Original Article

Potential functional variants of KIAA genes are associated with breast cancer risk in a case control study

Jing Zhou[1,2#], Congcong Chen[1,3#], Sijun Liu[4], Wen Zhou[1,3], Jiangbo Du[1,5], Yue Jiang[1,5], Juncheng Dai[1,3,5], Guangfu Jin[1,3,5], Hongxia Ma[1,3,5], Zhibin Hu[1,3,5], Jiaping Chen[1,3], Hongbing Shen[1,3,5]

[1]Department of Epidemiology, International Joint Research Center on Environment and Human Health, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, China; [2]Statistical Center, Information Department, Northern Jiangsu People's Hospital and Clinical Medical College of Yangzhou University, Yangzhou, China; [3]Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Medicine, Nanjing Medical University, Nanjing, China; [4]Department of Social Medicine and Health Education, School of Public Health, Nanjing Medical University, Nanjing, China; [5]State Key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing, China

Contributions: (I) Conception and design: J Chen, H Shen; (II) Administrative support: G Jin, H Ma, Z Hu, H Shen; (III) Provision of study materials or patients: H Shen; (IV) Collection and assembly of data: J Zhou, C Chen, W Zhou, J Du; (V) Data analysis and interpretation: C Chen, S Liu, J Dai, J Chen; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Jiaping Chen; Hongbing Shen. Department of Epidemiology, School of Public Health, Nanjing Medical University, 101 Long-Mian Road, Nanjing 211166, China. Email: jiapingchen@njmu.edu.cn; hbshen@njmu.edu.cn.

Background: KIAA genes identified in the Kazusa cDNA-sequencing project may play important roles in biological processes and are involved in carcinogenesis of many cancers. Genetic variants of KIAA genes are implicated in the abnormal expression of these genes and are linked to susceptibility of several human complex diseases.

Methods: The differentially expressed KIAA genes were screened and identified in The Cancer Genome Atlas (TCGA) database of breast cancer. A total of 48 variants located in the 28 KIAA genes were selected to investigate the associations between polymorphism and breast cancer in 1,032 cases and 1,063 cancer-free controls in a Chinese population.

Results: Two coding variants, which included a SNP rs2306369 in *KIAA1109* and a SNP rs1205434 in *KIAA1755*, were identified to be associated with the incidences of breast cancer. Logistic regression analysis showed that the SNP rs2306369 G allele was associated with a decreased risk of breast cancer (additive model: OR =0.81, 95% CI: 0.66–0.99, P=0.038), whereas the SNP rs1205434 A allele was involved with a higher risk of breast cancer (additive model: OR =1.19, 95% CI: 1.02–1.38, P=0.025). Further stratified analysis revealed that the SNP rs1205434 showed a significant difference for age at menarche strata (heterogeneity test P=0.009). Multiplicative interaction analysis indicated that there was positive multiplicative interaction between the SNP rs1205434 and menarche age (OR =1.09, 95% CI: 1.01–1.17, P=0.036). Additionally, expression quantitative trait loci analysis revealed that the SNP rs1205434 A allele could decrease the *KIAA1755* expression in the Genotype-Tissue Expression (GTEx) database (P=0.002). The Kaplan-Meier plotter showed that breast cancer patients with high *KIAA1755* expression have significantly better outcomes than those with low levels of expression (HR =0.84, 95% CI: 0.72–0.99, P=0.033).

Conclusions: The results indicate that the genetic variants (rs2306369 and rs1205434) in the coding region of *KIAA1109* and *KIAA1755* respectively may affect Chinese females' breast cancer susceptibility and act as potential predictive biomarkers for breast cancer.

Keywords: KIAA; breast cancer; susceptibility; genetic variants

Submitted Aug 26, 2020. Accepted for publication Jan 12, 2021.

doi: 10.21037/atm-20-6108

View this article at: http://dx.doi.org/10.21037/atm-20-6108

Introduction

Breast cancer, one of the commonest malignancies, is responsible for the highest cancer-related mortality rate in women. In 2018, an estimated 2,088,849 new cases and 626,679 deaths occurred worldwide (1). Breast cancer development involves multiple stages and is influenced by genetics and the environment (2-4). The most prevalent risk factors are those linked to menstruation (menarche at a young age, older age at menopause), reproduction (nulliparity, first birth at a late age, and fewer offspring), artificial hormones (oral contraceptives and hormone-replacement therapies) and alcohol intake (2). Genome-wide association studies (GWASs) are known to be powerful tools for dissecting the genetic architecture of complex diseases. To date, GWASs have identified over 150 loci for breast cancer susceptibility in several populations (3,4). Despite this, the information provided by GWAS remains insufficient. The heritability of breast cancer is estimated to be responsible for only 18% of the familial relative risk of breast cancer (3). Moreover, most of the breast cancer susceptibility SNPs that have been identified in breast cancer GWAS studies are located in intergenic or intronic regions of the genome, which may not generally impact the protein-coding regions as expectations (5,6).

KIAA genes were identified by the Kazusa cDNA-sequencing project. Sequences of human large cDNAs longer than 4 kb which direct synthesis of large proteins (>50 kDa) were characterized and deposited in the *HUGE* database (Human Unidentified Gene-Encoded large protein database) (7,8). Because KIAA genes were screened and identified mainly based upon the length of the mRNA (>4 kb) that encode large proteins (>1,000 amino acid residues), KIAA proteins possess many kinds of biological functions and join in important life activities. Some KIAA proteins are thought to play critical roles in multiple biological processes, such as DNA damage, centriole formation, cell migration and invasion (9-11). Recently, the potential importance of KIAA genes in cancer is beginning to be recognized and studied, which has attracted considerable attention as a target for various biological events in tumorigenesis. A growing number of evidences have shown that abnormal expression of KIAA genes were associated with the development and prognosis of various cancers, such as gastric cancer, esophageal squamous-cell carcinomas, colorectal cancer, lung cancer and breast cancer (12-16). Abnormal increase or loss of KIAA genes expression may contribute to tumor formation and development. Kang *et al*. observed that *KIAA1324* not only inhibited the invasiveness, growth, and tumorigenicity of gastric cancer cells but also blocked oncogenic GRP78 activities to promote apoptosis. Low *KIAA1324* expression was related to poor prognosis in patients with gastric cancer (12). *KIAA1377* has been found to be significantly amplified in esophageal squamous cell carcinoma with lymph node metastasis (13). *KIAA1199* has been identified as an oncogene in many cancers including breast cancer and associated with tumor invasion depth, TNM stage, and poor prognosis (14-16).

In recent year, polymorphisms in KIAA genes have drawn considerable attention in human complex diseases, such as cancers, diabetes, autoimmune diseases and cardiac disease, and extensive efforts have been made to elucidate the mechanisms responsible for their aberrant expressions (17-22). However, up to now, there are few records of the association of single nucleotide polymorphisms (SNPs) in KIAA genes and breast cancer risk. In this study, we hypothesized that there is an association between polymorphisms of crucial KIAA genes and the incidence of breast cancer in the Chinese population. To test this hypothesis, we identified a number of differentially expressed KIAA genes potentially associated with breast cancer from The Cancer Genome Atlas (TCGA) research database, and adopted a case-control design to investigate the occurrence of the SNPs among Chinese females (1,032 cases and 1,063 controls). We expected to accumulate useful data to reveal the correlation between genetic polymorphisms of key KIAA genes and breast cancer risk.

We present the following article in accordance with the MDAR checklist and STROBE reporting checklist (available at http://dx.doi.org/10.21037/atm-20-6108).

Methods

Study subjects

This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and received approval from the Institutional Review Board of Nanjing Medical University (FWA00001501). A total of 1,032 cases with clinically confirmed breast cancer and 1,063 cancer-free controls were included in the association analysis. All study participants were Chinese Han females, genetically unrelated to each other. The study design has been described in detail in a previous publication (23). Briefly, we included female patients who were histologically

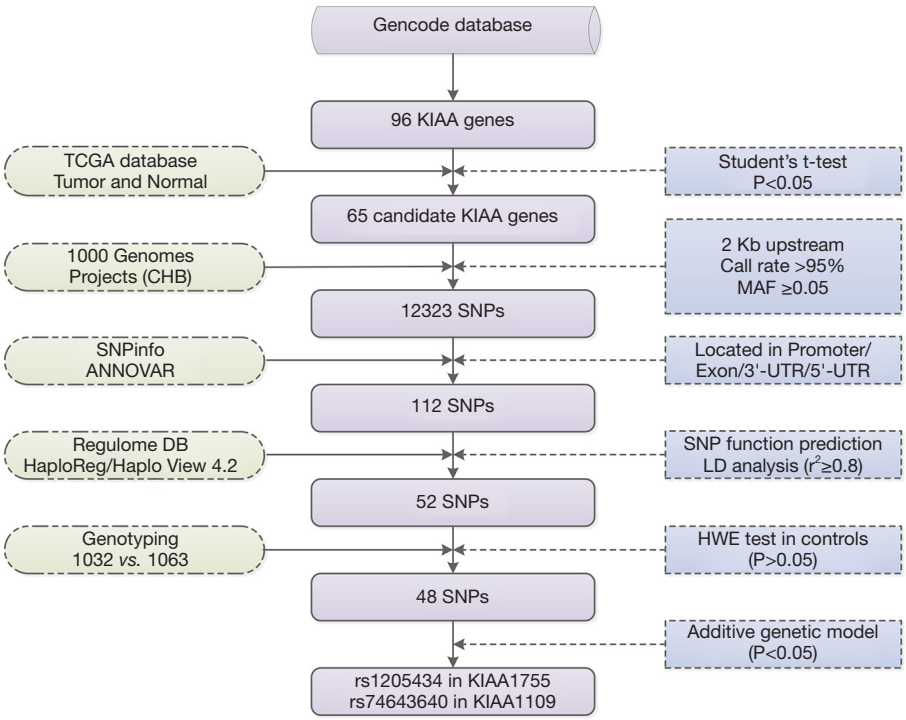

Figure 1 Flow chart for selecting SNPs in the KIAA genes. MAF, minor allele frequency; LD, linkage disequilibrium; UTR, untranslated region; CHB, Han Chinese in Beijing, China.

diagnosed with breast cancer at the First Affiliated Hospital of Nanjing Medical University, Cancer Hospital of Jiangsu Province, and Gulou Hospital, between 2004 and 2010. Controls were recruited from a noninfectious diseases' community screening program between 2004 and 2006 in Jiangsu Province. We obtained informed consent from each participant before recruitment. Each participant was interviewed using a structured questionnaire to collect information on age, menstrual and reproductive history, and environmental exposure history. Clinicopathologic data comprising estrogen-receptor (ER) and progesterone-receptor (RP) statuses of patients were collected from their medical records. The demographic distribution in terms of ages between cases and controls was similar (±5 years).

SNP selection and genotyping

A total of 96 KIAA genes were selected from the Gencode databases (Table S1). Then we detected 65 of 96 KIAA genes with differential expression levels in normal samples compared with tumor samples using the breast cancer TCGA database (Student's t-test, P<0.05) (Table S2).

A detailed workflow chart of the strategy has been stated as *Figure 1*. We screened the SNPs in 65 candidate KIAA genes from the 1000 Genomes Project according to the following criteria: (I) Population data for the Chinese Han population of Beijing in the 1000 Genomes Project was used to screen SNPs within the candidate KIAA gene regions and 2 kb upstream. A total of 12,323 SNPs remained after the following two quality control criteria were applied: call rate >95% and minor allele frequency ≥0.05. (II) We excluded SNPs that were not within a promoter, exon, or 3′-/5′-untranslated region. (III) HaploView 4.2 software was employed to extract SNPs that had linkage disequilibrium in the Chinese Han population (r^2 ≥0.8). HaploReg (http://archive.broadinstitute.org/mammals/haploreg/haploreg.php) and RegulomeDB (http://regulome.stanford.edu/index) were used to evaluate SNPs for regulatory elements and potential functions.

In total, 52 SNPs located in KIAA gene regions were selected for further genotyping. We performed genotyping using the Illumina HumanExome Beadchip array (Illumina

Inc.) in 1,064 breast cancer cases and 1,125 cancer-free controls. Genotyping was performed without knowledge of the individual's case or control status, and approximately equal numbers of case and control samples were tested during each assay, with two blank controls. Genotype calling was performed using Illumina's GenTrain clustering algorithm (version 1.0) in GenomeStudio (V2011.1). Thirty-two cases and sixty-two control subjects were excluded by filtering unqualified genetic variants and genetic relatedness. The study design has been described in detail in a previous publication (23). Finally, 1,032 breast cancer cases and 1,063 cancer-free controls were genotyped for SNPs by Illumina HumanExome Beadchip array. Meanwhile, Hardy-Weinberg equilibrium was used to filter SNPs with P values less than 0.05 in control groups and 4 SNPs were excluded. Lastly, 48 SNPs located in the 28 KIAA genes were left for further analysis.

Statistical analyses

For each SNP, a χ^2 test using one degree of freedom was employed to examine Hardy-Weinberg equilibrium (HWE) in controls. We used a χ^2 test and student's t-test to calculate the differences between the chosen categorical and continuous variables, respectively, the genotypic frequency of each of the 52 SNPs in the controls and cases, and differences in the distribution of demographic characteristics. Confidence intervals (CIs, 95%) and odds ratios (ORs) were established using logistical regression for the associations between a given SNP and breast cancer risk adjusted for menopausal status, age at menarche, as well as current age. Heterogeneity of associations between subgroups was assessed by the χ^2-based Q test. Interaction analysis was performed by adding an interaction item in logistic regression model. All statistical analyzes were performed with R software (version 3.2.3) for Statistical Computing. Statistical significance was indicated by a two-sided P value of <0.05.

Bioinformatics analysis

SNPs influencing the protein stability of KIAA genes were analyzed using I-Mutant (http://folding.biofold.org/i-mutant/i-mutant2.0.html) and MUpro (http://mupro.proteomics.ics.uci.edu/).

We assessed the eQTL of the candidate SNPs using the Gene-Tissue Expression (GTEx) project (http://www.gtexportal.org/). Fully processed, filtered, and normalized gene expression matrices (in BED format), as well as covariates, were download from the GTEx website. Then, genotype information from the GTEx Analysis V8 (dbGaP Accession phs000424.v8.p2) was applied from the dbGap database for eQTL analysis.

Online Kaplan-Meier plotter

The Kaplan-Meier plotter (http://kmplot.com/analysis/) was used to investigate how KIAA-gene expression affected the survival of breast cancer patients. We obtained the expression data for the candidate genes and overall patient survival data from the public databases. We analyzed the gene's prognostic values by allocating the patient samples into two groups, which were then compared using Kaplan-Meier curves. The hazard ratios (HRs), the corresponding 95% confidence intervals (CIs), and log-rank P values were generated.

Results

Characteristics of the study population

The demographic characteristics of 1,032 breast cancer cases and 1,063 cancer-free controls are described in Table S3. There were no significant differences in the distribution of ages between the cancer and control groups (P=0.078). The analysis showed that women with breast cancer went through menarche at a significantly younger age than the control group (P<0.0001), and birthed their first live child at a significantly elder age than the control group (P<0.0001).

Association between candidate SNPs in KIAA genes regions and breast cancer risk

This study primarily aimed to investigate the possible associations between 48 predominant SNPs (r^2<0.8) in 28 candidate KIAA genes and the risk of breast cancer. Through the additive model, we found two SNPs, rs1205434 and rs2306369, that were significantly related to breast cancer risk in our population (P<0.05) (*Table 1*). According to the additive model, there was a significant association between rs1205434 in the *KIAA1755* coding region and increased breast cancer incidence (rs1205434 adjusted OR =1.19, 95% CI: 1.02–1.38, P=0.025). In contrast, the model showed there was a significant association between rs2306369 in the *KIAA1109* coding region and a decrease in breast cancer incidence (rs2306369 adjusted OR =0.81,

Table 1 Associations between selected variants in KIAA gene family members and breast cancer risk

SNP	Gene	Func annot	Allele[a]	Call Rate	MAFb (case/control)	HWE[c] (control)	Cases[d] (N=1,032)	Controls[d] (N=1,063)	OR (95% CI)[e]	P value[e]
rs351260	KIAA0141	Missense	G/A	100.00	0.15/0.16	1.000	742/261/29	742/293/28	0.94 (0.79–1.11)	0.439
rs61742251	KIAA0226	Missense	A/G	100.00	0.04/0.04	1.000	947/82/3	979/83/1	0.99 (0.72–1.35)	0.927
rs1408184	KIAA0226L	Missense	A/G	100.00	0.45/0.43	0.060	330/483/219	366/489/208	1.07 (0.94–1.21)	0.303
rs16889440	KIAA0319	3'UTR	G/A	100.00	0.11/0.10	0.381	823/195/14	864/192/7	1.10 (0.90–1.36)	0.359
rs807534	KIAA0319	Missense	A/G	100.00	0.07/0.08	0.055	894/136/2	898/163/2	0.81 (0.63–1.04)	0.102
rs4504469	KIAA0319	Missense	G/A	100.00	0.17/0.15	0.398	720/283/29	773/263/27	1.13 (0.95–1.34)	0.167
rs2287790	KIAA0556	Missense	G/A	99.95	0.07/0.07	0.824	888/138/5	912/145/6	0.98 (0.77–1.25)	0.861
rs72771666	KIAA0825	Missense	G/A	99.86	0.06/0.05	1.000	917/111/2	950/109/3	1.04 (0.79–1.38)	0.764
rs29910	KIAA0825	Missense	G/A	99.90	0.06/0.06	0.786	901/126/3	940/119/4	1.11 (0.85–1.44)	0.443
rs17370297	KIAA0922	Missense	A/G	100.00	0.23/0.23	0.664	617/359/56	635/370/58	0.99 (0.85–1.15)	0.895
rs12485064	KIAA0930	5'UTR	G/A	99.95	0.39/0.36	0.689	367/528/136	435/495/133	1.13 (0.99–1.29)	0.081
rs11634652	KIAA1024	Missense	C/A	100.00	0.09/0.09	1.000	856/167/9	883/172/8	1.06 (0.85–1.32)	0.635
rs2297773	KIAA1024	Missense	A/G	100.00	0.16/0.16	1.000	736/272/24	753/284/26	0.99 (0.83–1.17)	0.886
rs2306369	KIAA1109	Missense	A/G	100.00	0.10/0.12	0.318	833/188/11	814/237/12	0.81 (0.66–0.99)	0.038
rs2297776	KIAA1161	Missense	A/T	100.00	0.09/0.09	0.847	854/172/6	884/172/7	1.03 (0.82–1.28)	0.824
rs12441101	KIAA1199	Missense	A/G	100.00	0.09/0.09	0.474	850/171/11	869/187/7	0.99 (0.80–1.23)	0.930
rs79868722	KIAA1211	Missense	A/G	100.00	0.23/0.21	0.709	613/364/55	667/353/43	1.13 (0.97–1.31)	0.127
rs6823339	KIAA1211	Missense	C/A	98.95	0.07/0.06	0.117	882/136/4	920/130/1	1.08 (0.83–1.39)	0.580
rs3796547	KIAA1211	Missense	G/A	100.00	0.27/0.27	0.875	558/392/82	571/418/74	1.06 (0.92–1.22)	0.443
rs10828663	KIAA1217	Missense	G/A	99.62	0.15/0.13	0.592	743/262/24	794/248/16	1.14 (0.95–1.37)	0.153
rs16924863	KIAA1217	Synonymous	G/C	100.00	0.04/0.04	0.138	956/75/1	991/69/3	1.17 (0.84–1.63)	0.366
rs4634233	KIAA1239	Missense	C/A	100.00	0.07/0.06	0.411	895/128/9	942/116/5	1.19 (0.93–1.53)	0.171
rs12686794	KIAA1432	Synonymous	C/G	100.00	0.05/0.06	1.000	937/94/1	949/111/3	0.83 (0.63–1.11)	0.217
rs3739308	KIAA1456	Missense	A/G	100.00	0.04/0.05	1.000	951/80/1	969/92/2	0.87 (0.64–1.19)	0.390
rs502882	KIAA1456	Synonymous	A/G	100.00	0.22/0.23	0.794	632/342/58	633/377/53	1.00 (0.86–1.16)	0.978
rs3739998	KIAA1462	Synonymous	G/C	100.00	0.22/0.20	0.850	634/349/49	675/343/45	1.09 (0.94–1.27)	0.262
rs2185724	KIAA1462	Synonymous	A/G	100.00	0.27/0.28	0.706	546/407/79	548/427/88	0.98 (0.85–1.13)	0.798
rs117872916	KIAA1522	Missense	G/A	100.00	0.12/0.11	1.000	804/214/14	843/208/12	1.08 (0.88–1.31)	0.472
rs2278911	KIAA1524	Missense	A/G	100.00	0.42/0.42	0.315	343/513/176	344/536/183	0.99 (0.87–1.12)	0.864
rs60797311	KIAA1549	Missense	G/A	99.95	0.09/0.09	0.331	855/167/9	885/173/5	1.02 (0.82–1.28)	0.842
rs2774960	KIAA1549	Missense	G/A	99.90	0.22/0.23	0.388	627/356/47	625/387/51	0.90 (0.77–1.05)	0.165
rs2718131	KIAA1549	Synonymous	C/G	99.95	0.10/0.11	0.274	833/190/8	839/215/9	0.88 (0.72–1.09)	0.241
rs2273117	KIAA1549L	Synonymous	G/A	100.00	0.35/0.35	0.067	435/471/126	442/508/113	1.00 (0.87–1.14)	0.970
rs1033543	KIAA1549L	Synonymous	G/A	100.00	0.28/0.29	0.367	531/428/73	536/447/80	0.97 (0.84–1.12)	0.706

Table 1 (*continued*)

Table 1 (*continued*)

SNP	Gene	Func annot	Allele[a]	Call Rate	MAFb (case/control)	HWE[c] (control)	Cases[d] (N=1,032)	Controls[d] (N=1,063)	OR (95% CI)[e]	P value[e]
rs6926980	KIAA1586	Missense	G/A	100.00	0.12/0.12	0.881	785/239/8	831/217/15	1.08 (0.88–1.31)	0.466
rs61740375	KIAA1586	Missense	A/G	100.00	0.10/0.09	0.693	838/190/4	887/170/6	1.16 (0.92–1.45)	0.203
rs3795504	KIAA1614	Synonymous	C/A	100.00	0.25/0.23	0.230	578/388/66	632/366/65	1.14 (0.98–1.31)	0.084
rs12608744	KIAA1683	Synonymous	G/A	99.71	0.19/0.19	0.076	656/347/27	698/312/49	1.00 (0.86–1.18)	0.960
rs2277921	KIAA1683	Missense	G/A	99.95	0.34/0.36	0.255	427/504/100	433/505/125	0.95 (0.83–1.09)	0.502
rs8110972	KIAA1683	Missense	T/A	99.95	0.46/0.45	1.000	276/556/199	322/527/214	1.04 (0.92–1.19)	0.535
rs1079166	KIAA1683	5'UTR	A/G	98.62	0.39/0.41	0.949	375/492/149	372/505/173	0.93 (0.81–1.06)	0.255
rs3746471	KIAA1755	Missense	G/A	99.90	0.41/0.39	0.607	353/503/175	390/514/158	1.11 (0.98–1.26)	0.109
rs760998	KIAA1755	Missense	G/A	99.95	0.50/0.48	0.759	245/534/252	279/537/247	1.09 (0.96–1.24)	0.173
rs16987188	KIAA1755	Missense	G/A	100.00	0.04/0.04	1.000	943/86/3	970/91/2	0.97 (0.72–1.31)	0.849
rs1205434	KIAA1755	Missense	C/A	99.43	0.24/0.21	1.000	588/381/58	655/354/47	1.19 (1.02–1.38)	0.025
rs2290477	KIAA2018	Missense	C/A	100.00	0.14/0.13	0.502	759/256/17	804/238/21	1.10 (0.91–1.32)	0.334
rs9866806	KIAA2018	Missense	G/C	100.00	0.22/0.21	0.135	618/364/50	676/333/54	1.12 (0.96–1.30)	0.142
rs77442027	KIAA2026	Missense	G/A	100.00	0.05/0.05	1.000	925/106/1	964/97/2	1.07 (0.80–1.43)	0.650

[a]Major/minor allele; [b]Minor allele frequency (MAF); [c]P values for the Hardy-Weinberg equilibrium (HWE) test; [d]Major homozygote/heterozygote/rare homozygote between cases and controls; [e]Logistic regression analysis with adjustment for age, age at menarche, menopausal status and two principle components in the additive model. Func annot, function annotation; OR, odds ratio; CI, confidence interval.

Table 2 Associations between the two SNPs (rs1205434, rs2306369) and breast cancer risk in co-dominant and dominant genetic models

Genotype	Case	Control	OR (95% CI)[a]	P[a]
SNP rs1205434 (C>A)				
CC	588	655	1	
AC	381	354	1.20 (0.99, 1.45)	0.062
AA	58	47	1.38 (0.91, 2.10)	0.126
Dominant model			1.22 (1.02, 1.46)	0.033
SNP rs2306369 (A>G)				
GG	833	814	1	
GA	188	237	0.78 (0.62, 0.97)	0.028
AA	11	12	0.87 (0.37, 2.03)	0.745
Dominant model			0.78 (0.63, 0.97)	0.028

[a]OR with its 95% CI and P values of co-dominant model, dominant model were derived from logistic regression adjusted for age, age at menarche, menopause status and two principle components. OR, odds ratio; CI, confidence interval.

95% CI: 0.66–0.99, P=0.038). All statistical analyzes were adjusted for possible influential confounders, i.e., age, age at menarche, and menopausal status.

Additionally, under the dominant and codominant models, correlations between the risk of breast cancer and the two SNPs (rs1205434 and rs2306369) were also calculated. *Table 2* lists further details. The results showed that the SNP rs1205434 was significantly correlated with increased breast cancer individual risk under the dominant model (OR =1.22, 95% CI: 1.02–1.46, P=0.033); whereas the SNP rs2306369 was significantly correlated with decreased breast cancer individual risk under both the codominant and dominant models (codominant: OR =0.78, 95% CI: 0.62–0.97, P=0.028; dominant: OR =0.78, 95% CI: 0.63–0.97, P=0.028).

We also performed stratification analyses of rs1205434 and rs2306369 based on the age group, age at menarche, age at first live birth, age at natural menopause, and menopause status. As shown in *Table 3*, the SNPs rs1205434 and rs2306369 exhibited different effects among women with later natural menopause and at different menopause stages (P<0.05). Additionally, rs1205434 showed a significant difference for age at menarche strata (heterogeneity test P=0.009). Further multiplicative interaction analysis indicated that there may exist an interaction effect between rs1205434 and menarche age (*Table 4*). Women who both carry CA/AA genotype of rs1205434 and menarche over 16 years would prone to develop breast cancer (OR =1.09, 95% CI: 1.01–1.17, P=0.036).

Potential functional impact of KIAA genes variants

In silico annotation showed that the SNPs for *KIAA1109* (rs2306369) and *KIAA1755* (rs1205434) were non-synonymous. The two SNPs were identified as those that decrease protein stability using I-Mutant and Mupro online bioinformatics analyzes (Table S4).

We examined if the two SNPs were eQTLs in the GTEx database. As shown in *Figure 2*, the SNP rs1205434 was in the eQTL for *KIAA1755* ($P=2.2\times10^{-3}$) in normal breast tissue. While, according to eQTL analysis, rs2306369 was not significantly associated with the expression of its host gene *KIAA1109* and nearby genes (±1MB).

KIAA gene expression and breast cancer survival

We assessed the relationships between KIAA gene expression and the prognosis of patients with breast cancer using the Kaplan-Meier plotter (http://kmplot.com/analysis/), which analyzes a combination of follow-up information and gene expression data (24). Patients were allocated to high- and low-expression groups based on KIAA gene expression mean values. The log-rank of the Kaplan Meier curve showed that the outcomes of patients with breast cancer who had high *KIAA1755* expression were significantly better than those who had low *KIAA1755* expression (HR =0.84, 95% CI: 0.72–0.99, P=0.033), while *KIAA1109* expression showed no significant differences between the two groups (*Figure 3*).

Discussion

In this case-control study, we investigated 48 SNPs of KIAA genes linked to incidences of breast cancer risk. Two coding missense variant SNPs (rs2306369 and rs1205434) in the coding regions of *KIAA1109* and *KIAA1755* respectively were significantly associated with susceptibility to breast cancer in a Han Chinese population. Moreover, we found that rs1205434 genotype correlated with *KIAA1755* expression by eQTL analysis. Additionally, elevated *KIAA1755* expression may be associated with better breast cancer patient outcomes through a Kaplan-Meier plotter-based analysis.

The SNP rs2306369 is a coding variant (encoding p. Thr4440Ala) of the *KIAA1109* gene. *KIAA1109*, which has been named Tweek, is conserved in many eukaryotes, from nematodes to vertebrates, although little is known of its function. *KIAA1109* deficiency was found to be lethal in mice, who died prior to weaning; one possible function of the protein is in synaptic vesicle recycling, as has been found in Drosophila (25). Gueneau *et al*. reported finding *KIAA1109* variants in individuals with severe brain-development disorders and arthrogryposis, and the report discusses the possible involvement of *KIAA1109* in cell-cycle control mechanisms, especially those of the central nervous system (26). Additionally, mutations in *KIAA1109* have been linked to survival in patients with endometrial cancer and esophageal squamous-cell carcinomas. Furthermore, bioinformatics studies showed KIAA1109 was involved in regulating NIK/NF-kappaB signaling (27,28).

KIAA1109 is encoded on chromosomal region 4q27; this region and the KIAA1109-interleukin 2 (IL2)-IL21 block, in particular, was identified as a possible locus of risk in the development of a number of common inflammatory disorders, such as type I diabetes, ulcerative colitis, systemic lupus erythematosus, celiac disease, Chron's

Table 3 The associations of two variants with breast cancer risk in subgroups divided by characteristics

Subgroups	rs1205434				P_{het}[b]	rs2306369				P_{het}[b]
	Case	Control	OR (95% CI)[a]	P[a]		Case	Control	OR (95% CI)[a]	P[a]	
Age (years)										
≤51	345/224/33	314/190/20	1.12 (0.92–1.38)	2.66E-01	0.458	495/106/4	406/118/5	0.79 (0.59–1.04)	8.95E-02	0.680
>51	243/157/25	341/164/27	1.26 (1.01–1.57)	3.95E-02		338/82/7	408/119/7	0.85 (0.64–1.14)	2.82E-01	
Age at menarche (years)										
≤16	460/281/40	377/226/29	1.04 (0.87–1.24)	6.91E-01	0.009	633/144/8	490/140/8	0.8 (0.63–1.01)	6.32E-02	0.958
>16	128/100/18	278/128/18	1.59 (1.22–2.07)	5.67E-04		200/44/3	324/97/4	0.79 (0.54–1.13)	2.00E-01	
Age at first live birth (years)										
≤24	213/118/22	300/169/23	1.06 (0.84–1.34)	6.17E-01	0.178	287/65/3	376/113/6	0.76 (0.55–1.04)	8.64E-02	0.614
>24	340/236/32	321/162/22	1.32 (1.07–1.64)	1.14E-02		492/111/8	388/114/6	0.84 (0.64–1.11)	2.24E-01	
Age at natural menopause (years)										
≤49	91/66/6	142/81/10	1.13 (0.79–1.61)	4.95E-01	0.160	142/21/1	190/41/4	0.66 (0.38–1.09)	1.10E-01	0.985
>49	138/107/16	191/80/14	1.58 (1.17–2.13)	2.68E-03		209/50/3	203/79/4	0.65 (0.44–0.95)	2.81E-02	
Menopausal status										
Premenopausal	300/166/30	284/161/21	1.04 (0.84–1.3)	6.97E-01	0.815	401/95/3	362/105/3	0.85 (0.63–1.15)	2.93E-01	0.224
Natural menopausal	235/179/22	263/127/18	1.39 (1.1–1.77)	6.62E-03		360/73/5	305/100/6	0.65 (0.48–0.88)	6.55E-03	
Unnatural menopausal	42/26/4	7/5/0	1.41 (0.5–4.74)	5.44E-01		54/16/2	11/1/0	5.04 (0.89–104.45)	1.50E-01	
ER status										
Positive	273/181/24		1.14 (0.94–1.38)	1.72E-01	0.519	390/85/4	814/237/12	0.75 (0.57–0.97)	2.91E-02	0.322
Negative	206/139/25		1.25 (1.02–1.53)	3.09E-02		296/72/5	814/237/12	0.9 (0.69–1.18)	4.72E-01	
PR status										
Positive	280/182/28		1.15 (0.95–1.39)	1.46E-01	0.569	398/87/7	814/237/12	0.81 (0.63–1.04)	1.08E-01	0.944
Negative	198/139/21		1.25 (1.01–1.53)	3.53E-02		288/70/2	814/237/12	0.82 (0.62–1.09)	1.79E-01	

[a]Derived from the logistic regression model after adjusting for age, age at menarche and menopausal status as appropriate assuming an additive genetic model. [b]Heterogeneity test between subgroups. OR, odds ratio; CI, confidence interval; ER, estrogen receptor; PR, progesterone receptor.

Table 4 Multiplicative interaction analysis of rs1205434 and menarche age for breast cancer risk

Variables	rs1205434 (cases/controls)						OR (95% CI)[a]	P[a]
	CC genotype		CA genotype		AA genotype			
	N	%	N	%	N	%		
Age at menarche								
≤16	460/377	54.96/46.04	281/226	55.42/44.58	40/29	57.97/42.03	1.04 (0.87–1.24)	6.91E-01
>16	128/278	31.53/68.47	100/128	43.86/56.14	18/18	50.00/50.00	1.59 (1.22–2.07)	5.67E-04
SNP × Menarche Age							1.09 (1.01–1.17)	3.56E-02

[a]Derived from the logistic regression model after adjusting for age, age at menarche and menopausal status. OR, odds ratio; CI, confidence interval.

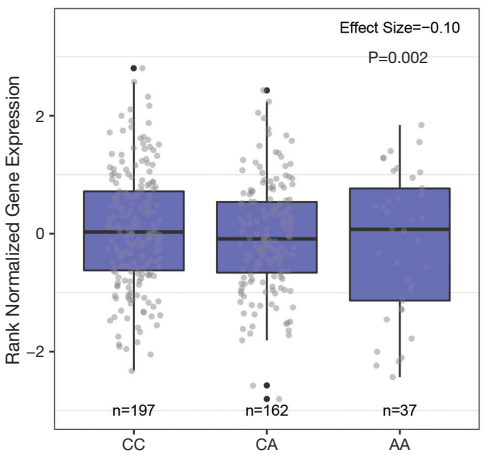

Figure 2 eQTL analysis of the rs1205434 in breast mammary tissues with the Gene-Tissue Expression (GTEx) database. Boxplots showing the effects of the genotypes of rs1205434 on *KIAA1755* expression levels (P=0.002).

disease, psoriasis, and rheumatoid arthritis, via GWASs and candidate gene studies (19-21,29-32). In addition, the SNP rs13151961 in the intron of *KIAA1109* gene has been implicated in male susceptibility to prostate cancer in those with a family history of the disease (17). The results reported herein reveal a significant association between the coding variant SNP rs2306369 and a decreased risk of breast cancer. Moreover, we found that the risk effects of rs2306369 variant genotypes were statistically significant in some subgroups, such as in women with later menopause, natural menopause, and ER-positive individuals. MUpro and I-Mutant software predicted that the protein structure stability of KIAA1109 would be decreased by the missense variant rs2306369. Detailed functional assays are required to validate these findings.

The missense variant rs1205434 is located within the *KIAA1755* gene and encodes p.Lys339Asn. *KIAA1755* is predicted to encode an uncharacterized protein and little known its function. Antony *et al*. found that *KIAA1755* is a tumor-specific antigen found in the ascitic fluid of women with ovarian cancer (33). In this study, we investigated *KIAA1755* expression and its effect on breast cancer prognosis through the online Kaplan-Meier plotter. The results showed that patients with high expression of *KIAA1755* have relatively improved survival rates. Furthermore, levels of *KIAA1755* were lower in tumors compared with in adjacent normal tissue, according to the TCGA breast cancer database. The results suggest *KIAA1755* may play a tumor-suppressor role in breast cancer.

In the current literature, some reports state that *KIAA1755* genetic variants may be associated with heart rate and hypertension. For example, the SNP rs6123471 in the 3′-UTR of KIAA1755 was associated, at a genome-wide significance level, with heart rate variability in individuals of European ancestry (22). Exome chip meta-analyses identified a relationship between a low-frequency and non-synonymous single-nucleotide variant, rs41282820, at locus *KIAA1755*, and heart rate (34). Loss-of-function mutations in the *KIAA1755* DNA sequence elevated eicosapentaenoate levels. Eicosapentaenoate, an essential fatty acid, has been found to increase essential hypertension in patients (35). The results of the present study revealed a significant association between non-synonymous SNP rs1205434 and an increase in the incidence of breast cancer. Moreover, we found that the risk effects of rs1205434 variant genotypes were statistically significant in some subgroups, such as older women and those with later menarche, later at first

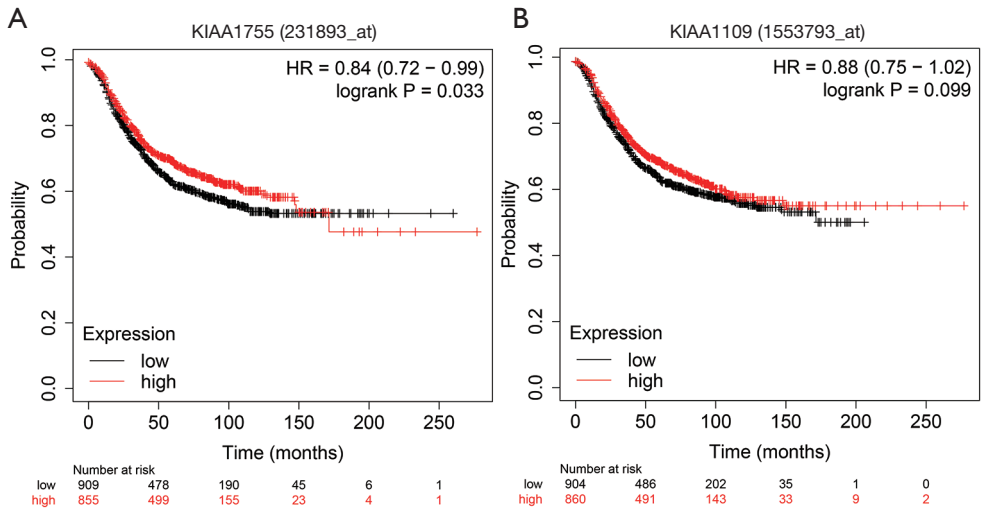

Figure 3 Kaplan-Meier survival curve of breast cancer patients according to the mRNA expression level of *KIAA1755* and *KIAA1109*. (A) The Kaplan Meier curve shows that breast cancer patients with high expression of *KIAA1755* had significant better outcome than those with low expression (P=0.033); (B) the Kaplan Meier curve shows that the outcome of breast cancer patients had no significant differences between the high expression of KIAA1109 group and the low expression of *KIAA1109* group

live birth, later menopause, etc. Additionally, multiplicative interaction analysis indicated that there may exist an interaction effect between rs1205434 and menarche age. According to the online bioinformatics analysis, rs1205434 (Lys339Asn) showed benign amino acid changes, and it probably damages KIAA1755 stability. Furthermore, compared with the C allele of rs1205434, the A allele showed a significant association with decreased KIAA1755 mRNA levels in mammary tissues with eQTL analysis. Together these results thus suggested that the rs1205434 may be a potential predictive biomarker for breast cancer occurrence, although further functional assays will be needed to confirm this result.

In conclusion, we studied the relationships between 48 SNPs of KIAA genes and the incidence of breast cancer in a population of Chinese Han women. Our study showed that the representative genetic variants rs2306369 and rs1205434 in the respective coding regions of *KIAA1109* and *KIAA1755* are candidate SNPs for use as markers for breast cancer susceptibility in this population. However, a few limitations of this study should be addressed. Firstly, we conducted one stage case control study with 1032 breast cancer cases and 1063 healthy controls. Studies with large sample size and more diverse different populations are needed to validate the results. Secondly, although we found bioinformatics evidence, the biological effect of the two SNPs (rs2306369 and rs1205434) on *KIAA1109* and *KIAA1755* expression and function were not investigated in this study. Further studies, incorporating functional evaluations, of diverse ethnic populations are warranted to confirm these findings.

Acknowledgments

We thank the study participants and research staff for their contributions and commitment to this study.
Funding: This study was supported by grants from the National Natural Science Foundation of China (81521004), the Priority Academic Program for the Development of Jiangsu Higher Education Institutions (Public Health and Preventive Medicine), and the Top-notch Academic Programs Project of Jiangsu Higher Education Institutions (PPZY2015A067).

Footnote

Reporting Checklist: The authors have completed the MDAR checklist and STROBE reporting checklist. Available at http://dx.doi.org/10.21037/atm-20-6108

Data Sharing Statement: Available at http://dx.doi.org/10.21037/atm-20-6108

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/

atm-20-6108). The authors declare that they have no conflict of interest regarding the publication of this paper.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Institutional Review Board of Nanjing Medical University (FWA00001501) and informed consent was taken from each participant before recruitment.

References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394-424.
2. Baselga J, Norton L. Focus on breast cancer. Cancer Cell 2002;1:319-22.
3. Milne RL, Kuchenbaecker KB, Michailidou K, et al. Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. Nat Genet 2017;49:1767-78.
4. Michailidou K, Lindström S, Dennis J, et al. Association analysis identifies 65 new breast cancer risk loci. Nature 2017;551:92-4.
5. Fachal L, Aschard H, Beesley J, et al. Fine-mapping of 150 breast cancer risk regions identifies 191 likely target genes. Nat Genet 2020;52:56-73.
6. Ferreira MA, Gamazon ER, Al-Ejeh F, et al. Genome-wide association and transcriptome studies identify target genes and risk loci for breast cancer. Nat Commun 2019;10:1741.
7. Suyama M, Nagase T, Ohara O. HUGE: a database for human large proteins identified by Kazusa cDNA sequencing project. Nucleic Acids Res 1999;27:338-9.
8. Kikuno R, Nagase T, Nakayama M, et al. HUGE: a database for human KIAA proteins, a 2004 update integrating HUGEppi and ROUGE. Nucleic Acids Res 2004;32:D502-4.
9. Polato F, Rusconi P, Zangrossi S, et al. DRAGO (KIAA0247), a New DNA Damage–Responsive, p53-Inducible Gene That Cooperates With p53 as Oncosuppressor. J Natl Cancer Inst 2014;106:dju053.
10. Evensen NA, Kuscu C, Nguyen HL, et al. Unraveling the Role of KIAA1199, a Novel Endoplasmic Reticulum Protein, in Cancer Cell Migration. J Natl Cancer Inst 2013;105:1402-16.
11. Tsuchiya Y, Yoshiba S, Gupta A, et al. Cep295 is a conserved scaffold protein required for generation of a bona fide mother centriole. Nat Commun 2016;7:12567.
12. Kang JM, Park S, Kim SJ, et al. KIAA1324 Suppresses Gastric Cancer Progression by Inhibiting the Oncoprotein GRP78. Cancer Res 2015;75:3087-97.
13. Zheng ST, Yang CC, Liu Q, et al. KIAA1377 is associated with lymph node metastasis in esophageal squamous cell carcinoma. Oncol Lett 2016;12:5223-8.
14. Xu J, Liu Y, Wang X, et al. Association between KIAA1199 overexpression and tumor invasion, TNM stage, and poor prognosis in colorectal cancer. Int J Clin Exp Pathol 2015;8:2909-18.
15. Jami MS, Hou J, Liu M, et al. Functional proteomic analysis reveals the involvement of KIAA1199 in breast cancer growth, motility and invasiveness. BMC Cancer 2014;14:194.
16. Deng F, Lei J, Zhang X, et al. Overexpression of KIAA1199: An independent prognostic marker in nonsmall cell lung cancer. J Cancer Res Ther 2017;13:664-8.
17. Tindall EA, Hoang HN, Southey MC, et al. The 4q27 locus and prostate cancer risk. BMC Cancer 2010;10:69.
18. Cheng I, Plummer SJ, Neslund-Dudas C, et al. Prostate cancer susceptibility variants confer increased risk of disease progression. Cancer Epidemiol Biomarkers Prev.2010;19:2124-32.
19. Zhernakova A, Alizadeh BZ, Bevova M, et al. Novel association in chromosome 4q27 region with rheumatoid arthritis and confirmation of type 1 diabetes point to a general risk locus for autoimmune diseases. Am J Hum Genet 2007;81:1284-8.
20. Sawalha AH, Kaufman KM, Kelly JA, et al. Genetic association of interleukin-21 polymorphisms with systemic lupus erythematosus. Ann Rheum Dis 2008;67:458-61.
21. Liu Y, Helms C, Liao W, et al. A genome-wide association

study of psoriasis and psoriatic arthritis identifies new disease loci. PLoS Genet 2008;4:e1000041.
22. Nolte IM, Munoz ML, Tragante V, et al. Genetic loci associated with heart rate variability and their effects on cardiac disease risk. Nat Commun 2017;8:15805.
23. Zhou W, Jiang Y, Zhu M, et al. Low-frequency nonsynonymous variants in FKBPL and ARPC1B genes are associated with breast cancer risk in Chinese women. Mol Carcinog 2017;56:774-80.
24. Lánczky A, Nagy A, Bottai G, et al. For a general citation of the KM-plotter, please use: miRpower: a web-tool to validate survival-associated miRNAs utilizing expression data from 2,178 breast cancer patients. Breast Cancer Res Treat 2016;160:439-46.
25. Verstreken P, Ohyama T, Haueter C, et al. Tweek, an evolutionarily conserved protein, is required for synaptic vesicle recycling. Neuron 2009;63:203-15.
26. Gueneau L, Fish RJ, Shamseldin HE, et al. KIAA1109 Variants Are Associated with a Severe Disorder of Brain Development and Arthrogryposis. Am J Hum Genet 2018;102:116-32.
27. Qing T, Zhu S, Suo C, et al. Somatic mutations in ZFHX4 gene are associated with poor overall survival of Chinese esophageal squamous cell carcinoma patients. Sci Rep 2017;7:4951.
28. Qiao Z, Jiang Y, Wang L, et al. Mutations in KIAA1109, CACNA1C, BSN, AKAP13, CELSR2, and HELZ2 Are Associated With the Prognosis in Endometrial Cancer. Front Genet 2019;10:909.
29. van Heel DA, Franke L, Hunt KA, et al. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. Nat Genet 2007;39:827-9.
30. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661-78.
31. Festen EA, Goyette P, Scott R, et al. Genetic variants in the region harbouring IL2/IL21 associated with ulcerative colitis. Gut 2009;58:799-804.
32. Hollis-Moffatt JE, Gearry RB, Barclay ML, et al. Consolidation of evidence for association of the KIAA1109-TENR-IL2-IL21 rs6822844 variant with Crohn's disease. Am J Gastroenterol 2010;105:1204-5.
33. Antony F, Deantonio C, Cotella D, et al. High-throughput assessment of the antibody profile in ovarian cancer ascitic fluids. Oncoimmunology 2019;8:e1614856.
34. van den Berg ME, Warren HR, Cabrera CP, et al. Discovery of novel heart rate-associated loci using the Exome Chip. Hum Mol Genet 2017;26:2346-63.
35. Yazdani A, Yazdani A, Elsea SH, et al. Genome analysis and pleiotropy assessment using causal networks with loss of function mutation and metabolomics. BMC Genomics 2019;20:395.

Cite this article as: Zhou J, Chen C, Liu S, Zhou W, Du J, Jiang Y, Dai J, Jin G, Ma H, Hu Z, Chen J, Shen H. Potential functional variants of KIAA genes are associated with breast cancer risk in a case control study. Ann Transl Med 2021;9(7):549. doi: 10.21037/atm-20-6108

Table S1 Selected the KIAA genes from the Gencode databases

Chr	Gene	Position (hg19)
1	*KIAA1324*	109656480-109745853
1	*KIAA2013*	11979648-11986485
1	*KIAA0907*	155882834-155904191
1	*KIAA0040*	175126123-175162135
1	*KIAA1614*	180882290-180920750
1	*KIAA1522*	33207486-33240571
1	*KIAA0319L*	35899091-36023014
1	*KIAA0754*	39876151-39882154
1	*KIAA1107*	92632542-92650280
2	*KIAA1715*	176788620-176867567
2	*KIAA1841*	61293006-61390108
2	*KIAA1841*	61293006-61390108
2	*KIAA1211L*	99410309-99552722
3	*KIAA1524*	108268716-108308491
3	*KIAA2018*	113367232-113415493
3	*KIAA1407*	113682984-113775460
3	*KIAA1257*	128628709-128721533
3	*KIAA0226*	197398264-197476598
3	*KIAA1143*	44779153-44803154
4	*KIAA1109*	123073488-123283913
4	*KIAA0922*	154387498-154557863
4	*KIAA1430*	186080819-186130658
4	*KIAA1239*	37245842-37451087
4	*KIAA1211*	57036361-57194791
4	*KIAA0232*	6783102-6885897
5	*KIAA1024L*	129083772-129102425
5	*KIAA0141*	141303373-141321612
5	*KIAA1191*	175773064-175788971
5	*KIAA0947*	5420777-5490347
5	*KIAA0825*	93488671-93918288
6	*KIAA1919*	111580551-111592370
6	*KIAA0408*	127772427-127780536
6	*KIAA1244*	138483058-138665800
6	*KIAA0319*	24544332-24646383
6	*KIAA1586*	56911347-56920023

Table S1 (*continued*)

http://dx.doi.org/10.21037/atm-20-6108

Table S1 (*continued*)

Chr	*Gene*	Position (hg19)
6	*KIAA1009*	84833960-84937353
6	*KIAA1009*	84833960-84937353
7	*KIAA1549*	138516126-138666064
7	*KIAA1147*	141356611-141401953
7	*KIAA0087*	26572740-26578407
7	*KIAA0895*	36363830-36429734
7	*KIAA1324L*	86506222-86689015
8	*KIAA0196*	126036502-126103920
8	*KIAA0196-AS1*	126052926-126057231
8	*KIAA1456*	12803151-12889012
8	*KIAA1875*	145162629-145173218
8	*KIAA1429*	95499921-95565757
9	*KIAA0368*	114122972-114247025
9	*KIAA1958*	115249127-115431677
9	*KIAA1984*	139690802-139702193
9	*KIAA1984-AS1*	139698379-139703300
9	*KIAA0020*	2720469-2844241
9	*KIAA1161*	34366668-34376851
9	*KIAA1045*	34957484-34982541
9	*KIAA1432*	5629025-5776557
9	*KIAA2026*	5881596-6007825
10	*KIAA1598*	118643742-118886097
10	*KIAA1217*	23983675-24836772
10	*KIAA1217*	23983675-24836772
10	*KIAA1462*	30301729-30404423
10	*KIAA1279*	70748487-70776738
11	*KIAA1377*	101785746-101871789
11	*KIAA1549L*	33563618-33695648
11	*KIAA1731*	93394805-93463113
12	*KIAA1033*	105501102-105562912
12	*KIAA1467*	13197218-13295455
12	*KIAA1551*	32112304-32146039
13	*KIAA0226L*	46916139-46964177
14	*KIAA0125*	106383838-106391825

Table S1 (*continued*)

http://dx.doi.org/10.21037/atm-20-6108

Table S1 (*continued*)

Chr	*Gene*	Position (hg19)
14	*KIAA0391*	35591052-35743271
14	*KIAA0586*	58894103-59015216
14	*KIAA0247*	70078313-70181859
14	*KIAA1737*	77564440-77583630
15	*KIAA0125P2*	21224717-21226180
15	*KIAA0101*	64657193-64679886
15	*KIAA1024*	79724858-79764632
15	*KIAA1199*	81071684-81244117
16	*KIAA0430*	15688243-15737023
16	*KIAA0556*	27561473-27791690
16	*KIAA0895L*	67209644-67217943
16	*KIAA0513*	85061375-85127836
17	*KIAA0100*	26941458-26972472
17	*KIAA0753*	6481468-6544077
17	*KIAA0753*	6481468-6544077
17	*KIAA0195*	73437240-73496163
18	*KIAA1328*	34409159-34812135
18	*KIAA1468*	59854491-59974355
19	*KIAA1683*	18367908-18385319
19	*KIAA0355*	34745442-34846491
20	*KIAA1755*	36838890-36889174
22	*KIAA1671*	25348697-25593415
22	*KIAA1658*	30814212-30814469
22	*KIAA1644*	44639547-44708731
22	*KIAA0930*	45586219-45636650
X	*KIAA1210*	118212598-118284542
X	*KIAA2022*	73952684-74145282

Chr: chromosome.

http://dx.doi.org/10.21037/atm-20-6108

Table S2 Expression of candidate KIAA genes obtained from breast cancer TCGA database

Gene	Tumor	Normal	Log_2 FC	P value	Call rate in tumor	Call rate in normal
KIAA0101	5.37	0.59	2.98	4.36E-44	1.00	1.00
KIAA1456	0.41	1.62	-1.76	3.05E-30	1.00	1.00
KIAA1614	0.34	0.75	-0.95	2.32E-29	1.00	1.00
KIAA1524	2.95	0.78	1.80	1.58E-25	1.00	1.00
KIAA1199	3.52	0.36	2.98	2.52E-23	1.00	1.00
KIAA1211	1.30	0.19	2.26	8.85E-23	1.00	1.00
KIAA1683	0.73	2.11	-1.42	2.88E-21	1.00	1.00
KIAA0355	5.72	8.44	-0.55	4.73E-21	1.00	1.00
KIAA1598	12.48	6.04	1.04	1.45E-18	1.00	1.00
KIAA0408	0.01	0.18	-1.29	3.09E-17	1.00	1.00
KIAA1841	2.05	1.27	0.65	7.55E-15	1.00	1.00
KIAA1107	1.46	2.23	-0.58	2.86E-13	1.00	1.00
KIAA1109	5.69	8.59	-0.58	3.97E-13	1.00	1.00
KIAA0141	9.36	12.23	-0.38	1.54E-12	1.00	1.00
KIAA0196	26.53	15.78	0.75	4.85E-12	1.00	1.00
KIAA1257	0.39	0.11	1.24	2.03E-11	1.00	1.00
KIAA1024	0.54	0.23	0.95	1.51E-10	1.00	1.00
KIAA1377	2.41	3.14	-0.37	2.42E-10	1.00	1.00
KIAA0226	4.93	3.81	0.36	8.04E-10	1.00	1.00
KIAA1671	7.17	11.14	-0.63	1.77E-09	1.00	1.00
KIAA0895	2.35	1.44	0.67	2.60E-09	1.00	1.00
KIAA0226L	0.91	1.43	-0.60	3.25E-09	1.00	1.00
KIAA1644	0.57	0.23	1.02	3.46E-09	1.00	1.00
KIAA1429	14.23	9.11	0.64	4.86E-09	1.00	1.00
KIAA0430	10.96	13.80	-0.33	6.84E-09	1.00	1.00
KIAA1244	11.88	6.31	0.90	8.36E-09	1.00	1.00
KIAA1239	0.06	0.37	-1.54	2.02E-08	1.00	1.00
KIAA1467	18.66	5.65	1.71	2.80E-08	1.00	1.00
KIAA1009	2.12	2.78	-0.38	3.44E-08	1.00	1.00
KIAA1279	15.71	12.31	0.35	5.40E-08	1.00	1.00
KIAA1522	40.19	28.43	0.50	6.70E-08	1.00	1.00
KIAA1045	0.21	0.31	-0.41	3.95E-07	1.00	1.00
KIAA0319	0.31	0.05	1.47	4.81E-07	1.00	1.00
KIAA0556	4.64	3.08	0.58	8.55E-07	1.00	1.00
KIAA1549L	1.57	0.85	0.81	2.84E-06	1.00	1.00

Table S2 (*continued*)

http://dx.doi.org/10.21037/atm-20-6108

Table S2 (*continued*)

Gene	Tumor	Normal	Log_2 FC	P value	Call rate in tumor	Call rate in normal
KIAA1328	0.88	1.10	-0.29	3.31E-06	1.00	1.00
KIAA0922	5.54	6.41	-0.21	4.77E-06	1.00	1.00
KIAA0513	2.76	1.94	0.49	7.40E-06	1.00	1.00
KIAA0907	9.26	6.81	0.44	7.69E-06	1.00	1.00
KIAA1143	7.65	8.95	-0.22	1.10E-05	1.00	1.00
KIAA1462	6.19	8.49	-0.45	2.90E-05	1.00	1.00
KIAA2013	19.72	16.36	0.27	5.48E-05	1.00	1.00
KIAA0319L	14.99	11.40	0.39	7.59E-05	1.00	1.00
KIAA1217	16.70	21.42	-0.36	9.23E-05	1.00	1.00
KIAA1161	7.94	5.09	0.63	1.15E-04	1.00	1.00
KIAA0040	32.41	19.93	0.70	2.18E-04	1.00	1.00
KIAA1211L	4.43	3.17	0.47	2.96E-04	1.00	1.00
KIAA0232	10.23	11.45	-0.16	4.56E-04	1.00	1.00
KIAA2022	0.45	0.28	0.54	5.83E-04	1.00	1.00
KIAA2018	3.44	3.90	-0.17	6.84E-04	1.00	1.00
KIAA1586	2.57	2.86	-0.15	8.22E-04	1.00	1.00
KIAA1024L	0.01	0.00	0.11	1.67E-03	1.00	1.00
KIAA1549	2.09	1.51	0.45	1.81E-03	1.00	1.00
KIAA1755	1.32	1.52	-0.19	3.16E-03	1.00	1.00
KIAA1958	1.40	1.11	0.31	4.61E-03	1.00	1.00
KIAA0368	21.25	22.68	-0.09	4.87E-03	1.00	1.00
KIAA1432	4.83	5.26	-0.12	7.61E-03	1.00	1.00
KIAA1324L	3.47	3.54	-0.03	8.55E-03	1.00	1.00
KIAA0930	10.33	9.37	0.14	1.65E-02	1.00	1.00
KIAA0247	21.92	19.20	0.19	2.01E-02	1.00	1.00
KIAA0391	0.93	0.82	0.16	2.83E-02	1.00	1.00
KIAA2026	5.78	6.37	-0.14	3.23E-02	1.00	1.00
KIAA1731	2.25	1.86	0.26	4.09E-02	1.00	1.00
KIAA0825	0.56	0.41	0.38	4.17E-02	1.00	1.00
KIAA1324	50.32	42.35	0.25	4.52E-02	1.00	1.00

FC, fold change.

http://dx.doi.org/10.21037/atm-20-6108

Table S3 Demographic and selected variables in breast cancer cases and cancer-free controls

Variables	Cases[a] (N=1,032)	Controls[b] (N=1,063)	P[c]
Age, years (mean ± SD)	50.87±11.43	51.74±11.23	0.078
Age at menarche, years (mean ± SD)	15.22±1.91	16.17±1.96	<0.0001
Age at first live birth, years (mean ± SD)	25.6±3.25	24.69±3.35	<0.0001
Age at natural menopause, years (mean ± SD)	49.72±3.5	49.59±3.91	0.617
Menopausal status			<0.0001
Premenopausal	499(49.45)	470(52.63)	
Natural menopausal	438(43.41)	411(46.02)	
Unnatural menopausal	72(7.14)	12(1.34)	
Estrogen receptor (ER)[d]			
Positive	460(55.22)		
Negative	373(44.78)		
Progesterone receptor (PR)[d]			
Positive	469(56.57)		
Negative	360(43.43)		

[a]Cases were consecutively recruited from the First Affiliated Hospital of Nanjing Medical University, the Cancer Hospital of Jiangsu Province and the Gulou Hospital, Nanjing, China, from Jan 2004 to April 2010; [b]Controls were randomly selected from a cohort of more than 30,000 participants in a community-based screening program for non-infectious diseases conducted in Jiangsu Province; [c]T-tests and $\chi 2$ tests were used for continuous or categorical variables, respectively; [d]ER and PR status information was available in 833/829 breast cancer cases.

Table S4 In silico analysis for SNPs function annotation

SNP	Chr: Position[a]	Gene	Allele[b]	MAF[c]	Regulome DB Score	I-Mutant/Mupro	HaploReg
rs2306369	4: 123268859	KIAA1109 (missense)	A/G	0.15	5	Decreased	DNAse, Motifs changed, Enhancer histone marks
rs1205434	20: 36869516	KIAA1755 (missense)	C/A	0.27	5	Decreased	Motifs changed, Enhancer histone marks, Selected eQTL hits

[a]Based on NCBI build 37 of the human genome. [b]Reference allele/effect allele. [c]minor allele frequency in 1000 Genomes Project East Asian data. SNP: single nucleotide polymorphism, Chr: chromosome, MAF: minor allele frequency, eQTL: expression Quantitative Trait Loci.

http://dx.doi.org/10.21037/atm-20-6108