Two recent GWAS studies for pancreatic cancer using Caucasian populations have identified associations with genome-wide significance on chromosomes 9p34.2 (*ABO*), 13q22.1, 1q32 (*NR5A2*) and 5p15.33 (*CLPTM1L-TERT*), and highlighted that accumulation of these common genetic risk variants with modest effects are likely to play an important role on this complex disease, either individually or in interaction with environmental factors [19–22]. As the ethnicity is one of the critical factors in the pathogenesis of the genetic diseases with complex gene-gene and gene-environmental interactions, we (Biobank Japan (BBJ) in The University of Tokyo and National Cancer Center (NCC) Japan) combined samples of 991 cases with pancreatic cancer and 5209 controls (Table S1), attempted to identify common genetic variations associated with susceptibility to pancreatic cancer in the Japanese population.

Results

After the standard quality control of the genotype results (Table S2), association analysis was performed for 420,236 SNPs using logistic regression analysis on the basis of allelic, dominant and recessive models after adjustment of age, sex and smoking status for each individual. The Q-Q plot for this GWAS based on allelic *P*-values by logistic regression revealed no significant population stratification with genomic inflation factor λ of 1.026 (Figure 1).

We successfully identified three genomic regions, 6p25.3, 12p11.21 and 7q36.2, shown to be significantly associated (*P*-value$<5.0\times10^{-7}$) with increased risk of pancreatic cancer in Japanese population as indicated in the Manhattan plot in Figure 2 (referred to ref. 23).

The most significantly-associated SNP, rs9502893 (*P*-value of 3.30×10^{-7}, per-allele odds ratio (OR) of 1.29 with 95% confidence interval (CI) of 1.17–1.43), is located within a 75-kb linkage disequilibrium (LD) block on chromosome 6p25.3 (Table 1). This LD block includes *FOXQ1* (forkhead box (Fox) Q1) gene, which is located 25 kb upstream to this marker SNP (Figure 3a). Imputation analysis also revealed modest association at SNPs located near to or on the *FOXQ1* gene suggesting it to be one of the causative genes for pancreatic cancer (Figure 3a and Table S3).

The second significantly-associated SNP, rs708224, located in the second intron of the gene *BICD1* (Bicaudal-D homolog 1) on chromosome 12p11 (*P*-value of 3.30×10^{-7}, per-allele OR of 1.32 with 95% CI of 1.19–1.47) (Table 1). The 80-kb LD block showing the association corresponds to the second intron of *BICD1* as revealed by the imputation analysis shown in Figure 3b (Table S3).

The third locus is marked by rs6464375, rs7779540, rs6973850 and rs1048768 in the first intron of *DPP6* gene. These SNPs indicated suggestive associations only under recessive model with minimum *P*-value of 4.41×10^{-7} (OR of 3.73 with 95%CI of 2.24–6.21) as shown in Table 1 and Figure 3c.

Discussion

Here we present results of GWAS analysis on 991 cases with pancreatic cancer and 5209 controls. Our study represents the first GWAS attempt to identify common variants associated with pancreatic cancer in Japanese population and successfully identified SNPs located on chromosomal loci of 6p25.3, 12p11.21 and 7q36.2 are significantly associated with increased risk of pancreatic cancer in Japanese population.

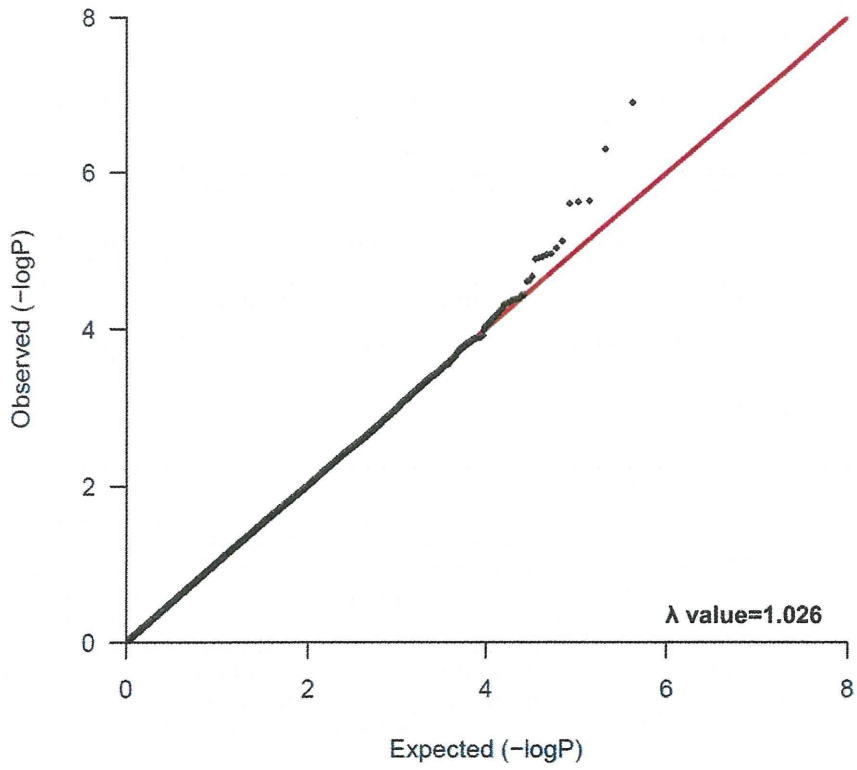

Figure 1. Q-Q plot for GWAS of pancreatic cancer in Japanese population. This Q-Q plot is based on logistic regression allelic *P*-values after standard quality control. (genomic inflation factor λ = 1.026).
doi:10.1371/journal.pone.0011824.g001

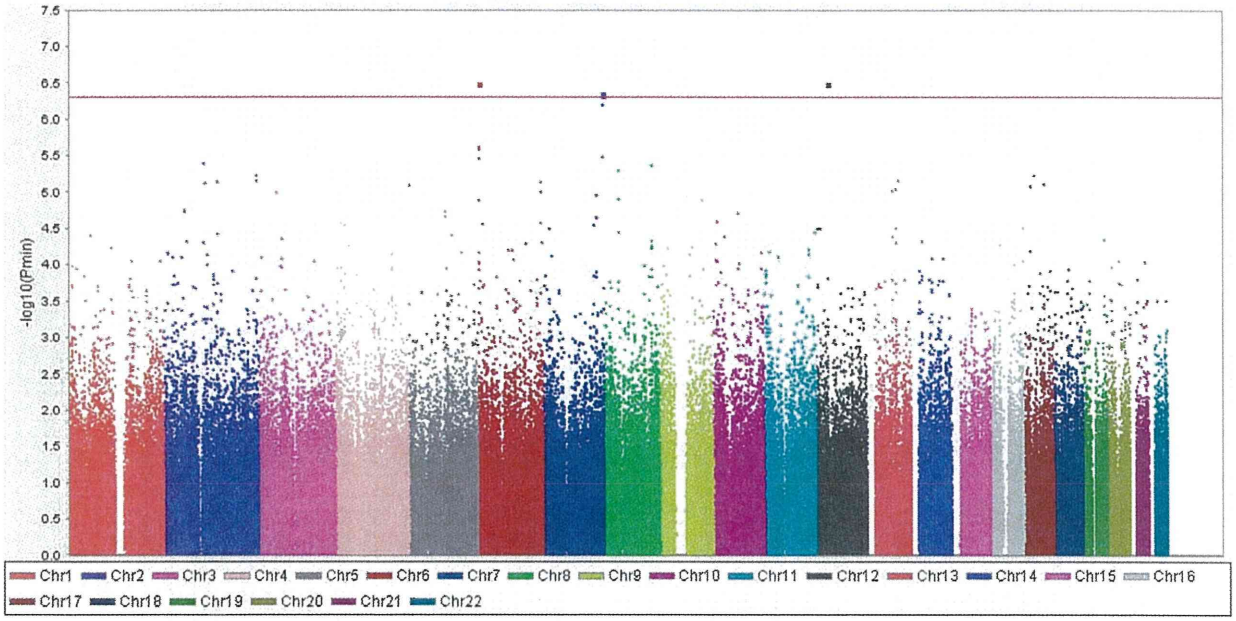

Figure 2. Manhattan plot for GWAS of pancreatic cancer in Japanese population. The plot is based on logistic regression model after correction of age, sex and smoking status. The P_{min} indicates the minimum *P*-value from logistic regression analysis for three models: allelic, dominant and recessive. Red line indicates genome-wide significant level (*P*-value = 5×10^{-7}).
doi:10.1371/journal.pone.0011824.g002

It is known that the development of the common disease is caused by the accumulation of common genetic variants, and each of this variant has a very modest effect on the risk (for example OR of <1.2). In order to detect such small fraction, GWAS involving much larger populations (5000–10000) should be required. Our study was expected to identify SNPs with moderate effects (i.e OR>1.4). Hence SNPs that show very modest effect might have failed to be identified through this study.

The most significantly associated SNP in this GWAS, rs9502893 (*P*-value = 3.30×10^{-7}, OR = 1.29) is located within a 75 kb LD block which encompasses gene *FOXQ1* on chromosome loci 6p25.3. *FOXQ1* encodes for protein forkhead box (Fox) Q1. The Fox family of transcription factors consists of at least 43 members and mutations in Fox genes can cause significant effects on human common disease and cancers [24,25]. A Fox member, FoxM1, is well-known to be associated with oncogenesis of pancreatic cancer. Down-regulation of this protein results in the inhibition of migration, invasion and angiogenesis in pancreatic cancer cells [26]. Furthermore, a recent study showed that FoxQ1 is overexpressed in pancreatic cancer, suggesting its role in pancreatic cancer tumorigenesis [27]. Although the SNP that we identified is approximately 25 kb downstream to this gene, the associated SNP may 'tag' the causative variant located on the expression regulatory region of the gene and subsequently alter expression of the gene. However, further study is needed to elucidate a precise biological role and mechanism of the gene function with regard to pancreatic carcinogenesis.

The second most significantly associated SNP, rs708224 (*P*-value = 3.30×10^{-7}, OR = 1.32) is located within the *BICD1* gene. This gene encodes a protein Bicaudal-D homolog 1, which plays a role in vacuolar trafficking. Previous studies reported substantial evidences indicating a link between vacuolar gene and shorter telomeres in yeast model [28–30]. In addition, Mangino et al. suggested that genetic variations within the *BICD1* gene could alter its transcriptional levels and in turn influence telomere length in humans [31]. Several recent studies have documented reduced telomere length in pancreatic ductal adenocarnoma specimens, suggesting telomeric dysfunction in pancreatic cancer cells [32–34]. Thus, it is of importance to determine the functional consequences of rs708224 and/or variations linked to this SNP in the pathogenesis of pancreatic cancer.

Several SNPs located in the first intron of *DPP6* indicated suggestive associations with an increased risk of pancreatic cancer in this study. *DPP6* encodes protein dipeptidyl-peptidase 6, which binds to specific voltage-gated potassium channels and alters their expression and biophysical properties. A recent study on core signaling pathways in human pancreatic cancers found three somatic mutations in DPP6 among 24 pancreatic cancer samples examined by detailed sequence analyses. This report also suggested that DPP6 might play a crucial role in regulation of invasion of pancreatic cancer cells [35]. Hence, our study strengthens the risk of DPP6 in pancreatic cancer and warrants further screening on this gene to confirm its association with pancreatic cancer.

Recent GWAS reports have indicated several loci on chromosomes 9p34.2, 13q22.1, 1q32.1 and 5p15.33 to be associated with an increased risk of pancreatic cancer in Caucasian population [21,22]. Among the significantly associated SNPs, rs9543325 on chromosome 13q22.1 showed moderate association in our study populations (*P*-value (allelic model) of 1.69×10^{-4}; OR of 1.21 with 95%CI of 1.10–1.34) (Table S4). On the other hand, SNPs on chromosomes 9p34.2 (rs505922) and 1q32.1 (rs3790844) showed a weak association in our study populations (*P*-values of 3.69×10^{-2} and 1.24×10^{-2}; ORs of 1.11 and 1.14 with 95% CI of 1.01–1.22 and 1.03–1.27, respectively) (Table S4). We were unable to replicate the remaining loci (*SHH* and two loci on chromosomes 5p15.33 and 15q14) in these reports, probably because most of these associated SNPs are either non-polymorphic or possess very low allelic frequencies (MAF = 0.01) in Japanese population. The power of our study was not sufficient enough to detect positive associations for

Table 1. SNPs that show suggestive association with increase risk of pancreatic cancer in Japanese population.

CHR*	SNP	Position*	Risk allele	RAF		Allelic				Dominant				Recessive				P_{min}	Gene	Relativeloc*
				Case	Control	*P*-value	OR	L95	U95	*P*-value	OR	L95	U95	*P*-value	OR	L95	U95			
6	rs9502893	1285189	G	0.411	0.351	3.30E-07	1.29	1.17	1.43	2.97E-05	1.36	1.18	1.57	2.18E-05	1.50	1.24	1.80	3.30E-07	*FOXQ1*	25196
12	rs708224	32327676	A	0.718	0.656	3.30E-07	1.32	1.19	1.47	8.54E-07	1.42	1.23	1.63	2.09E-03	1.46	1.15	1.86	3.30E-07	*BICD1*	0
7	rs6464375	153256776	A	0.116	0.103	1.15E-01	1.13	0.97	1.32	7.36E-01	1.03	0.87	1.22	4.41E-07	3.73	2.24	6.21	4.41E-07	*DPP6*	0
7	rs7779540	153253595	A	0.116	0.103	1.08E-01	1.14	0.97	1.33	7.12E-01	1.03	0.87	1.23	4.58E-07	3.72	2.23	6.20	4.58E-07	*DPP6*	0
7	rs6973850	153269181	A	0.116	0.106	2.23E-01	1.10	0.94	1.29	9.76E-01	1.00	0.84	1.18	6.27E-07	3.64	2.19	6.04	6.27E-07	*DPP6*	0
6	rs11242679	1282311	A	0.366	0.311	2.40E-06	1.28	1.15	1.42	1.15E-05	1.37	1.19	1.58	2.07E-03	1.39	1.13	1.71	2.40E-06	*FOXQ1*	22318
6	rs7750826	1281867	G	0.365	0.311	2.57E-06	1.28	1.15	1.41	1.30E-05	1.37	1.19	1.57	1.98E-03	1.39	1.13	1.71	2.57E-06	*FOXQ1*	21874
7	rs10487687	153271407	A	0.150	0.136	8.99E-02	1.13	0.98	1.30	6.76E-01	1.03	0.88	1.21	3.35E-06	2.66	1.76	4.02	3.35E-06	*DPP6*	0
6	rs11242674	1252846	A	0.355	0.301	3.46E-06	1.28	1.15	1.41	9.64E-06	1.37	1.19	1.58	4.59E-03	1.37	1.10	1.69	3.46E-06	*FOXQ1*	−4829
2	rs6711606	101288602	A	0.135	0.116	1.27E-02	1.20	1.04	1.39	1.86E-01	1.12	0.95	1.32	4.02E-06	2.81	1.81	4.37	4.02E-06	*RNF149*	0
8	rs10088262	124834883	A	0.374	0.341	3.42E-03	1.16	1.05	1.28	4.30E-06	1.40	1.21	1.61	3.98E-01	0.91	0.74	1.13	4.30E-06	*FAM91A1*	−15180
8	rs7832232	38588460	A	0.483	0.454	1.43E-02	1.13	1.03	1.25	7.63E-01	0.98	0.84	1.14	5.10E-06	1.45	1.24	1.71	5.10E-06	*RNF5P1*	−10528
2	rs6736997	235279936	A	0.372	0.328	2.95E-04	1.20	1.09	1.33	4.96E-02	1.15	1.00	1.33	5.85E-06	1.57	1.29	1.91	5.85E-06	*ARL4C*	−209504
17	rs225190	27901771	G	0.410	0.360	5.99E-06	1.26	1.14	1.39	1.92E-04	1.32	1.14	1.52	2.37E-04	1.43	1.18	1.72	5.99E-06	*MYO1D*	0
2	rs4663158	235263691	A	0.397	0.352	1.39E-04	1.21	1.10	1.34	2.89E-02	1.17	1.02	1.35	6.91E-06	1.53	1.27	1.85	6.91E-06	*ARL4C*	−193259
13	rs2039553	79197723	A	0.291	0.268	1.32E-02	1.15	1.03	1.28	4.39E-01	1.06	0.92	1.21	7.01E-06	1.73	1.36	2.19	7.01E-06	*NDFIP2*	171501
2	rs1427593	137271694	A	0.110	0.080	1.55E-05	1.42	1.21	1.66	7.10E-06	1.49	1.25	1.77	4.30E-01	1.31	0.67	2.58	7.10E-06	*THSD7B*	−193238
6	rs3016539	162156065	A	0.903	0.871	1.67E-05	1.42	1.21	1.67	7.28E-06	1.50	1.26	1.79	2.90E-01	1.36	0.77	2.43	7.28E-06	*PARK2*	0
2	rs12615966	104745389	A	0.112	0.097	1.40E-02	1.22	1.04	1.42	1.59E-01	1.13	0.95	1.35	7.44E-06	3.15	1.91	5.21	7.44E-06	*LOC284998*	−6744
17	rs2257205	53803296	A	0.378	0.327	1.58E-05	1.25	1.13	1.38	7.74E-06	1.38	1.20	1.59	2.97E-02	1.25	1.02	1.53	7.74E-06	*RNF43*	0
5	rs6879627	2162901	G	0.575	0.522	8.12E-06	1.25	1.14	1.39	4.66E-04	1.31	1.12	1.52	1.57E-04	1.42	1.18	1.69	8.12E-06	*LOC731559*	225138
17	rs4924935	18694595	G	0.269	0.228	8.80E-05	1.25	1.12	1.40	8.15E-06	1.37	1.19	1.58	5.06E-01	1.11	0.82	1.48	8.15E-06	PRPSAP2	(7622
17	rs1737947	18772157	G	0.252	0.212	3.88E-05	1.27	1.13	1.43	8.49E-06	1.37	1.19	1.58	2.89E-01	1.19	0.87	1.62	8.49E-06	PRPSAP2	0
13	rs1886449	72830115	A	0.424	0.383	2.61E-04	1.21	1.09	1.33	5.62E-02	1.15	1.00	1.33	9.24E-06	1.51	1.26	1.80	9.24E-06	*LOC730242*	−206271
13	rs1585440	65379816	C	0.761	0.713	9.28E-06	1.30	1.16	1.45	3.09E-05	1.35	1.17	1.55	7.36E-03	1.50	1.12	2.03	9.28E-06	*LOC387933*	−118820
3	rs4683235	46476935	A	0.215	0.173	9.93E-06	1.31	1.16	1.48	7.53E-05	1.34	1.16	1.54	2.24E-03	1.70	1.21	2.38	9.93E-06	*LTF*	0

Odds ratios, 95% confidence limits and P-values were obtained using logistic regression analysis according to allelic, dominant and recessive model after adjustment of age, sex and smoking.
RAF, risk allele frequency; OR, odds ratio; L95, U95, lower and upper confidence limits; P_{min}, minimum *P*-value among three genetic models.
*Position and relative loci (Relativeloc) are based on NCBI Human Genome Build 36.
doi:10.1371/journal.pone.0011824.t001

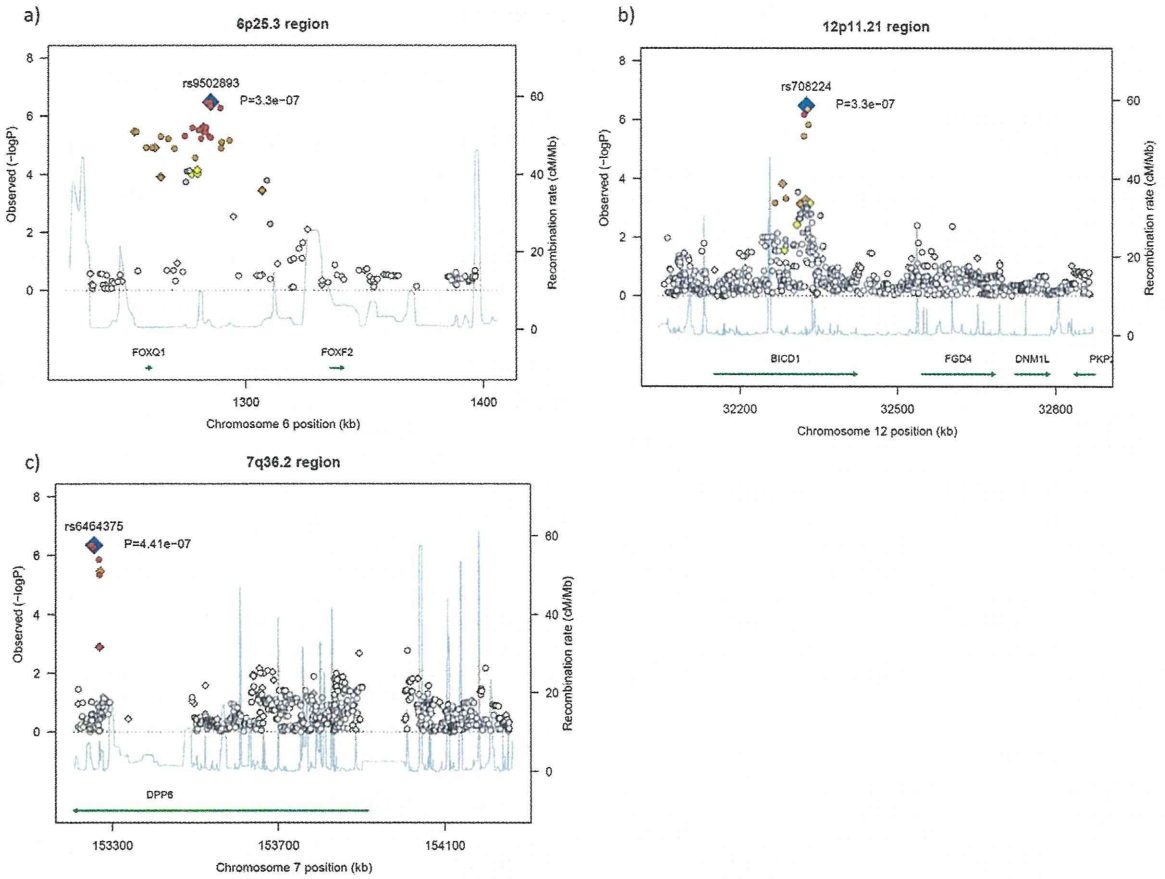

Figure 3. Regional association plots for three pancreatic cancer risk loci. (a) 6p25.3 region, SNP rs9502893 located 25 kb upstream to gene *FOXQ1*. (b) 12p11.21 region, SNP rs708224 is located at the second intron of gene *BICD1*. (c) 7q36.2 region, SNP rs6464375 is located at the first intron of gene *DPP6 transcript variant 3*. Each of the marker SNPs is marked by a blue diamond. SNPs that are genotyped in the Illumina platform are plotted as diamonds; Imputed SNPs are plotted as circles. The color intensity reflects the extent of LD with the marker SNP, red ($r^2 \geq 0.8$), orange ($0.5 \leq r^2 < 0.8$), yellow ($0.2 \leq r^2 < 0.5$) and white ($r^2 < 0.2$). Light blue line indicated local recombination rate.
doi:10.1371/journal.pone.0011824.g003

these variants with the low allelic frequency. Such ethnic difference in genetic architecture of disease susceptibility is not rare. For example, two recent GWAS reported common variants on *KCNQ1* gene associated with type 2 diabetes mellitus in Japanese population, but European GWAS were unable to identify the associations due to the low allelic frequency of these variants in the population [36,37]. In addition, identification of susceptibility loci may be also influenced by the differences in the LD structure across different populations and by potential interaction with other genetic variants and environmental factors [38].

In summary, this study represents the first GWAS to identify common variants possibly associated with pancreatic cancer in Japanese population. Our study confirmed the association from the Caucasian GWAS studies and revealed several novel possible candidate associated loci that were not detected in the previous Caucasian GWAS studies. Nevertheless, further additional replications are required to confirm or exclude the current findings.

Materials and Methods

Case and control subjects

A total of 331 and 675 cases that were clinically and/or histologically diagnosed to have an invasive pancreatic ductal adenocarcinoma were obtained from Biobank Japan (http://biobankjp.org) at the Institute of Medical Science, The University of Tokyo as well as National Cancer Center Hospital, respectively. The control samples consisted of Japanese volunteers that were obtained from Osaka-Midosuji Rotary Club, Osaka, Japan ($n = 906$) as well as from staff members in Keio University, Japan, who participated in its health-check program ($n = 677$). In addition, individuals who were registered in Biobank Japan as subjects with various diseases except cancer ($n = 3,728$) (those having pulmonary tuberculosis, chronic hepatitis-B, keroid, drug-induced skin rash, peripheral artery disease, arrhythmia, stroke and myocardial infarction) were used as controls. All samples were obtained after obtaining the written informed consent. This project was approved by the ethics committee at The Institute of Medical Sciences, The University of Tokyo, National Cancer Center and Keio University. Individuals who had clinical history of diabetes mellitus (a possible confounding factor for pancreatic cancer) were excluded from these control sets. For sample quality control, we excluded five cases with call rate < 0.98. After performing principal component analysis, we excluded outliers of 10 cases and 102 controls, who did not belong to the major Japanese cluster (Hondo cluster) (Figure S1) [39]. We eventually performed the association study based on 991 cases and 5209 controls (Table S1). Power calculation showed that our study

would have over 90% power to detect a per-allele OR of 1.4 or greater for an allele with 30% frequency at the genome-wide significance level ($\alpha = 5\times10^{-7}$).

SNP genotyping and quality control

All the individuals were genotyped using either Illumina Infinium HumanHap550v3 or Illumina Infinium Human610-Quad DNA Analysis Genotyping BeadChip. SNPs common in the two platforms were used for further analysis. We applied SNP quality control for all sets of samples as follows; SNP call rate should be >0.99 in both cases and controls, and *P*-value of Hardy-Weinberg equilibrium test should be $>1.0\times10^{-6}$ in controls. SNPs with minor allele frequency (MAF) of <0.01 in both case and control samples were excluded from the further analysis (Table S2).

Statistical analysis

We analyzed each SNP using logistic regression adjusted for age (continuous), sex and smoking status (current/former, never). *P*-values and OR with 95%CI were calculated for allelic, dominant and recessive models. We used the minimum *P*-values obtained from three models to evaluate the statistical significance of the association. All OR were reported with respect to the risk allele. All the statistical analyses were performed using R statistical environment version 2.9.0 (http://www.r-project.org/) or PLINK 1.06 (http://pngu.mgh.harvard.edu/purcell/plink/). R statistical environment version 2.9.0 was employed to draw Q-Q plot and regional association plot.

Genotype Imputation

We performed genotype imputation analysis for each set of samples by utilizing a Hidden Markov model as programmed in MACH version 1.0 (http://www.sph.umich.edu/csg/abecasis/mach/index.html). To infer untyped and missing genotypes around the candidate chromosomal loci, we provided genotypes from our own samples together with haplotypes for reference samples (Japanese from Tokyo, JPT) from HapMap database (http://hapmap.ncbi.nlm.nih.gov/). SNPs with low genotyping rate (<99%), showing deviations from Hardy-Weinberg equilibrium ($<1.0\times10^{-6}$), or MAF (<0.01) were excluded from the analysis. MACH version 1.0 was used to estimate haplotypes, map crossover and error rates using 50 iterations of the Markov chain Monte Carlo algorithm. By utilizing the genotype information from the HapMap database, maximum likelihood genotypes were generated. For quality control, we retained imputed SNPs with the estimated r^2 of >0.3. We also picked up a total of 17 SNPs (*P*-value<0.001) to verify the association using Invader and TaqMan genotyping methods (data not shown).

References

1. Kelsen DP, Portenoy R, Thaler H, Tao Y, Brennan M (1997) Pain as a predictor of outcome in patients with operable pancreatic carcinoma. Surgery 122: 53–59.
2. Catanzaro A, Richardson S, Veloso H, Isenberg GA, Wong RC, et al. (2003) Long-term follow-up of patients with clinically indeterminate suspicion of pancreatic cancer and normal EUS. Gastrointest Endosc 58: 836–840.
3. Anderson KE, Mack T, Silverman D (2006) Cancer of the pancreas. In: Schottenfeld D, Fraumeni JF, Jr., eds. Cancer Epidemiology and Prevention. New York: Oxford University Press. pp 721–762.
4. Stevens RJ, Roddam AW, Beral V (2007) Pancreatic cancer in type 1 and young-onset diabetes: systematic review and meta-analysis. Br J Cancer 96: 507–509.
5. Lowenfels AB, Maisonneuve P, Cavallini G, Ammann RW, Lankisch PG, et al. (1993) Pancreatitis and the risk of pancreatic cancer. International Pancreatitis Study Group. N Engl J Med 328: 1433–1437.
6. Del Chiaro M, Zerbi A, Falconi M, Bertacca L, Polese M, et al. (2007) Cancer risk among the relatives of patients with pancreatic ductal adenocarcinoma. Pancreatology 7: 459–469.
7. McWilliams RR, Rabe KG, Olswold C, De Andrade M, Petersen GM (2005) Risk of malignancy in first-degree relatives of patients with pancreatic carcinoma. Cancer 104: 388–394.
8. Fernandez E, La Vecchia C, D'Avanzo B, Negri E, Franceschi S (1994) Family history and the risk of liver, gallbladder, and pancreatic cancer. Cancer Epidemiol Biomarkers Prev 3: 209–212.
9. Tersmette AC, Petersen GM, Offerhaus GJ, Falatko FC, Brune KA, et al. (2001) Increased risk of incident pancreatic cancer among first-degree relatives of patients with familial pancreatic cancer. Clin Cancer Res 7: 738–744.
10. Yan L, McFaul C, Howes N, Leslie J, Lancaster G, et al. (2005) Molecular analysis to detect pancreatic ductal adenocarcinoma in high-risk groups. Gastroenterology 128: 2124–2130.
11. Caldas C, Hahn SA, da Costa LT, Redston MS, Schutte M, et al. (1994) Frequent somatic mutations and homozygous deletions of the p16 (MTS1) gene in pancreatic adenocarcinoma. Nat Genet 8: 27–32.
12. Barton CM, Staddon SL, Hughes CM, Hall PA, O'Sullivan C, et al. (1991) Abnormalities of the p53 tumour suppressor gene in human pancreatic cancer. Br J Cancer 64: 1076–1082.

Supporting Information

Table S1 Sample characteristic of this study.
Found at: doi:10.1371/journal.pone.0011824.s001 (0.02 MB XLS)

Table S2 Total number of SNPs excluded according to each quality control criteria.
Found at: doi:10.1371/journal.pone.0011824.s002 (0.02 MB XLS)

Table S3 Imputation analysis around significantly associated SNPs.
Found at: doi:10.1371/journal.pone.0011824.s003 (0.04 MB XLS)

Table S4 Association study of SNPs which shown to be significantly associated with increased risk of pancreatic cancer in Caucasian population in Japanese.
Found at: doi:10.1371/journal.pone.0011824.s004 (0.02 MB XLS)

Figure S1 Principal component analysis for GWAS of pancreatic cancer in Japanese population. a) Principal component analysis for GWAS of pancreatic cancer in Japanese population refer to four HapMap population control subjects including CEU indicates Caucasians from Utah; YRI, Nigerians from Yoruba; CHB, Han Chinese from Beijing and JPT, Japanese from Tokyo. b) Principal component analysis of study subjects referred only to Asian populations. We utilized samples from the homogenous case-control (Hondo) cluster.
Found at: doi:10.1371/journal.pone.0011824.s005 (9.43 MB TIF)

Acknowledgments

We express our heartfelt gratitude to all the patients who participate in this study. We would like to thank Dr Yoichiro Kamatani for his constructive comments and suggestions. Our thankfulness also goes to the member of The Rotary Club of Osaka-Midosuji District 2660 Rotary International in Japan for making this study possible. We thank Drs. Hideki Ueno, Masafumi Ikeda, Chigusa Morizane, Yoshihiro Sakamoto, Minoru Esaki, Tomoo Kosuge and Nobuyoshi Hiraoka for ascertainment of the patients and their clinico-pathological information at the National Cancer Center Hospital. We also would like to express our gratefulness to Miss Kumi Matsuda for her outstanding technical assistance.

Author Contributions

Conceived and designed the experiments: SKL AK HZ MK YD NK TY YN HS. Performed the experiments: SKL AK MK SO HS. Analyzed the data: SKL AK HZ AS AT MK NK SC HT TY YN HS. Contributed reagents/materials/analysis tools: SKL AK HZ AS AT MK NK HH KS TO TY YN. Wrote the paper: SKL AK HZ TY YN.

13. Berrozpe G, Schaeffer J, Peinado MA, Real FX, Perucho M (1994) Comparative analysis of mutations in the p53 and K-ras genes in pancreatic cancer. Int J Cancer 58: 185–191.
14. Hahn SA, Schutte M, Hoque AT, Moskaluk CA, da Costa LT, et al. (1996) DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. Science 271: 350–3.
15. Hruban RH, Iacobuzio-Donahue C, Wilentz RE, Goggins M, Kern SE (2001) Molecular pathology of pancreatic cancer. Cancer J 7: 251–258.
16. Hahn SA, Greenhalf B, Ellis I, Sina-Frey M, Rieder H, et al. (2003) BRCA2 germline mutations in familial pancreatic carcinoma. J Natl Cancer Inst 95: 214–221.
17. Wong T, Howes N, Threadgold J, Smart HL, Lombard MG, et al. (2001) Molecular diagnosis of early pancreatic ductal adenocarcinoma in high-risk patients. Pancreatology 1: 486–509.
18. Jones S, Hruban RH, Kamiyama M, Borges M, Zhang X, et al. (2009) Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. Science 324: 217.
19. MacLeod SL, Chowdhury P (2006) The genetics of nicotine dependence: relationship to pancreatic cancer. World J Gastroenterol 12: 7433–7439.
20. Milne RL, Greenhalf W, Murta-Nascimento C, Real FX, Malats N (2009) The inherited genetic component of sporadic pancreatic adenocarcinoma. Pancreatology 9: 206–214.
21. Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, Fuchs CS, Petersen GM, et al. (2009) Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. Nat Genet 41: 986–990.
22. Petersen GM, Amundadottir L, Fuchs CS, Kraft P, Stolzenberg-Solomon RZ, et al. (2010) A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nat Genet 42: 224–228.
23. Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661–78.
24. Myatt SS, Lam EW (2007) The emerging roles of forkhead box (Fox) proteins in cancer. Nat Rev Cancer 7: 847–859.
25. Hannenhalli S, Kaestner KH (2009) The evolution of Fox genes and their role in development and disease. Nat Rev Genet 10: 233–240.
26. Wang Z, Banerjee S, Kong D, Li Y, Sarkar FH (2007) Down-regulation of Forkhead Box M1 transcription factor leads to the inhibition of invasion and angiogenesis of pancreatic cancer cells. Cancer Res 67: 8293–8300.
27. Cao D, Hustinx SR, Sui G, Bala P, Sato N, et al. (2004) Identification of novel highly expressed genes in pancreatic ductal adenocarcinomas through a bioinformatics analysis of expressed sequence tags. Cancer Biol Ther 3: 1081–1089.
28. Askree SH, Yehuda T, Smolikov S, Gurevich R, Hawk J, et al. (2004) A genome-wide screen for Saccharomyces cerevisiae deletion mutants that affect telomere length. Proc Natl Acad Sci U S A 101: 8658–8663.
29. Gatbonton T, Imbesi M, Nelson M, Akey JM, Ruderfer DM, et al. (2006) Telomere Length as a Quantitative Trait: Genome-Wide Survey and Genetic Mapping of Telomere Length-Control Genes in Yeast. PLoS Genet 2: e35. doi:10.1371/journal.pgen.0020035.
30. Rog O, Smolikov S, Krauskopf A, Kupiec M (2005) The yeast VPS genes affect telomere length regulation. Curr Genet 47: 18–28.
31. Mangino M, Brouilette S, Braund P, Tirmizi N, Vasa-Nicotera M, et al. (2008) A regulatory SNP of the BICD1 gene contributes to telomere length variation in humans. Hum Mol Genet 17: 2518–2523.
32. Büchler P, Conejo-Garcia JR, Lehmann G, Müller M, Emrich T, et al. (2001) Real-time quantitative PCR of telomerase mRNA is useful for the differentiation of benign and malignant pancreatic disorders. Pancreas 22: 331–340.
33. Kobitsu K, Tsutsumi M, Tsujiuchi T, Suzuki F, Kido A, et al. (1997) Shortened telomere length and increased telomerase activity in hamster pancreatic duct adenocarcinomas and cell lines. Mol Carcinog 18: 153–159.
34. van Heek NT, Meeker AK, Kern SE, Yeo CJ, Lillemoe KD, et al. (2002) Telomere shortening is nearly universal in pancreatic intraepithelial neoplasia. Am J Pathol 161: 1541–1547.
35. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, et al. (2008) Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science 321: 1801–1806.
36. Unoki H, Takahashi A, Kawaguchi T, Hara K, Horikoshi M, et al. (2008) SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. Nat Genet 40: 1098–1102.
37. Yasuda K, Miyake K, Horikawa Y, Hara K, Osawa H, et al. (2008) Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. Nat Genet 40: 1092–1097.
38. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, et al. (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 9: 356–369.
39. Yamaguchi-Kabata Y, Nakazono K, Takahashi A, Saito S, Hosono N, et al. (2008) Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. Am J Hum Genet 83: 445–456.

Breast Cancer Res Treat (2010) 123:885–893
DOI 10.1007/s10549-010-0804-4

EPIDEMIOLOGY

Genetic polymorphisms in folate and alcohol metabolism and breast cancer risk: a case–control study in Thai women

Suleeporn Sangrajrang · Yasunori Sato · Hiromi Sakamoto · Sumiko Ohnami · Thiravud Khuhaprema · Teruhiko Yoshida

Received: 23 December 2009 / Accepted: 10 February 2010 / Published online: 24 February 2010

Abstract Dietary folate as well as polymorphic variants in one-carbon metabolism genes may modulate risk of breast cancer through aberrant DNA methylation and altered nucleotide synthesis and repair. Alcohol is well recognized as a risk factor for breast cancer, and interactions with one-carbon metabolism has also been suggested. The purpose of this study is to test the hypothesis that genetic polymorphisms in the folate and alcohol metabolic pathway are associated with breast cancer risk. Twenty-seven single nucleotide polymorphisms (SNPs) in the *MTR*, *MTRR*, *MTHFR*, *TYMS*, *ADH1C*, *ALDH2*, *GSTP1*, *NAT1*, *NAT2*, *CYP2E1 DRD2*, *DRD3*, and *SLC6A4* were genotyped. Five hundred and seventy patients with histopathogically confirmed breast cancer and 497 controls were included in the present study. Association of genotypes with breast cancer risk was evaluated using multivariate logistic regression to estimate odds ratios (OR) and their 95% confidence intervals (95% CI). Increased risk was observed for homozygotes at the *MTR* SNPs (rs1770449 and rs1050993) with the OR = 2.21 (95% CI 1.18–4.16) and OR = 2.24 (95% CI 1.19–4.22), respectively. A stratified analysis by menopausal status indicated the association between the *NAT2* SNP (rs1799930) and breast cancer was mainly evident in premenopausal women (OR 2.70, 95% CI 1.20–6.07), while the *MTRR* SNP (rs162049) was significant in postmenopausal women (OR 1.61, 95% CI 1.07–2.44). Furthermore, SNPs of the genes that contribute to alcohol behavior, *DRD3* (rs167770), *DRD2* (rs10891556), and *SLC6A4* (rs140701), were also associated with an increased risk of breast cancer. No gene–gene or gene–environment interactions were observed in this study. Our results suggest that genetic polymorphisms in folate and alcohol metabolic pathway influence the risk of breast cancer in Thai population.

Keywords Breast cancer · Folate · Alcohol · Metabolizing enzyme · Single nucleotide polymorphisms

S. Sangrajrang (✉) · T. Khuhaprema
Research Division, National Cancer Institute, Rama VI Road, Ratchathewi, Bangkok 10400, Thailand
e-mail: sulee@health.moph.go.th

Y. Sato · H. Sakamoto · S. Ohnami · T. Yoshida
Genetic Division, National Cancer Center Research Institute, Tokyo, Japan

Introduction

Breast cancer is the second most common cancer in Thai women and the incidence is still increasing [1]. A wide variety of genetic damage induced by endogenous metabolites and exogenous hazards may contribute to the etiology of breast cancer. Folate is an important nutrient required for DNA synthesis, and it is also involved in the methionine metabolic pathway, which is crucial for DNA methylation [2]. At least 30 different enzymes are involved in this complex pathway including methylenetetrahydrofolate reductase (MTHFR), methionine synthase (MTR), methionine synthase reductase (MTRR), and thymidylate synthase (TYMS). Defects or polymorphic variations in the folate metabolic pathway may influence cancer susceptibility [3].

Evidence level seems to be high that alcohol increases risk of breast cancer [4]. A pooled analysis of 4,335 breast cancer cases and >300,000 controls suggested that intake of 2–5 drinks/day increased risk by roughly 40% [5]. The underlying mechanisms are not firmly established [6] but may include an influence on circulating levels of estrogens [7], immune function, enhanced permeability of chemical

carcinogens, decreased absorption of essential nutrients [8], or through metabolism of alcohol to acetaldehyde, a known carcinogen [9].

A relative folate deficiency may develop in individuals who chronically consume more than moderate amounts of alcohol because of the negative effects of alcohol on folate metabolism, including malabsorption, increased excretion, or enzymatic suppression [10]. The potential for high folate intake to counteract the elevated risk of breast cancer associated with alcohol consumption has been illustrated by data from a number of studies [11–14].

To further investigate the role of these pathways in mammary carcinogenesis, we analyzed the association between breast cancer and 27 SNPs in 13 key genes involved in one-carbon and alcohol metabolism on 570 cases and 497 controls in Thai women. We also investigated gene–gene and gene–environment interactions.

Materials and methods

Study population

Cases were all new incident breast cancer patients histologically diagnosed at the National Cancer Institute in Bangkok and at the hospital in Khon Kaen province of North Eastern of Thailand during the period of May 2002–March 2004 and August 2005–August 2006, with a participation rate of 99.6% (600/602). Controls were randomly selected from healthy women who visited patients admitted to the same hospitals for diseases other than breast or ovarian cancer. All of the 642 control individuals were recruited during the same study period as the case ascertainment. The participation rate among visitors who were asked to participate was 98.9% (642/649). Informed consent was obtained from all participants and a structured questionnaire was administered by trained interviewers to collect information on demographic and anthropometric data, reproductive and medical history, residential history, physical activity and occupation as well as diet (see Table 1). Lifestyle exposure parameters were reported as follows: tobacco smoking: less than or equal to 6 months of smoking in life, or if she smokes longer than 6 months, the sum of cigarettes smoked is less than or equal to 50 per 6 months (non-smoker) versus more than 50 cigarettes over a 6-month period (smoker); involuntary tobacco smoking: less versus more than or equal to 1 h of exposure per day; alcohol consumption: less versus more than or equal to once a week for at least 6 months. Approximately 7 ml of blood were collected from participants, but 30 cases and 145 controls refused to give blood samples. In total, blood samples of 570 cases and 497 controls were included in the genotype analysis, resulting in a participation rate of 95.0% (cases) and 77.4% (controls). The study was approved by the ethical review committee for research in human subjects, Ministry of Public Health, Thailand and by the ethics committee of National Cancer Center, Japan.

Genotyping analysis

Genomic DNA was isolated from buffy coats using a QIAmp DNA blood kit (Qiagen, Hilden, Germany). DNA concentrations were measured by PicoGreen dsDNA qualification kits (Molecular Probes, Leiden, The Netherlands). All SNPs were analyzed by TaqMan 5′ nuclease assay using the ABI PRISM 7900 HT Sequence Detection System (Applide Biosystems LLC, Foster city, CA, USA). Oligonucleotide primers and the dual labeled allele specific probes were designed by ABI. PCR were performed in 384-well plates with each plate containing four control samples. A set of three 384-well plates were prepared to accommodate 570 cases and 497 control subjects and used for genotyping. Genomic DNA (5 ng) was amplified in a total volume of 5 μl in the presence of 100 μM of each of the dNTPs, 3 pmols of each of appropriate primers, 2 pmols of each of the corresponding dual labeled probes, and 0.025 units of Taq DNA polymerase. PCR cycling consisted of 40 cycles at 94°C for 15 s, 55–60°C for 15 s, and 72°C for 15 s. The results of 5% blindly repeated samples were at least 99% concordant each other. Genotyping success rate for individual polymorphisms averaged 95%.

Statistical analyses

Hardy–Weinberg equilibrium (HWE) testing was used as one of the measures for a quality control for genotyping, and allele and genotype frequencies were calculated. The multivariate logistic regression analyses were applied to evaluate differences in genotype distributions, and the odds ratios (OR) and their 95% confidence intervals (CI) were calculated after adjustment for the following covariates: age, body mass index (BMI), smoking, pregnancy and breast feeding, family history of breast cancer in the first-degree relatives, education, and menopausal status. Alcohol consumption was not included, because it was not associated with the breast cancer in our study (Table 2). For some SNPs, additional tests were also performed by Fisher's exact test on allelic contingency table and by Cochran–Armitage trend test to confirm the observed genotype-specific associations.

A stratified analysis was performed as an exploratory, adjunct analysis. Selected strata were menopausal, pregnancy, breast feeding, oral contraceptive use, estrogen receptor, and progesterone receptor status. The adjusted ORs and 95% CIs for OR were calculated by the

Table 1 Selected characteristics of study population

Characteristics	Cases	Controls	*P* value
Age, *n* (mean ± SD)	570 (46.0 ± 10.6)	497 (43.2 ± 12.4)	<0.0001
Age at menarche, *n* (mean ± SD)	561 (14.7 ± 1.9)	496 (14.86 ± 1.8)	0.61
Age at menopause, *n* (mean ± SD)	99 (46.6 ± 5.5)	85 (47.2 ± 5.7)	0.46
Age at first pregnancy, *n* (mean ± SD)	425 (23.1 ± 5.4)	331 (22.8 ± 4.9)	0.48
Pregnancy ($n = 1{,}067$), *n* (%)			
No	145 (25.4)	166 (33.4)	0.004
Yes	425 (74.6)	331 (66.6)	
Breast feeding ($n = 736$), *n* (%)			
No	59 (14.4)	27 (8.3)	0.011
Yes	351 (85.6)	299 (91.7)	
Oral contraceptive use ($n = 1{,}061$), *n* (%)			
No	310 (54.5)	288 (58.5)	0.19
Yes	259 (45.5)	204 (41.5)	
Menopausal status ($n = 1{,}062$), *n* (%)			
Premenopausal	325 (57.4)	316 (63.7)	0.038
Postmenopausal	241 (42.6)	180 (36.3)	
Body mass index ($n = 1{,}067$), *n* (mean ± SD kg/m^2)	570 (24.1 ± 4.4)	497 (23.0 ± 4.1)	3.7×10^{-5}[a]
BMI < 20 kg/m^2 (%)	87 (15.3)	111	
20 ≤ BMI < 25	273 (47.9)	269	
25 ≤ BMI	210 (36.8)	117	
Tobacco smoking ($n = 1{,}066$), *n* (%)			
No	552 (97.0)	489 (98.4)	0.16
Yes	17 (3.0)	8 (1.6)	
Involuntary smoking ($n = 1{,}040$), *n* (%)			
No	489 (88.7)	456 (93.3)	0.013
Yes	62 (11.3)	33 (6.7)	
Alcohol consumption ($n = 1{,}067$), *n* (%)			
No	525 (92.1)	469 (94.4)	0.18
Yes	45 (7.9)	28 (5.6)	
Family history of breast cancer in first-degree relatives ($n = 1{,}067$), *n* (%)			
No	544 (95.4)	487 (98.0)	0.026
Yes	26 (4.6)	10 (2.0)	
Education ($n = 1{,}067$), *n* (%)			
≤9 years	398 (69.8)	262 (52.7)	1.1×10^{-8}
>9 years	172 (30.2)	235 (47.3)	

[a] The *P* value for the BMI was calculated by Fisher's exact test on the 2 × 3 contingency table

multivariate logistic regression analyses. In addition, gene–gene interactions and gene–environment interactions were evaluated by logistic model including an interaction term between genes and those between gene and environmental factors.

The statistical significance was defined as $P \leq 0.05$, and adjustment for multiple testing, which is absolutely necessary in the validation phase of association studies, was not performed due to an exploratory hypothesis-generating nature of the study.

All statistical analyses were carried out using the Statistical Analysis System (SAS) software Version 9.1 (SAS Institute Inc, Cary, NC), and partially the R suite (http://www.r-project.org/).

Results

Characteristics of the study population were compared by case–control status as shown in Table 1. The mean age of controls (43.2 ± 12.4 years) was significantly lower ($P < 0.01$) than that of breast cancer patients (46.0 ± 10.6 years). Pregnancy, menopausal status, breast feeding, BMI, involuntary tobacco smoking, family history of breast cancer, and

Table 2 SNPs analyzed in this study

Gene name[a]	rs number	Nucleotide exchange[b]	Annotation	MAF[c]	HWE[d]
MTR	3795708	T > C		39.44	0.213
	1770449	A > G		21.40	0.283
	1050993	G > A		21.27	0.344
MTRR	1801394	A > G	Ile22Met	31.13	0.916
	326121	C > T		38.34	0.774
	162049	G > A		48.67	0.652
	10380	C > T	His595Tyr	14.48	0.854
	327592	T > C		15.78	0.609
MTHFR	1801133	C > T	Ala222Val	13.55	0.437
	1801131	A > C	Glu429Ala	25.87	0.025
TYMS	16948322	C > T		31.70	0.212
	2298581	G > C		37.03	0.052
ADH1C	4147542	G > A		28.50	0.262
ALDH2	4646778	C > A		23.20	0.702
	671	G > A	Glu487Lys	9.05	1.000
DRD2	7117915	G > A		36.47	0.922
	10891556	G > T		16.84	0.873
DRD3	2087017	A > G		44.50	0.068
	167770	A > G		20.42	0.400
SLC6A4	140701	A > G		23.88	0.456
GSTP1	612020	C > T		8.57	0.153
	1695	A > G		26.38	0.560
NAT1	7845127	C > T		28.27	1.000
NAT2	1041983	C > T		47.81	0.648
	1799930	G > A	Arg197Gln	31.68	0.020
CYP2E1	2249695	C > T		41.12	0.776
	2031920	C > T		15.96	0.057

[a] According to the NCBI database SNP

[b] More frequent to less frequent allele

[c] Minor allele frequency

[d] *P* value for Hardy–Weinberg equilibrium is calculated in the control population by Fisher's exact test

education were different between cases and controls. However, as to oral contraceptive use, smoking, and alcohol consumption, no significant differences were found between cases and controls.

Frequencies of variant alleles among the control population are shown in Table 2. All SNP frequencies were in Hardy–Weinberg equilibrium (HWE) among controls, except two SNPs in *MTHFR* (rs1801131) and *NAT2* (rs1799930). The result of association analysis of individual SNPs is shown in Table 3. Homozygotes of minor alleles of *MTR* SNPs (rs1770449 and rs1050993) were associated with an increased risk of breast cancer with OR = 2.21 (95% CI 1.18–4.16) and 2.24 (95% CI 1.19–4.22), respectively. The SNP in *DRD3* (rs167770) was found associated with an increased risk among heterozygote carriers (OR 1.36, 95% CI 1.03–1.80). Although the *DRD3* (rs167770) SNP did not show a statistically significant association for the minor allele homozygotes, this SNP was significant when tested for allelic (OR 1.24, 95% CI 1.01–1.53; $P = 0.045$) and recessive (OR 1.38, 95% CI 1.07–1.77; $P = 0.014$) models.

A stratified analysis by menopausal status suggested tendencies that *NAT2* (rs1799930), *DRD2* (rs10891556), and *SLC6A4* (rs140701) polymorphisms are related to breast cancer risk in premenopausal women (OR 2.70, 95% CI 1.20–6.07; OR 1.62, 95% CI 1.03–2.56; and OR 1.56, 95% CI 1.01–2.41, respectively) (Table 4). Among postmenopausal women, an increased risk of breast cancer was suggested for *MTRR* (rs162049) and *DRD3* (rs167770) polymorphisms (OR 1.61, 95% CI 1.07–2.44 and OR 1.59, 95% CI 1.11–2.28, respectively).

Analyses on gene–environment interactions were performed between polymorphisms and alcohol consumption, oral contraceptive use and body mass index. They were exploratory analyses, but no strong interactions were identified (Table 5). Two-gene interactions were also analyzed between SNPs of the *DRD3* (rs167770) and other genes showing significant association individually in Table 3. None of gene–gene interactions was statistically significant (Table 6).

Further, *DRD3*, and *MTR* genotype were correlated with ER and PR status. Fifty-three percent of breast cancer cases were ER-positive tumors (167/314) and 39% were PR-positive tumors (118/301). The associations between *DRD3* (rs167770) and *MTR* (rs1770449 and rs1050993) polymorphisms and breast cancer risk were not different among the ER/PR status (data not shown).

Discussion

Genetic variation in enzymes and other proteins involved in folate and alcohol metabolisms are rational candidates for studying the impact of both genetic and environmental effects and their interactions on breast cancer risk. As a systematic candidate gene approach, we analyzed 12 SNPs in four folate metabolism genes and 15 SNPs in nine alcohol metabolism and behavior genes; these SNPs were selected by our previous study to catalog candidate genes, which are potentially subjected to gene–environment interactions with regard to cancer susceptibilities among Japanese population [15]. Only the SNPs that have minor allele frequencies higher than 5% were evaluated in the study. Of the 27 SNPs, seven suggested an association with breast cancer risk (*MTR* rs1770449 and rs1050993, *MTRR* rs162049, *DRD2* rs10891556, *DRD3* rs167770, *SLC6A4* rs140701, and *NAT2* rs1799930) in an overall or stratified analysis.