- and area-adjusted HR (95% CI)	1.00	1.02	(0.77 to 1.36)	1.05	(0.72 to 1.53)	1.09	(0.80 to 1.49)	0.80	(0.53 to 1.21)	0.58	(0.27 to 1.26)	1.02 (0.52 to 2.02)	0.41
Multivariate HR (95% CI)†	1.00	1.12	(0.80 to 1.58)	1.23	(0.79 to 1.92)	1.23	(0.85 to 1.78)	0.83	(0.51 to 1.35)	0.70	(0.28 to 1.76)	0.86 (0.34 to 2.17)	0.66
	Total green tea intake assessed by five-year follow-up questionnaire‡												
	0	1 to 119 ml per day		120 to 299 ml per day		300 to 599 ml per day		600 to 959 ml per day		960 to 1319 ml per day		1320 ml + per day	*P* for trend
No. of cases	13	44		48		85		105		28		27	
Person-years	18,312	50,715		58,499		91,109		114,255		44,745		35,166	
Age- and area-adjusted HR (95% CI)	1.00	1.25	(0.67 to 2.32)	1.25	(0.67 to 2.32)	1.47	(0.81 to 2.67)	1.48	(0.82 to 2.67)	1.02	(0.52 to 2.00)	1.20 (0.61 to 2.36)	0.88
Multivariate HR (95% CI)§	1.00	1.08	(0.52 to 2.23)	1.26	(0.62 to 2.56)	1.59	(0.81 to 3.14)	1.43	(0.73 to 2.83)	0.89	(0.40 to 1.97)	1.29 (0.60 to 2.79)	0.70

,†,§ Adjusted for age (continuous), area (10 public health centers), age at menarche (continuous), menopausal status at baseline (premenopausal women, age at menopause for postmenopausal women (- 47, 48 to 50, 51 to 53, 54+; †,§ - 44, 45 to 49, 50 to 54, 55+)), number of births (0, 1, 2, 3, 4, 5+), age at first birth (-21, 22 to 25, 26 to 29, 30+, nulliparous), height (continuous), BMI (continuous), alcohol intake (non-drinkers, occasional drinkers, -150 (g per week) and 150+ (g per week) among regular drinkers (ethanol)), smoking status (never, past, current), *leisure time physical activity (no, one to three days per month, more than one day per week), †,§ daily physical activity (metabolic equivalent-hours per day) (continuous), exogenous hormone use (never, past, current), family history of breast cancer (yes, no), oolong tea intake (*less than one cup per week, one to four cups per week, one or more cups per day; †,§ less than 1 cup per week, one to six cups per week, one or more cups per day), black tea intake (*less than one cup per week, one to four cups per week, one or more cups per day; †,§ less than one cup per week, one to six cups per week, one or more cups per day), coffee intake (*less than one cup per week, one to four cups per week, one to two cups per day, three or more cups per day; †,§ less than one cup per week, one to six cups per week, one or more cups per day), †,§ canned coffee intake (less than one cup per week, one to six cups per week, one or more cups per day), and †Sencha and Bancha/Genmaicha intake.

Models for Sencha or Bancha/Genmaicha intake did not include these variables, respectively.

‡ Total green tea intake was defined as the sum of Sencha and Bancha/Genmaicha intake (ml per day).

Table 3 Hazard ratio and 95% confidence interval of breast cancer according to subgroup analyses

	Green tea intake assessed by baseline questionnaire								
	Less than one cup per day	One to two cups per day		Three to four cups per day		Five or more cups per day		P for trend
All subjects								
No. of cases	154	117		160		150		
Age- and area-adjusted HR (95% CI)	1.00	1.01	(0.79 to 1.29)	1.08	(0.86 to 1.36)	1.02	(0.81 to 1.30)	0.75
Multivariate HR (95% CI)*	1.00	1.04	(0.80 to 1.35)	1.08	(0.84 to 1.39)	1.03	(0.80 to 1.34)	0.76
ER+PR+ breast cancer								
No. of cases	29	25		32		32		
Age- and area-adjusted HR (95% CI)	1.00	1.25	(0.73 to 2.17)	1.18	(0.69 to 2.00)	1.05	(0.62 to 1.79)	0.93
Multivariate HR (95% CI)*	1.00	1.33	(0.75 to 2.38)	1.35	(0.77 to 2.36)	1.02	(0.57 to 1.83)	0.92
ER- PR- breast cancer								
No. of cases	23	16		20		16		
Age- and area-adjusted HR (95% CI)	1.00	1.10	(0.58 to 2.11)	1.09	(0.58 to 2.05)	0.82	(0.42 to 1.60)	0.61
Multivariate HR (95% CI)*	1.00	1.13	(0.55 to 2.32)	1.23	(0.63 to 2.41)	0.81	(0.39 to 1.69)	0.69
Premenopausal women								
No. of cases	81	59		71		51		
Age- and area-adjusted HR (95% CI)	1.00	1.01	(0.72 to 1.43)	1.13	(0.81 to 1.58)	0.95	(0.66 to 1.37)	0.99
Multivariate HR (95% CI)*	1.00	1.05	(0.73 to 1.49)	1.12	(0.79 to 1.58)	0.97	(0.66 to 1.41)	0.99
Postmenopausal women								
No. of cases	70	56		86		96		
Age- and area-adjusted HR (95% CI)	1.00	1.01	(0.71 to 1.45)	1.04	(0.75 to 1.45)	1.05	(0.76 to 1.45)	0.76
Multivariate HR (95% CI)*	1.00	1.01	(0.67 to 1.50)	1.02	(0.71 to 1.48)	1.08	(0.75 to 1.55)	0.67

* Adjusted for age (continuous), area (10 public health centers), age at menarche (continuous), menopausal status at baseline (premenopausal women, age at menopause for postmenopausal women (-47, 48 to 50, 51 to 53, 54+)), number of births (0, 1, 2, 3, 4, 5+), age at first birth (-21, 22 to 25, 26 to 29, 30+, nulliparous), height (continuous), BMI (continuous), alcohol intake (non-drinkers, occasional drinkers, -150 (g/week) and 150+ (g/week) among regular drinkers (ethanol)), smoking status (never, past, current), leisure time physical activity (no, one to three days per month, more than one day per week), exogenous hormone use (never, past, current), family history of breast cancer (yes, no), oolong tea intake (less than one cup per week, one to four cups per week, one or more cups per day), black tea intake (less than one cup per week, one to four cups per week, one or more cups per day) and coffee intake (less than one cup per week, one to four cups per week, one to two cups per day, three or more cups per day).

tea less than one cup per week to those who drank 10 or more cups per day. This strength argues against the possibility that the observed absence of associations with breast cancer risk is attributable to insufficient variation in green tea intake. Our findings therefore suggest that green tea intake within a usual drinking habit is unlikely to reduce the risk of breast cancer.

The other major strength of the present study was its prospective design, in which information was collected before the subsequent diagnosis of breast cancer, thereby avoiding the exposure recall bias inherent to case-control studies. Subjects were selected from the general population, the sample was large, the response rate to the questionnaire (more than 80%) was acceptable for study settings such as this, and the loss to follow-up (0.4%) was negligible. Furthermore, the cancer registry in the study population was of sufficient quality to reduce the possibility of misclassification of the outcome.

Several limitations of this study warrant mention. First, since green tea intake was assessed by self-administered questionnaire, misclassification may have been unavoidable, albeit that our validation study showed relatively high validity [20]. Changes in green tea intake during the follow-up period may also have caused misclassification. To assess the potential influence of this misclassification, we calculated HRs for women who drank more than 720 ml per day in both the baseline and five-year follow-up questionnaires versus those who did not drink green tea in either questionnaire. Although we found no association, we cannot deny the possibility that changes in green tea intake during the follow-up period influenced the findings, particularly considering that the questionnaires used in the baseline and five-year follow-up surveys were different. These misclassifications due to inaccurate measurement would in turn have attenuated the true association, which might be one reason for our results. However, we previously found an inverse association between green tea intake and the risk of distal gastric cancer among women [21], and between green tea intake and the risk of advance prostate cancer among men in the JPHC study using the same analytic approach, namely that exposure status was not updated during follow-up [22]. These findings would also argue against the possibility that the observed absence of associations with breast cancer risk was attributable to inaccurate measurement of green tea intake.

Table 4 Hazard ratio and 95% confidence interval of breast cancer according to oolong tea, black tea and coffee intake as assessed by baseline survey

	Oolong tea intake						
	Less than one cup per week	One to four cups per week		One or more cups per day			P for trend
No. of cases	336	153		76			
Person-years	454,964	148,821		93,630			
Age- and area-adjusted HR (95% CI)	1.00	1.37	(1.13 to 1.67)	1.09	(0.85 to 1.41)		0.08
Multivariate HR (95% CI)*	1.00	1.34	(1.09 to 1.66)	0.98	(0.74 to 1.30)		0.40
	Black tea intake						
	Less than one cup per week	One to four cups per week		One or more cups per day			P for trend
No. of cases	441	111		24			
Person-years	557,038	140,312		24,776			
Age- and area-adjusted HR (95% CI)	1.00	1.00	(0.81 to 1.24)	1.29	(0.85 to 1.96)		0.45
Multivariate HR (95% CI)*	1.00	0.84	(0.67 to 1.07)	1.30	(0.84 to 2.02)		0.80
	Coffee intake						
	Less than one cup per week	One to four cups per week		One to two cups per day		Three or more cups per day	P for trend
No. of cases	161	180		173		63	
Person-years	233,697	213,979		210,969		69,940	
Age- and area-adjusted HR (95% CI)	1.00	1.20	(0.97 to 1.49)	1.18	(0.94 to 1.48)	1.30 (0.95 to 1.77)	0.09
Multivariate HR (95% CI)*	1.00	1.15	(0.91 to 1.46)	1.12	(0.87 to 1.43)	1.22 (0.87 to 1.71)	0.26

* Adjusted for age (continuous), area (10 public health centers), age at menarche (continuous), menopausal status at baseline (premenopausal women, age at menopause for postmenopausal women (-47, 48 to 50, 51 to 53, 54+)), number of births (0, 1, 2, 3, 4, 5+), age at first birth (-21, 22 to 25, 26 to 29, 30+, nulliparous), height (continuous), BMI (continuous), alcohol intake (non-drinkers, occasional drinkers, -150 (g/week) and 150+ (g/week) among regular drinkers (ethanol)), smoking status (never, past, current), leisure time physical activity (no, one to three days per month, more than one day per week), exogenous hormone use (never, past, current), family history of breast cancer (yes, no), green tea intake (less than one cup per day, one to two cups per day, three to four cups per day, five or more cups per day), oolong tea intake (less than one cup per week, one to four cups per week, one or more cups per day), black tea intake (less than one cup per week, one to four cups per week, one or more cups per day) and coffee intake (less than one cup per week, one to four cups per week, one to two cups per day, three or more cups per day). Models for oolong tea, black tea, or coffee intake did not include these variables, respectively.

Second, in spite of a reasonably large cohort population (53,793 women) and long follow-up period (average 13.6 year), the number of breast cancer cases was relatively low (n = 581) in the baseline data analyses, reflecting the low incidence rate in Japan (age-standardized rate per 100,000 world population in 2002, 32.7 in Japan and 101.1 in United States for comparison) [23]. Although we found no association between green tea intake and breast cancer risk (HR for five or more cups per day versus less than one cup per week = 1.12 in the baseline data analyses), the 95% CI was 0.81 to 1.56, which was a relatively wide. Our relatively small sample size therefore cannot deny the possibility of a smaller increase or decrease in risk.

Third, although we measured and adjusted for several potential confounders in the statistical model as far as possible, the effects of confounding by unmeasured variables and residual confounding cannot be totally discarded.

Fourth, information on hormone receptor status was available for 45 to 47% cases in the present study. The major reason for the unknown cases is that collection of this information began in 2002, while data for 1990 to 2002 were obtained by retrospective review of medical records or pathology reports. The relatively small number of cases weakened the statistical power to detect the association. In addition, despite the generally high agreement between the enzyme-linked immunoassay and immunohistochemical techniques, differences in classification as well as interlaboratory variation may result in misclassification, which, if present, may have also attenuated the true association. These methodological limitations might partly explain our findings, which showed no inverse association between green tea intake and the risk of breast cancer regardless of hormone receptor-defined subtype. Thus, further research to clarify the differential associations by hormone receptor-defined subtype is warranted.

Our findings contradict those of three case-control studies, which showed an inverse association [9-11]. Given that all three previous prospective studies showed no association [12-14], these findings might have been influenced by recall and selection bias stemming from the case-control design. Meanwhile, for the three

case-control studies conducted outside Japan [9-11], inconsistent findings might be in part explained by differences in the type of green tea and drinking method. In general, green tea contains catechins, minerals, and vitamins, and their contents vary among green tea types: *Sencha*, for example, one of most popular green teas in Japan, contains higher levels of tannin, vitamin C and folate than *Bancha/Genmaicha* [24]. The present study, however, found no overall association between green tea intake and the risk of breast cancer regardless of type. Moreover, levels of tea polyphenol and other nutrients in green tea varies according to preparation, amount of green tea leaves, frequency of renewing a tea batch in the pot, water temperature, brewing time, and so on. Although two Chinese case-control studies took account of several of these conditions, including green tea leaf amount, brew strength, and tea batch renewal frequency [9,10], no study has directly measured prediagnostic biomarkers of tea polyphenols. In this regard, we conducted a nested cased-control study within the JPHC study and found no overall association between plasma tea polyphenols and the risk of breast cancer [25]. Differences in tea polyphenols level due to green tea type and drinking method are therefore unlikely to explain these inconsistent findings, in any major way at least.

Alternatively, the inconsistencies reported in previous studies might be explained by possible effect modification by dietary factors and genetic polymorphisms [11,14,26]. In a case-control study in Asian-Americans, a risk-reducing effect of green tea was primarily observed among subjects whose soy intake was low [11], while a nested case-control study in Singapore showed a protective effect of green tea against breast cancer among women with high-activity genotypes of the *methylenetetrahydrofolate reductase* (*MTHFR*) and *thymidylate synthase* (*TYMS*) genes [14]. This association was stronger among those whose dietary folate intake was low. Subsequent studies, including our present study, however, have failed to replicate effect modification by soy and folate intake [9,13]. On the other hand, given a lack of information on genetic polymorphisms, the possibility cannot be excluded that green tea drinking is protective among subgroups of women.

Moreover, a recent cohort study in Shanghai showed a time-dependent interaction between green tea consumption and age of breast cancer onset: HRs for women who started green tea drinking at 25 years of age or younger versus non-green tea drinkers were 0.69 for premenopausal breast cancer and 1.61 for postmenopausal breast cancer [16]. This complex observation might also provide an explanation for the inconsistencies reported in previous studies. However, the present study was not able to examine time-dependent interactions because of a lack of information on green tea drinking habits, such as age at the beginning of green tea drinking and years of drinking.

Although oolong tea, black tea, and coffee are less popular than green tea in Japan, they are also rich in phenolic compounds and have been hypothesized to have a protective effect against the development of breast cancer: oolong tea and black tea contain higher levels of theaflavins and thearubigins than catechins, and coffee contains a high level of chlorogenic acid [1-3]. We also found no association of oolong tea, black tea, coffee and canned coffee intake with breast cancer risk. In addition to the relatively small number of women who drank oolong tea, black tea, and coffee, the potential influence of green tea intake on our findings was not fully excluded, albeit that we adjusted for green tea intake. Our findings for black tea, however, are consistent with those of a previous meta-analysis [27]. In contrast, our findings for coffee intake contradict a recent meta-analysis, which suggested a small risk-reducing effect of coffee intake on breast cancer: for example, relative risk (95% CI) per increment of two cups per day was 0.98 (0.96 to 1.00) based on 18 studies [28]. In addition to this small risk reduction, the small variation in coffee intake in our participants might have contributed in part to our findings.

Conclusions

In this population-based prospective cohort study, we found no overall association between green tea intake and the risk of breast cancer among Japanese women. Our findings suggest that drinking green tea as a beverage is unlikely to reduce the risk of breast cancer regardless of green tea type and number of cups within a usual drinking habit.

Abbreviations

CI: confidence interval; EGCG: (-)-epigallocatechin-3-gallate; ER: estrogen receptor; HR: hazard ratio; JPHC Study: Japan Public Health Center-based Prospective Study; PHC: public health centers; PR: progesterone receptor.

Acknowledgements

This study was supported by Grants-in-Aid for Cancer Research and for the Third Term Comprehensive Ten-Year Strategy for Cancer Control from the Ministry of Health, Labour and Welfare of Japan, and Grants-in-Aid for Scientific Research on Priority Areas (17015049) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

We wish to thank all staff members in each study area and in the central offices for their cooperation and technical assistance. We also wish to thank the Iwate, Aomori, Ibaraki, Niigata, Osaka, Kochi, Nagasaki and Okinawa Cancer Registries for their provision of incidence data.

Study group members: Members of the JPHC Study Group (principal investigator: S. Tsugane): S. Tsugane, M. Inoue, T. Sobue, and T. Hanaoka, Research Center for Cancer Prevention and Screening, National Cancer Center, Tokyo; J. Ogata, S. Baba, T. Mannami, A. Okayama, and Y. Kokubo, National Cardiovascular Center, Suita; K. Miyakawa, F. Saito, A. Koizumi, Y. Sano, I. Hashimoto, T. Ikuta, and Y. Tanaba, Iwate Prefectural Ninohe Public Health Center, Ninohe; Y. Miyajima, N. Suzuki, S. Nagasawa, Y. Furusugi, and N. Nagai, Akita Prefectural Yokote Public Health Center, Yokote; H. Sanada, Y. Hatayama, F. Kobayashi, H. Uchino, Y. Shirai, T. Kondo, R. Sasaki, Y. Watanabe, Y. Miyagawa, Y. Kobayashi, and M. Machida, Nagano Prefectural Saku Public

Health Center, Saku; Y. Kishimoto, E. Takara, T. Fukuyama, M. Kinjo, M. Irei, and H. Sakiyama, Okinawa Prefectural Chubu Public Health Center, Okinawa; K. Imoto, H. Yazawa, T. Seo, A. Seiko, F. Ito, F. Shoji, and R. Saito, Katsushika Public Health Center, Tokyo; A. Murata, K. Minato, K. Motegi, and T. Fujieda, Ibaraki Prefectural Mito Public Health Center, Mito; T. Abe, M. Katagiri, M. Suzuki, and K. Matsui, Niigata Prefectural Kashiwazaki and Nagaoka Public Health Center, Kashiwazaki and Nagaoka; M. Doi, A. Terao, Y. Ishikawa, and T. Tagami, Kochi Prefectural Chuo-higashi Public Health Center, Tosayamada; H. Doi, M. Urata, N. Okamoto, F. Ide, and H. Sueta, Nagasaki Prefectural Kamigoto Public Health Center, Arikawa; H. Sakiyama, N. Onga, H. Takaesu, and M. Uehara, Okinawa Prefectural Miyako Public Health Center, Hirara; F. Horii, I. Asano, H. Yamaguchi, K. Aoki, S. Maruyama, M. Ichii, and M. Takano, Osaka Prefectural Suita Public Health Center, Suita; S. Matsushima and S. Natsukawa, Saku General Hospital, Usuda; M. Akabane, Tokyo University of Agriculture, Tokyo; M. Konishi, K. Okada, and I. Saito, Ehime University, Toon; H. Iso, Osaka University, Suita; Y. Honda, K. Yamagishi, S. Sakurai, and N. Tsuchiya, Tsukuba University, Tsukuba; H. Sugimura, Hamamatsu University, Hamamatsu; Y. Tsubono, Tohoku University, Sendai; the late M. Kabuto, National Institute for Environmental Studies, Tsukuba; S. Tominaga, Aichi Cancer Center Research Institute, Nagoya; M. Iida W. Ajiki, and A. Ioka, Osaka Medical Center for Cancer and Cardiovascular Disease, Osaka; S. Sato, Osaka Medical Center for Health Science and Promotion, Osaka; N. Yasuda, Kochi University, Nankoku; K. Nakamura, Niigata University, Niigata; S. Kono, Kyushu University, Fukuoka; K. Suzuki, Research Institute for Brain and Blood Vessels Akita, Akita; the late Y. Takashima and M. Yoshida, Kyorin University, Mitaka; E. Maruyama, Kobe University, Kobe; the late M. Yamaguchi, Y. Matsumura, S. Sasaki, and S. Watanabe, National Institute of Health and Nutrition, Tokyo; T. Kadowaki, Tokyo University, Tokyo; M. Noda and T. Mizoue, International Medical Center of Japan, Tokyo; Y. Kawaguchi, Tokyo Medical and Dental University, Tokyo; H. Shimizu, Sakihae Institute, Gifu.

Author details
[1]Epidemiology and Prevention Division, Research Center for Cancer Prevention and Screening, National Cancer Center, 5 - 1 - 1 Tsukiji, Chuo - ku, Tokyo, 104 - 0045, Japan. [2]Departments of Nutrition and Epidemiology, Harvard School of Public Health, and Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 665 Huntington Avenue, Boston, MA 02115, USA.

Authors' contributions
MoI, Mal, SS, NS, TY, TS and ST were involved with the study concept and design. MoI, Mal, SS, NS, TY, TS and ST participated in the acquisition of data. MoI, Mal, SS, NS, TY, TS, WW and ST contributed to the analyses and interpretation of data. MoI conducted the statistical analyses and wrote the manuscript. All authors participated in the interpretation of results and critical revision of the manuscript for important intellectual content. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Received: 21 May 2010 Revised: 12 August 2010
Accepted: 28 October 2010 Published: 28 October 2010

References

1. Wang Y, Ho CT: **Polyphenolic chemistry of tea and coffee: a century of progress.** *J Agric Food Chem* 2009, **57**:8109-8114.
2. Yang CS, Lambert JD, Sang S: **Antioxidative and anti-carcinogenic activities of tea polyphenols.** *Arch Toxicol* 2009, **83**:11-21.
3. Cabrera C, Artacho R, Gimenez R: **Beneficial effects of green tea - a review.** *J Am Coll Nutr* 2006, **25**:79-99.
4. Schneider C, Segre T: **Green tea: potential health benefits.** *Am Fam Physician* 2009, **79**:591-594.
5. Yang CS, Maliakal P, Meng X: **Inhibition of carcinogenesis by tea.** *Annu Rev Pharmacol Toxicol* 2002, **42**:25-54.
6. Komori A, Yatsunami J, Okabe S, Abe S, Hara K, Suganuma M, Kim SJ, Fujiki H: **Anticarcinogenic activity of green tea polyphenols.** *Jpn J Clin Oncol* 1993, **23**:186-190.
7. Curado MP, Edwards B, Shin HR, Storm H, Ferlay J, Heanue M, Boyle P: **Cancer Incidence in Five Continents.** In *IARC Scientific Publications No. 160. Volume IX.* Lyon: IARC; 2007.
8. Satoh K, Sakamoto Y, Ogata A, Nagai F, Mikuriya H, Numazawa M, Yamada K, Aoki N: **Inhibition of aromatase activity by green tea extract catechins and their endocrinological effects of oral administration in rats.** *Food Chem Toxicol* 2002, **40**:925-933.
9. Shrubsole MJ, Lu W, Chen Z, Shu XO, Zheng Y, Dai Q, Cai Q, Gu K, Ruan ZX, Gao YT, Zheng W: **Drinking green tea modestly reduces breast cancer risk.** *J Nutr* 2009, **139**:310-316.
10. Zhang M, Holman CD, Huang JP, Xie X: **Green tea and the prevention of breast cancer: a case-control study in Southeast China.** *Carcinogenesis* 2007, **28**:1074-1078.
11. Wu AH, Yu MC, Tseng CC, Hankin J, Pike MC: **Green tea and risk of breast cancer in Asian Americans.** *Int J Cancer* 2003, **106**:574-579.
12. Nagano J, Kono S, Preston DL, Mabuchi K: **A prospective study of green tea consumption and cancer incidence, Hiroshima and Nagasaki (Japan).** *Cancer Causes Control* 2001, **12**:501-508.
13. Suzuki Y, Tsubono Y, Nakaya N, Suzuki Y, Koizumi Y, Tsuji I: **Green tea and the risk of breast cancer: pooled analysis of two prospective studies in Japan.** *Br J Cancer* 2004, **90**:1361-1363.
14. Inoue M, Robien K, Wang R, Van Den Berg DJ, Koh WP, Yu MC: **Green tea intake, MTHFR/TYMS genotype and breast cancer risk: the Singapore Chinese Health Study.** *Carcinogenesis* 2008, **29**:1967-1972.
15. Ogunleye AA, Xue F, Michels KB: **Green tea consumption and breast cancer risk or recurrence: a meta-analysis.** *Breast Cancer Res Treat* 2010, **119**:477-484.
16. Dai Q, Shu XO, Li H, Yang G, Shrubsole MJ, Cai H, Ji B, Wen W, Franke A, Gao YT, Zheng W: **Is green tea drinking associated with a later onset of breast cancer?** *Ann Epidemiol* 2010, **20**:74-81.
17. Kono S, Ikeda M, Tokudome S, Kuratsune M: **A case-control study of gastric cancer and diet in northern Kyushu, Japan.** *Jpn J Cancer Res* 1988, **79**:1067-1074.
18. Imai K, Suga K, Nakachi K: **Cancer-preventive effects of drinking green tea among a Japanese population.** *Prev Med* 1997, **26**:769-775.
19. Watanabe S, Tsugane S, Sobue T, Konishi M, Baba S: **Study design and organization of the JPHC study.** *J Epidemiol* 2001, **11**:S3-7.
20. Tsubono Y, Kobayashi M, Sasaki S, Tsugane S: **Validity and reproducibility of a self-administered food frequency questionnaire used in the baseline survey of the JPHC Study Cohort I.** *J Epidemiol* 2003, **13**:S125-133.
21. Sasazuki S, Inoue M, Hanaoka T, Yamamoto S, Sobue T, Tsugane S: **Green tea consumption and subsequent risk of gastric cancer by subsite: the JPHC Study.** *Cancer Causes Control* 2004, **15**:483-491.
22. Kurahashi N, Sasazuki S, Iwasaki M, Inoue M, Tsugane S: **Green tea consumption and prostate cancer risk in Japanese men: a prospective study.** *Am J Epidemiol* 2008, **167**:71-77.
23. Ferlay J, Bray F, Pisani P, Parkin DM: **GLOBOCAN 2002: Cancer Incidence, Mortality and Prevalence Worldwide, IARC CancerBase No. 5, version 2.0.** Lyon: IARCPress; 2004.
24. The Council for Science and Technology, Ministry of Education, Culture, Sports, Science and Technology, Japan: *Standard Tables of Food Composition in Japan.* Fifth revised and enlarged edition. Tokyo: National Printing Bureau; 2005.
25. Iwasaki M, Inoue M, Sasazuki S, Miura T, Sawada N, Yamaji T, Shimazu T, Willett WC, Tsugane S: **Plasma tea polyphenol levels and subsequent risk of breast cancer among Japanese women: a nested case-control study.** *Breast Cancer Res Treat* 2010, **124**:827-834.
26. Yuan JM, Koh WP, Sun CL, Lee HP, Yu MC: **Green tea intake, ACE gene polymorphism and breast cancer risk among Chinese women in Singapore.** *Carcinogenesis* 2005, **26**:1389-1394.
27. Sun CL, Yuan JM, Koh WP, Yu MC: **Green tea, black tea and breast cancer risk: a meta-analysis of epidemiological studies.** *Carcinogenesis* 2006, **27**:1310-1315.
28. Tang N, Zhou B, Wang B, Yu R: **Coffee consumption and risk of breast cancer: a metaanalysis.** *Am J Obstet Gynecol* 2009, **200**:290.

doi:10.1186/bcr2756
Cite this article as: Iwasaki *et al.*: Green tea drinking and subsequent risk of breast cancer in a population to based cohort of Japanese women. *Breast Cancer Research* 2010 12:R88.

Breast Cancer Res Treat (2010) 124:827–834
DOI 10.1007/s10549-010-0916-x

EPIDEMIOLOGY

Plasma tea polyphenol levels and subsequent risk of breast cancer among Japanese women: a nested case–control study

Motoki Iwasaki · Manami Inoue · Shizuka Sasazuki · Tsutomu Miura · Norie Sawada · Taiki Yamaji · Taichi Shimazu · Walter C. Willett · Shoichiro Tsugane

Received: 20 April 2010 / Accepted: 22 April 2010 / Published online: 4 May 2010

Abstract Although many in vitro and animal studies have suggested a protective effect of green tea against breast cancer, findings from epidemiological studies have been inconsistent. No study has used prediagnostic biomarkers of tea polyphenols, which might play a protective role. A total of 24,226 women aged 40 to 69 years in the Japan Public Health Center-based Prospective Study who responded to the baseline questionnaire and provided blood in 1990–1995 were followed to December 2002. During a mean 10.6 years of follow-up, 144 newly diagnosed breast cancers were identified. Two matched controls for each case were selected from the cohort. Plasma levels of (-)-epigallocatechin (EGC), (-)-epicatechin (EC), (-)-epigallocatechin-3-gallate (EGCG), and (-)-epicatechin-3-gallate (ECG) were measured, and the odds ratio (OR) of breast cancer according to plasma level was estimated using a conditional logistic regression model. We found no statistically significant association between plasma tea polyphenol levels and breast cancer risk. Adjusted ORs for the highest versus lowest group were 0.90 (95% CI 0.42–1.96; P for trend = 0.98) for EGC, 0.95 (95% CI 0.43–2.08; P for trend = 0.86) for EC, 1.21 (95% CI 0.52–2.80; P for trend = 0.53) for EGCG, and 1.75 (95% CI 0.81–3.78; P for trend = 0.15) for ECG. Stratified analyses according to baseline menopausal status showed no remarkable difference between two strata. This nested case–control study found no overall association between plasma tea polyphenols and the risk of breast cancer in Japan.

Keywords Breast cancer · Plasma tea polyphenol · (-)-Epigallocatechin-3-gallate · Nested case–control study

Abbreviations

BMI	Body mass index
CI	Confidence interval
EC	(-)-Epicatechin
ECG	(-)-Epicatechin-3-gallate
EGC	(-)-Epigallocatechin
EGCG	(-)-Epigallocatechin-3-gallate
FFQ	Food frequency questionnaire
JPHC Study	Japan Public Health Center-based Prospective Study
OR	Odds ratio
PHC	Public health centers

This study is conducted for the Japan Public Health Center-based Prospective Study Group. The study group members are listed in Appendix.

M. Iwasaki (✉) · M. Inoue · S. Sasazuki · T. Miura · N. Sawada · T. Yamaji · T. Shimazu · S. Tsugane
Epidemiology and Prevention Division, Research Center for Cancer Prevention and Screening, National Cancer Center, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan
e-mail: moiwasak@ncc.go.jp

T. Miura
Department of Health and Nutrition, Faculty of Human Life, Jin-ai University, Echizen, Japan

W. C. Willett
Departments of Nutrition and Epidemiology, Harvard School of Public Health, Boston, MA, USA

W. C. Willett
Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Introduction

Green tea is one of the most popular beverages in Japan and China and is regularly consumed as a traditional habit and part of the culture. Due to its high polyphenol content, mainly catechins, there has been considerable interest in the impact of green tea intake on health outcomes, and possible benefits against the risk of cancer, cardiovascular diseases, and other conditions [1, 2]. Today, green tea extracts are a rapidly growing dietary supplement category in the United States.

As breast cancer risk is substantially lower in Asian than Western countries [3], the contribution of a high intake of green tea to low risk of breast cancer has been hypothesized. This hypothesis has been supported by in vitro and animal studies, which together have demonstrated various protective effects of green tea and tea polyphenols acting via strong antioxidant activity, inhibition of cell proliferation and angiogenesis, induction of apoptosis, and antiestrogenic properties [1, 4–7]. To date, however, findings from epidemiological studies have been inconsistent [8–13]. An inverse association was found in three case–control studies based on Asian-American and Chinese populations [8–10], whereas no association was observed in two cohort studies in Japan [11, 12] or one nested case–control study in Singapore [13]. This inconsistency might be in part explained by differences in study design and exposure measurements. Nevertheless, all previous studies assessed green tea intake by self-administered questionnaire or in-person interview [8–13], and no study has directly measured prediagnostic biomarkers of tea polyphenols. Tea polyphenol content in green tea varies according to preparation, type and amount of green tea leaves, frequency of renewing a tea batch in the pot, temperature of the boiled water, and brewing time, among others. To reduce misclassification due to these conditions and to better understand the role of tea polyphenols in the etiology of breast cancer, studies using prediagnostic biomarkers of tea polyphenols are required.

To clarify the effect of tea polyphenols on breast cancer risk, we measured plasma levels of (-)-epigallocatechin (EGC), (-)-epicatechin (EC), (-)-epigallocatechin-3-gallate (EGCG), and (-)-epicatechin-3-gallate (ECG) in a nested case–control study within a large-scale population-based prospective cohort in Japan. These compounds are all characteristic polyphenols, and EGCG in particular is the most abundant and biologically active catechin in green tea. As green tea intake varies widely among individual Japanese, this study provides a special opportunity to better understand the effect of relatively high-dose tea polyphenol levels achievable from habitual tea drinking on breast cancer risk.

Materials and methods

Study population

The Japan Public Health Center-based Prospective Study (JPHC Study), which began in 1990 for Cohort I and in 1993 for Cohort II, included 140,420 subjects (68,722 men and 71,698 women) living in municipalities supervised by 11 public health centers (PHC). Details of the study design have been described elsewhere [14]. The study protocol was approved by the institutional review board of the National Cancer Center, Tokyo, Japan.

The study population comprised registered Japanese residents living in each PHC area, aged 40–59 years in Cohort I and 40–69 years in Cohort II. In the present analysis, one PHC area was excluded since data on cancer incidence were not available. Thus, after exclusion of ineligible subjects ($n = 95$), we defined a population-based cohort of 67,426 women.

Questionnaire survey

A baseline self-administered questionnaire survey was conducted in 1990–1994. A total of 55,891 women (83%) returned the questionnaire, which contained questions concerning demographic characteristics, medical history, menstrual and reproductive history, anthropometric factors, physical activity, smoking and drinking habits, and diet.

Blood collection

Subjects voluntarily provided 10 ml of blood during health checkups in 1990–1995. Blood samples were divided into plasma and buffy layers and stored at −80°C until analysis. Among respondents to the baseline questionnaire, a total of 24,996 women (45%) donated blood.

Follow-up

All registered subjects were followed from the start of the study period to December 31, 2002. Data on residential relocation were obtained from residential registries. Among study subjects ($n = 24,996$), 1,289 subjects (5.2%) moved out of the study area and 5 (0.02%) were lost to follow-up within the study at-risk period.

Selection of cases and controls

Incidence data on breast cancer were collected for the JPHC cancer registry through two data sources, major local hospitals and population-based cancer registries. Death certificates were used to supplement information on cancer incidence. Site of origin and histologic type were coded by