International Journal of Clinical Oncology (2023) 28:664–679
https://doi.org/10.1007/s10147-023-02323-z

ORIGINAL ARTICLE

Impact of *MIR31HG* polymorphisms on risk of breast cancer in Chinese women

Ying Wei[1,2] · Xiaolin Wang[3] · Zhe Zhang[3] · Changtao Zhao[2] · Yuwei Chang[2] · Zhiqing Bian[2] · Xinhan Zhao[1]

Received: 16 June 2022 / Accepted: 1 March 2023 / Published online: 8 March 2023

Abstract

Background Breast cancer (BC) is one of the leading causes of death worldwide. This study explored the relationship between the *MIR31HG* gene polymorphisms and the risk of BC in Chinese women.

Methods Eight single nucleotide polymorphisms (SNPs) in *MIR31HG* were genotyped among 545 patients with BC and 530 healthy controls using Agena MassARRAY analysis. The PLINK software was used to calculate the odds ratio (OR) and 95% confidence intervals (CIs) via the logistic regression analysis. Multi-factor dimensionality reduction (MDR) analysis was performed to study the impact of SNP-SNP interaction on BC risk.

Results *MIR31HG* rs72703442-AA (OR 0.29, 95% CI 0.10–0.79, $p=0.026$), rs55683539-TT (OR 0.46, 95% CI 0.26–0.80, $p=0.012$) and rs2181559-AA (OR 0.59, 95% CI 0.40–0.89, $p=0.038$) were associated with a reduced risk of BC in Chinese women, as well as stratified results at age ≥ 52 years. Rs79988146 was correlated with estrogen receptor (ER) and progesterone receptor (PR)in Chinese female BC patients under various genetic models. Age at menarche stratification indicated that rs1332184 was associated with increased risk in BC patients, whereas stratification by number of births indicated that rs10965064 was associated with reduced risk in BC patients. MDR analysis showed that the best single-locus model for predicting of BC risk are rs55683539, which, rs55683539-CC group was a high risk group and rs55683539-TT group was a low risk group.

Conclusions The results indicated that the *MIR31HG* polymorphisms were associated with a reduced risk of BC in Chinese women.

Keywords *MIR31HG* · Polymorphism · Breast cancer · Risk · Case–control study

Abbreviations

BC	Breast cancer
SNPs	Single nucleotide polymorphisms
OR	Odds ratio
95% CI	95% Confidence interval
ER	Estrogen receptor
PR	Progesterone receptor
LN	Lymph node
lncRNA	Long non-coding RNA
HWE	Hardy–Weinberg equilibrium
FDR	False discovery rate

Introduction

Breast cancer (BC), which has to be highlighted in China as the same as many other countries, is becoming the heaviest disease burden and the main leading cause of cancer-related death in women [1–6]. In 2015, estimated 272,400 new cases of BC were diagnosed, and estimated 70,700 deaths were expected to occur with incidence increasing year by year in China [4]. The fact that BC, as a complex disease to tackle, is caused by a combination of various factors such as, age, lifestyle, family history, and serum hormone level [4, 5, 7, 8]. Among these factors

Joint first authors: Ying Wei and Xiaolin Wang.

✉ Xinhan Zhao
zhaoxinhan@mail.xjtu.edu.cn

[1] Department of Internal Medicine Oncology, The First Affiliated Hospital of Xi'an Jiaotong University School of Medicine, Xi'an, #227 West Yanta Road, 710061, Shaanxi, China

[2] Department of Internal Medicine Oncology, Yulin No.2 Hospital, Yulin 719000, Shaanxi, China

[3] Department of General Surgery, Yulin No.2 Hospital, Yulin 719000, Shaanxi, China

mentioned above, hereditary factor is considered to be the most important and crucial one, because 5–10% of cases are raised from genetic variation of susceptible genes [9]. Genome-wide association studies (GWAS) have reported significant effects of some gene polymorphisms on BC risk [10–14], but many new associations between single nucleotide polymorphisms (SNPs) and BC risk have not been explored.

Long non-coding RNAs (lncRNAs) are non-coding RNAs longer than 200 nucleotides. Recently, several studies have shown a link between genetic variants in lncRNA genes and breast cancer risk. Cui et al. found a SNP 2 kb upstream of H19 transcription start site that was associated with breast cancer risk in estrogen receptor (ER)-positive patients in the Chinese population [15]. Wu et al. studied risk associations among 22,977 cases and 105,974 controls of European ancestry and found several novel risk-loci that harbored lncRNA genes [16]. *MIR31HG*, identified as LNCHIFCAR/LOC554202/hsa-lnc-31, is located on chromosome 9 and produces a long non-coding RNA (lncRNA) which acts as a host gene for *MIR-31*. lncRNA with a lack of protein-coding function defined as RNA longer than 200 nucleotides, plays considerable and remarkable role in complex biological activities, including regulating gene expression through chromatin remodeling, controlling gene transcription, participating in post-transcriptional mRNA process, and mediating protein function or localization [17]. Their dysregulation seems to be contributed to the growth and progression of human tumors [18], and therefore *MIR31HG* is widely reported to be involved in the development of various cancers, such as colorectal cancer [19], bladder cancer [20], oral cancer [21], lung adenocarcinoma [17], pancreatic carcinoma [22], esophageal squamous cell carcinoma [23] and BC [6, 24, 25]. Studies have shown that as a non-coding oncogene, the down-regulated expression of *MIR31HG* can lead to diminished cell proliferation, migration, invasion and increased apoptosis in BC [25]. Through gene evaluated, a related study also explained that *MIR31HG* was regulated by promoter hypermethylation in triple-negative BC and participated in the regulatory mechanism of BC as an important determination of the invasion metastasis cascade [24]. These studies have shown that *MIR31HG* has an important role in BC, however, there is no report on the relationship between the polymorphisms of *MIR31HG* and the risk of BC.

This study aimed to reveal the impact of *MIR31HG* gene polymorphisms on the risk of BC in Chinese women through a case–control study, and to explore the association of *MIR31HG* polymorphisms with clinical characteristics of BC patients, which may provide a theoretical and experimental basis for further investigating the role of *MIR31HG* on BC carcinogenesis.

Materials and methods

Study population

The study was approved by the ethics committee of Shaanxi Provincial Cancer Hospital, and informed consents were delivered and signed by all participants. A total of 545 patients with BC (mean age 52.00 ± 9.89 years) were recruited and analyzed for the study in 2017 and 2018. The histopathological diagnosis was followed the classification of breast tumors by the World Health Organization (WHO5th), and clinical staging was based on American Joint Committeeon Cancer (AJCC) on breast cancer TNM staging system. It is necessary to exclude some patients who had family history of cancer, received radiotherapy, chemotherapy, or other treatments before the period of investigation. BC cases were categorized by estrogen receptor (ER), progesterone receptor (PR), lymph node (LN) metastasis, clinical stage, human epidermal growth factor receptor 2 (HER2), Ki67, tumor size, tumor location and distant metastasis, among which ER and PR test results > 1% were defined as ER-positive and PR-positive, and HER2 3+ and 2+ with fluorescence in situ hybridization (FISH) positive indicated HER2-positive. The Ki67 positive rate 20% was used as the cutoff point to divide patients with BC into low (< 20%) and high (> 20%) groups. During this period, 530 cancer-free controls (mean age: 51.66 ± 9.67 years) were enrolled from the healthcare of the hospital at the same time. The excluded criteria of controls were as follows: (1) no gynecological neoplasm, (2) no other history of solid cancers, and (3) no immune disorders (Table 1). Approximately 3–5 mL of venous blood sample was collected from each participant and then was placed into anti-coagulative tubes stored at − 80 °C until use. Demographic and clinic indicators were recorded by self-administered standardized questionnaires and medical records, respectively.

Extraction of genomic DNA and genotyping

Studies have shown that *MIR31HG* plays an important role in BC, but the correlation between *MIR31HG* polymorphisms and BC risk has not been reported. The selection of the eight candidate SNPs on the *MIR31HG* gene in this study is based on haplotype data or genotype data [26] and from the 1,000 Genome Projects (http://www.internationalgenome.org/) to select SNPs with a minor alleles frequency (MAF) greater than 0.05 in the global population.

Following the GoldMag-Mini extraction method (GoldMag Co, Ltd, Xi'an, China) strictly, genomic DNA was extracted from the venous blood. DNA concentration

Table 1 Clinical characteristics in cases and controls

Characteristic	Case (n = 545)	Control (n = 530)	*p*
Age mean ± SD	52.00 ± 9.89	51.66 ± 9.67	0.453
ER (%)			
Positive	371 (68.0%)	–	
Negative	166 (30.5%)	–	
Missing	8 (1.5%)	–	
PR (%)			
Positive	320 (58.7%)	–	
Negative	217 (39.8%)	–	
Missing	8 (1.5%)	–	
LN metastasis			
Yes	267 (49.0%)	–	
No	278 (51.0%)	–	
Clinical stage			
I/II	355 (65.1%)	–	
III/IV	156 (28.6%)	–	
HER2			
Positive	78 (14.3%)	–	
Negative	279 (51.2%)	–	
Missing	188 (34.5%)	–	
Ki67			
Low (≤ 20%)	147 (27.0%)	–	
High (> 20%)	363 (66.6%)	–	
Missing	35 (6.4%)	–	
Tumor location			
Right	258 (47.3%)	–	
Left	279 (51.2%)	–	
between	8 (1.5%)	–	
Tumor size			
≥ 2 cm	305 (56.0%)	–	
< 2 cm	135 (24.8%)	–	
Missing	105 (19.3%)	–	
Number of births			
> 1 time	246 (45.1%)	–	
≤ 1 time	231 (42.4%)	–	
Missing	68(12.5%)	–	
Age of menarche			
> 13 years old	332 (60.9%)	–	
≤ 13 years old	117 (21.5%)	–	
Missing	96 (17.6%)	–	
Menopausal status			
Post-	280 (51.4%)	–	
Pre-	164 (30.1%)	–	
Missing	101 (18.5%)	–	

P value was calculated by test. $P < 0.05$ indicated a significant difference

SD standard deviation, *ER* estrogen receptor, *PR* progesterone receptor, *LN* lymph node, *HER2*, human epidermal growth factor receptor 2

was measured by spectrometry (DU530 UV/VIS spectrophotometer, Beckman Instruments, Fullerton, CA, USA). The Agena Biosicence Assay Design Suite V2.0 software (http://agenacx.com/online-tools) was used to design the extended primer. The MassARRAY Nanodispenser (Agena Bioscience, San Diego, CA, USA) and the MassARRAY iPLEX platform (Agena Bioscience, San Diego, CA, USA) were used to genotype, and then Agena Bioscience TYPER software (version 4.0) was used to analyze the data [27].

Statistical analyses

Statistical analysis was set up using Microsoft Excel, SPSS 18.0 statistical package (SPSS, Chicago, IL, USA) and the PLINK 1.07 software. Hardy–Weinberg equilibrium (HWE) for *MIR31HG* genotype distributions of controls were accessed using Fisher's exact test. The demographic and clinical characteristics of study participants were evaluated by chi-squared test to compare the differences in genotypes and allele frequency distribution between the groups. Welch's T-test was used to compare ages between cases and controls. Logistic regression analysis was used to evaluate the genetic susceptibility of BC under five genetic models (allele, codominant, recessive, dominant, and additive models). Odds ratios (ORs) and 95% confidence intervals (CIs) from a logistic regression model were performed to analyze the relative risk. All *p* values were two sided and $p < 0.05$ was considered to be statistically significant. Multi-factor dimensionality reduction (MDR) analysis was performed to assess the impact of SNPs interactions on BC risk [28]. We used G*power 3.1.9.2 software to calculate the minimum sample size and actual power values required for this study [29].

Results

Demographics of study subjects and SNPs information

The minimum sample size of the case population and the control population calculated by G-power software is 107 and 103, respectively, and the actual power value is 0.95. Eight SNPs were genotyped in 545 patients with BC and 530 cancer-free controls. Demographics and clinical characteristics of all study subjects are displayed in Table 1, which showed that the sample size of the subject population recruited in this study is completely in line with statistical significance. The cases consisted of 371 (68.0%) ER positive tumors, 320 (58.7%) PR positive tumors, 267 (49.0%) LN metastasis positive tumors, 78 (14.3%) HER2 positive tumors, 355 (65.1%) I/II and 156 (28.6%) III/IV clinical stage, 147 (27.0%) low Ki67 status, 258 (47.3%)

right and 279 (51.2%) left in tumor locations. There is no significant difference between cases and controls in the distribution of ages ($p = 0.453$). Basic information including position, allele, role, and MAF of all eight *MIR31HG* SNPs between cases and controls is shown in Table 2. The genotype distributions of all SNPs were in accordance with HWE ($p > 0.05$).

Association between MIR131HG polymorphisms and BC risk

In addition, five multiple genetic models (allele, codominant, dominant, recessive, and additive models) were used to analyze the relationship between the candidate SNPs and the risk of BC in Chinese women. The results showed that under the polygenic model, the three candidate SNPs on the MIR131HG gene were significantly associated with BC risk ($p < 0.05$) (see Table 3). Specifically, rs72703442 was significantly associated with a lower risk of BC under both co-dominant (OR 0.29, 95% CI 0.10–0.79, $p = 0.026$) and recessive (OR 0.30, 95% CI 0.11–0.82, $p = 0.011$) models. Rs55683539 was significantly associated with a lower risk of BC in allelic (OR 0.76, 95% CI 0.62–0.93, $p = 0.007$), codominant (OR 0.46, 95% CI 0.26–0.80, $p = 0.012$), dominant (OR 0.78, 95% CI 0.61–0.99, $p = 0.040$), recessive (OR 0.49, 95% CI 0.29–0.84, $p = 0.008$), and additive (OR 0.76, 95% CI 0.62–0.93, $p = 0.007$) models. Rs2181559 was significantly associated with a lower risk of BC in allelic (OR 0.82, 95% CI 0.69–0.98, $p = 0.028$), codominant (OR 0.59, 95% CI 0.40–0.89, $p = 0.038$), recessive (OR 0.62, 95% CI 0.43–0.91, $p = 0.014$), and additive (OR 0.82, 95% CI 0.68–0.98, $p = 0.026$) models.

Age-stratified analysis

Then, according to the average age of the recruited subjects, we conducted a stratified analysis with 52 years as the age node to further explore the effect of the *MIR131HG* polymorphisms on the risk of BC in Chinese women (Table 4). In women aged ≥ 52 years, *MIR31HG* rs72703442 was significantly associated with reduced risk of BC under both codominant (OR 0.18, 95% CI 0.04–0.81, $p = 0.034$) and recessive (OR 0.18, 95% CI 0.04–0.83, $p = 0.010$) recessive models. Moreover, we detected rs55683539 in the allele (OR 0.68, 95% CI 0.51–0.90, $p = 0.007$) codominant (OR 0.68, 95% CI 0.48–0.98, $p = 0.015$), dominant (OR 0.64, 95% CI 0.45–0.91, $p = 0.012$), and additive (OR 0.65, 95% CI 0.49–0.88, $p = 0.004$) model and found a significant lower risk result. Meanwhile, women with A allele (compared with those carrying the T allele) and AA genotype (compared with those carrying the TT and TA genotype) for rs2181559 had a reduced risk BC in the allele (OR 0.76, 95% CI 0.56–0.94, $p = 0.014$) and co-dominant (OR 0.48, 95% CI 0.27–0.84, $p = 0.032$), recessive (OR 0.54, 95% CI 0.32–0.92, $p = 0.021$), and additive (OR 0.72, 95% CI 0.55–0.93, $p = 0.012$) models. However, in women < 52 years old, we did not find that candidate SNPs have an impact on the risk of BC in Chinese women ($p > 0.05$).

Stratified analysis of demographic and clinic indicators in case group

A stratified analysis of ER, PR, HER2, age at menarche, number of births and menopausal status in the case group was further analyzed. Both ER and PR stratified analyses indicated that rs79988146 was associated with ER positive and PR positive in BC patients under dominant (ER: OR 2.13, 95% CI 1.04–4.35, $p = 0.028$; PR: OR 1.94, 95% CI 1.03–3.62, $p = 0.033$) models (Table 5). Moreover, rs1332184 (allele: OR 0.56, 95% CI 0.35–0.89, $p = 0.014$; and additive: OR 0.54, 95% CI 0.33–0.88, $p = 0.009$), rs72703442 (allele: OR 0.39, 95% CI 0.20–0.78, $p = 0.005$; codominant: OR 0.29, 95% CI 0.13–0.64, $p = 0.003$; dominant OR 0.32, 95% CI 0.15–0.67, $p = 0.001$; and additive: OR 0.37, 95% CI 0.18–0.74, $p = 0.002$), rs55683539 (allele: OR 0.59, 95% CI 0.36–0.96, $p = 0.033$; dominant OR 0.51, 95% CI 0.29–0.91, $p = 0.019$; and additive: OR 0.60, 95% CI 0.37–0.98, $p = 0.032$) and rs2181559

Table 2 Allele frequencies in cases and controls among *MIR31HG* SNPs

SNP	Position	Allele A/B	Role	Minor allele Frequency		*p*-HWE
				Case	Control	
rs79988146	21461747	C/T	Intron	0.051	0.042	0.239
rs1332184	21504204	A/G	Intron	0.246	0.270	0.999
rs72703442	21515796	A/C	Intron	0.139	0.169	0.758
rs2025327	21531630	C/T	Intron	0.106	0.113	0.999
rs55683539	21542135	T/C	Intron	0.215	0.264	0.503
rs2181559	21543939	A/T	Intron	0.325	0.370	0.780
rs10965059	21544063	T/C	Intron	0.091	0.090	0.785
rs10965064	21553539	G/C	Intron	0.372	0.377	0.927

SNP single nucleotide polymorphism, *HWE* Hardy–Weinberg equilibrium, *A* minor allele, *B* major allele

Table 3 Genotypic model analysis of the relationship between *MIR31HG* SNPs and BC risk

SNP	Model	Genotype	Case, n (%)	Control, n (%)	OR (95% CI)	*p*
rs79988146	Allele	T	1034 (94.9%)	1015 (95.7%)	1	
		C	56 (5.1%)	45 (4.3%)	1.22 (0.82–1.83)	0.328
	Codominant	TT	487 (91.9%)	491 (90.1%)	1	0.570
		CT	41 (7.7%)	52 (9.5%)	1.26 (0.82–1.93)	
		CC	2 (0.4%)	2 (0.4%)	0.99 (0.14–7.07)	
	Dominant	TT	487 (91.9%)	491 (90.1%)	1	0.300
		CC/TC	43 (8.1%)	54 (9.9%)	1.25 (0.82–1.90)	
	Recessive	TT/TC	528 (99.6%)	543 (99.6%)	1	0.980
		CC	2 (0.4%)	2 (0.4%)	0.97 (0.14–6.93)	
	Additive	–	–	–	1.21 (0.82–1.80)	0.340
rs1332184	Allele	G	822 (75.4%)	771 (73.0%)	1	
		A	268 (24.6%)	285 (27.0%)	0.88 (0.73–1.07)	0.204
	Codominant	GG	281 (53.2%)	308 (56.5%)	1	0.430
		GA	209 (39.6%)	206 (37.8%)	0.90 (0.70–1.16)	
		AA	38 (7.2%)	31 (5.7%)	0.74 (0.45–1.23)	
	Dominant	GG	281 (53.2%)	308 (56.5%)	1	0.280
		GA/AA	247 (46.8%)	237 (43.5%)	0.88 (0.69–1.11)	
	Recessive	GG/GA	490 (92.8%)	514 (94.3%)	1	0.310
		AA	38 (7.2%)	31 (5.7%)	0.78 (0.48–1.27)	
	Additive	–	–	–	0.88 (0.72–1.07)	0.200
rs72703442	Allele	C	939 (86.2%)	881 (83.1%)	1	
		A	151 (13.8%)	179 (16.9%)	0.79 (0.63–1.00)	0.051
	Codominant	CC	367 (69.2%)	399 (73.2%)	1	**0.026**
		CA	147 (27.8%)	141 (25.9%)	0.88 (0.67–1.16)	
		AA	16 (3.0%)	5 (0.9%)	**0.29 (0.10–0.79)**	
	Dominant	CC	367 (69.2%)	399 (73.2%)	1	0.150
		CA/AA	163 (30.8%)	146 (26.8%)	0.82 (0.63–1.07)	
	Recessive	CC/CA	514 (97.0%)	540 (99.1%)	1	**0.011**
		AA	16 (3.0%)	5 (0.9%)	**0.30 (0.11–0.82)**	
	Additive	–	–	–	0.79 (0.62–1.00)	0.047
rs2025327	Allele	T	974 (89.4%)	940 (88.7%)	1	
		C	116 (10.6%)	120 (11.3%)	0.93 (0.71–1.22)	0.615
	Codominant	TT	416 (78.5%)	435 (79.8%)	1	0.860
		CT	108 (20.4%)	104 (19.1%)	0.92 (0.68–1.24)	
		CC	6 (1.1%)	6 (1.1%)	0.96 (0.31–2.99)	
	Dominant	TT	416 (78.5%)	435 (79.8%)	1	0.590
		CC/TC	114 (21.5%)	110 (20.2%)	0.92 (0.69–1.24)	
	Recessive	TT/TC	524 (98.9%)	539 (98.9%)	1	0.960
		CC	6 (1.1%)	6 (1.1%)	0.97 (0.31–3.03)	
	Additive	–	–	–	0.93 (0.71–1.22)	0.610
rs55683539	Allele	C	856 (78.5%)	780 (73.6%)	1	
		T	234 (21.5%)	280 (26.4%)	**0.76 (0.62–0.93)**	**0.007**
	Codominant	CC	290 (54.7%)	332 (60.9%)	1	**0.012**
		CT	200 (37.7%)	192 (35.2%)	0.84 (0.65–1.08)	
		TT	40 (7.5%)	21 (3.8%)	**0.46 (0.26–0.80)**	
	Dominant	CC	290 (54.7%)	332 (60.9%)	1	**0.040**
		TT/TC	240 (45.3%)	213 (39.1%)	**0.78 (0.61–0.99)**	
	Recessive	CC/TC	490 (92.5%)	524 (96.2%)	1	**0.008**
		TT	40 (7.5%)	21 (3.8%)	**0.49 (0.29–0.84)**	
	Additive	–	–	–	**0.76 (0.62–0.93)**	**0.007**
rs2181559	Allele	T	736 (67.5%)	668 (63.0%)	1	

Table 3 (continued)

SNP	Model	Genotype	Case, n (%)	Control, n (%)	OR (95% CI)	*p*
		A	354 (32.5%)	392 (37.0%)	**0.82 (0.69–0.98)**	**0.028**
	Codominant	TT	212 (40.0%)	241 (44.2%)	1	**0.038**
		TA	244 (46.0%)	254 (46.6%)	0.92 (0.71–1.18)	
		AA	74 (14.0%)	50 (9.2%)	0.59 (0.40–0.89)	
	Dominant	TT	212 (40.0%)	241 (44.2%)	1	0.160
		TA/AA	318 (60.0%)	304 (55.8%)	0.84 (0.66–1.07)	
	Recessive	TT/TA	456 (86.0%)	495 (90.8%)	1	**0.014**
		AA	74 (14.0%)	50 (9.2%)	**0.62 (0.43–0.91)**	
	Additive	–	–	–	**0.82 (0.68–0.98)**	**0.026**
rs10965059	Allele	C	983 (90.8%)	903 (91.0%)	1	
		T	99 (9.2%)	89 (9.0%)	1.02 (0.76–1.38)	0.888
	Codominant	CC	410 (82.7%)	450 (83.2%)	1	0.320
		CT	83 (16.7%)	83 (15.3%)	0.91 (0.65–1.27)	
		TT	3 (0.6%)	8 (1.5%)	2.43 (0.64–9.22)	
	Dominant	CC	410 (82.7%)	450 (83.2%)	1	0.820
		TT/TC	86 (17.3%)	91 (16.8%)	0.96 (0.70–1.33)	
	Recessive	CC/TC	493 (99.4%)	533 (98.5%)	1	0.160
		TT	3 (0.6%)	8 (1.5%)	2.47 (0.65–9.35)	
	Additive	–	–	–	1.02 (0.76–1.37)	0.890
rs10965064	Allele	C	684 (62.7%)	660 (62.3%)	1	
		G	406 (37.3%)	400 (37.7%)	0.98 (0.82–1.17)	0.815
	Codominant	CC	206 (38.9%)	218 (40.0%)	1	0.910
		CG	248 (46.8%)	248 (45.5%)	0.94 (0.73–1.22)	
		GG	76 (14.3%)	79 (14.5%)	0.98 (0.68–1.42)	
	Dominant	CC	206 (38.9%)	218 (40.0%)	1	0.700
		CG/GG	324 (61.1%)	327 (60.0%)	0.95 (0.75–1.22)	
	Recessive	CC/CG	454 (85.7%)	466 (85.5%)	1	0.940
		GG	76 (14.3%)	79 (14.5%)	1.01 (0.72–1.42)	
	Additive	–	–	–	0.98 (0.82–1.17)	0.820

P value was calculated by Wald Test adjusted by age

Bold indicated that $p < 0.05$ meant the data was statistically significant

SNP single nucleotide polymorphism, *OR* odds ratio, *95% CI* 95% confidence interval

(allele: OR 0.62, 95% CI 0.41–0.94, $p = 0.025$; dominant: OR 0.58, 95% CI 0.35–0.97, $p = 0.037$; and additive: OR 0.64, 95% CI 0.43–0.97, $p = 0.029$) were associated with negative HER2 status. The stratification of patients' menarche age showed that rs1332184 had a significant positive correlation with menarche age under allelic (OR 1.50, 95% CI 1.08–2.09, $p = 0.017$), codominant (OR 2.58 95% CI 1.15–5.79, $p = 0.048$), dominant (OR 1.55, 95% CI 1.01–2.37, $p = 0.044$), recessive (OR 2.24, 95% CI 1.03–4.91, $p = 0.049$), and additive (OR 1.52, 95% CI 1.08–2.12, $p = 0.016$) models. In addition, stratification of patients' reproductive times showed that rs55683539 and rs10965064 were significantly negatively correlated with patients' age at menarche under both codominant (rs55683539: OR 0.32, 95% CI 0.11–0.93, $p = 0.031$) and recessive (rs55683539: OR 0.29, 95% CI 0.10–0.86, $p = 0.017$; and rs10965064: OR 0.52, 95% CI 0.31–0.89, $p = 0.015$) models (Table 6). We also explore the association of *MIR31HG* SNPs with menopausal status of BC patients. However, no significant association was found.

Table 7 displayed the results of stratified analysis of LN metastasis, clinical stage, and tumor size in case group. Stratified analysis of LN metastasis presented a positive relationship between rs79988146 and LN metastasis under allelic (OR 1.79, 95% CI 1.03–3.11, $p = 0.038$), dominant (OR 1.88, 95% CI 1.05–3.36, $p = 0.032$), and additive (OR 1.76, 95% CI 1.01–3.08, $p = 0.042$) models. Stratified analysis of clinical stage displayed that rs1332184 was associated with the higher stage under allelic (OR 1.43, 95% CI 1.06–1.92, $p = 0.020$), codominant (OR 2.69, 95% CI 1.27–5.72, $p = 0.036$), recessive (OR 2.50, 95% CI 1.20–5.22, $p = 0.015$), and additive (OR 1.42, 95% CI 1.04–1.92, $p = 0.026$) models. Moreover, rs2181559 might

Table 4 Stratified analysis of the age on association between selected SNPs and BC risk

SNP	Model	Genotype	Age ≥ 52 years old		Age < 52 years old	
			OR (95% CI)	*p*	OR (95% CI)	*p*
rs79988146	Allele	T	1		1	
		C	1.41 (0.80–2.48)	0.227	1.05 (0.59–1.87)	0.858
	Codominant	TT	1	0.590	1	0.400
		CT	1.34 (0.72–2.47)		1.18 (0.64–2.15)	
		CC	1.72 (0.15–19.26)		/	
	Dominant	TT	1	0.310	1	0.710
		CC/TC	1.36 (0.75–2.47)		1.12 (0.62–2.03)	
	Recessive	TT/TC	1	0.670	1	0.210
		CC	1.67 (0.15–18.71)		0.00 (0.00-NA)	
	Additive	-	1.33 (0.77–2.32)	0.310	1.06 (0.60–1.89)	0.840
rs1332184	Allele	G	1		1	
		A	0.89 (0.68–1.17)	0.390	0.88 (0.67–1.16)	0.371
	Codominant	GG	1	0.460	1	0.510
		GA	1.04 (0.72–1.50)		0.81 (0.57–1.16)	
		AA	0.67 (0.34–1.33)		0.94 (0.44–2.01)	
	Dominant	GG	1	0.860	1	0.270
		GA/AA	0.97 (0.68–1.37)		0.83 (0.59–1.16)	
	Recessive	GG/GA	1	0.220	1	0.960
		AA	0.66 (0.34–1.29)		1.02 (0.48–2.16)	
	Additive	–	0.91 (0.69–1.20)	0.510	0.88 (0.66–1.17)	0.370
rs72703442	Allele	C	1		1	
		A	0.76 (0.55–1.06)	0.105	0.83 (0.59–1.16)	0.274
	Codominant	CC	1	**0.034**	1	0.510
		CA	0.93 (0.63–1.38)		0.84 (0.57–1.24)	
		AA	**0.18 (0.04–0.81)**		0.55 (0.13–2.34)	
	Dominant	CC	1	0.350	1	0.310
		CA/AA	0.84 (0.57–1.22)		0.82 (0.57–1.20)	
	Recessive	CC/CA	1	**0.010**	1	0.450
		AA	**0.18 (0.04–0.83)**		0.57 (0.14–2.44)	
	Additive	–	0.77 (0.55–1.07)	0.120	0.82 (0.58–1.16)	0.260
rs2025327	Allele	T	1		1	
		C	0.85 (0.59–1.24)	0.403	1.04 (0.7–1.55)	0.840
	Codominant	TT	1	0.770	1	0.870
		CT	0.86 (0.56–1.33)		1.02 (0.66–1.56)	
		CC	0.81 (0.21–3.07)		1.90 (0.17–21.32)	
	Dominant	TT	1	0.480	1	0.880
		CC/TC	0.86 (0.57–1.30)		1.03 (0.68–1.58)	
	Recessive	TT/TC	1	0.790	1	0.590
		CC	0.84 (0.22–3.16)		1.89 (0.17–21.23)	
	Additive	–	0.87 (0.60–1.27)	0.480	1.05 (0.70–1.58)	0.810
rs55683539	Allele	C	1		1	
		T	**0.68 (0.51–0.90)**	**0.007**	0.86 (0.65–1.13)	0.275
	Codominant	CC	1	**0.015**	1	0.190
		CT	**0.68 (0.48–0.98)**		1.02 (0.71–1.46)	
		TT	**0.38 (0.16–0.87)**		0.52 (0.25–1.09)	
	Dominant	CC	1	**0.012**	1	0.640
		TT/TC	**0.64 (0.45–0.91)**		0.92 (0.66–1.30)	
	Recessive	CC/TC	1	0.043	1	0.071
		TT	0.44 (0.19–1.00)		0.52 (0.25–1.07)	

Table 4 (continued)

SNP	Model	Genotype	Age ≥ 52 years old		Age < 52 years old	
			OR (95% CI)	*p*	OR (95% CI)	*p*
	Additive	–	**0.65 (0.49–0.88)**	**0.004**	0.86 (0.65–1.13)	0.280
rs2181559	Allele	T	1		1	
		A	**0.76 (0.56–0.94)**	**0.014**	0.93 (0.72–1.19)	0.554
	Codominant	TT	1	**0.032**	1	0.430
		TA	0.79 (0.54–1.15)		1.07 (0.75–1.53)	
		AA	**0.48 (0.27–0.84)**		0.73 (0.41–1.31)	
	Dominant	TT	1	0.056	1	1.000
		TA/AA	0.71 (0.50–1.01)		1.00 (0.71–1.40)	
	Recessive	TT/TA	1	**0.021**	1	0.220
		AA	**0.54 (0.32–0.92)**		0.71 (0.41–1.23)	
	Additive	–	**0.72 (0.55–0.93)**	**0.012**	0.93 (0.72–1.20)	0.570
rs10965059	Allele	C	1		1	
		T	0.87 (0.57–1.33)	0.526	1.21 (0.79–1.87)	0.383
	Codominant	CC	1	0.330	1	0.410
		CT	0.72 (0.45–1.16)		1.09 (0.68–1.75)	
		TT	1.54 (0.27–8.61)		3.74 (0.41–33.83)	
	Dominant	CC	1	0.230	1	0.540
		TT/TC	0.75 (0.47–1.20)		1.16 (0.73–1.84)	
	Recessive	CC/TC	1	0.560	1	0.200
		TT	1.65 (0.30–9.21)		3.70 (0.41–33.36)	
	Additive	–	0.82 (0.54–1.25)	0.350	1.21 (0.79–1.85)	0.380
rs10965064	Allele	C	1		1	
		G	0.97 (0.75–1.24)	0.784	0.99 (0.78–1.26)	0.943
	Codominant	CC	1	0.890	1	0.720
		CG	1.04 (0.72–1.52)		0.89 (0.62–1.29)	
		CC	0.92 (0.54–1.58)		1.07 (0.64–1.78)	
	Dominant	CC	1	0.940	1	0.700
		CG/GG	1.01 (0.71–1.45)		0.93 (0.66–1.32)	
	Recessive	CC/CG	1	0.680	1	0.590
		GG	0.90 (0.54–1.48)		1.14 (0.71–1.82)	
	Additive	–	0.98 (0.76–1.26)	0.880	1.00 (0.79–1.27)	0.990

P value was calculated by Wald Test adjusted by age

SNP single nucleotide polymorphism, *OR* odds ratio, *95% CI* 95% confidence interval

*$P < 0.05$ indicates statistical significance

be associated with larger tumor size of BC patients (additive: OR 1.39, 95% CI 1.00–1.94, $p = 0.046$).

MDR analysis for the effect of MIR31HG SNP-SNP interaction on BC risk

The Dendrogram and the Fruchterman-Reingold describe the interactions between these SNPs (Fig. 1A, B). Short connections among nodes represent stronger redundant interactions (Fig. 1A). A negative value for the two-locus entropy indicates an antagonistic effect, and a positive value indicates a synergistic effect (Fig. 1B). MDR analysis showed that candidate SNPs interaction is associated with BC risk (Table 8). The optimal single-locus model for predicting BC risk is rs55683539 [testing accuracy (TA): 0.5038, cross-validation consistency (CVC): 5/10], which, rs55683539-CC group was a high risk group and rs55683539-TT group was a low risk group increase the BC risk. Among the multi-locus models, predicting the best combination of BC risk is through rs79988146, rs1332184, rs72703442, rs2025327, rs55683539, rs2181559, rs10965059, and rs10965064 combination of eight-locus model [TA: 0.5179, CVC: 10/10]. The combination of all high-risk genotypes was associated with an increased risk of BC compared with that of low-risk genotypes.

Table 5 Stratified analysis of ER, PR and HER2 in case group

SNP	Model	Genotype	ER		PR		HER2	
			OR (95% CI)	*p*	OR (95% CI)	*p*	OR (95% CI)	*p*
rs79988146	Allele	T	1		1		1	
		C	1.88 (0.96–3.69)	0.061	1.74 (0.96–3.15)	0.064	1.15 (0.51–2.61)	0.733
	Codominant	TT	1	0.050	1	0.090	1	0.860
		CT	2.29 (1.09–4.84)		2.00 (1.05–3.80)		1.08 (0.46–2.52)	
		CC	0.55 (0.03–8.97)		0.95 (0.06–15.51)		/	
	Dominant	TT	1	**0.028**	1	**0.033**	1	
		CC/TC	**2.13 (1.04–4.35)**		**1.94 (1.03–3.62)**		1.08 (0.46–2.52)	0.860
	Recessive	TT/TC	1	0.630	1	0.930	1	
		CC	0.50 (0.03–8.22)		0.88 (0.05–14.30)		/	/
	Additive	–	1.89 (0.97–3.71)	0.048	1.79 (0.99–3.25)	0.046	1.08 (0.46–2.52)	0.860
rs1332184	Allele	G	1		1		1	
		A	1.12 (0.83–1.52)	0.451	1.05 (0.79–1.40)	0.720	**0.56 (0.35–0.89)**	**0.014**
	Codominant	GG	1	0.510	1	0.930	1	**/**
		GA	1.26 (0.85–1.86)		1.07 (0.74–1.54)		0.69 (0.40–1.18)	
		AA	1.01 (0.46–2.22)		1.09 (0.51–2.33)		/	
	Dominant	GG	1	0.300	1	0.690	1	0.045
		GA/AA	1.22 (0.84–1.77)		1.07 (0.76–1.52)		0.58 (0.34–1.00)	
	Recessive	GG/GA	1	0.840	1	0.890	1	/
		AA	0.92 (0.42–2.00)		1.06 (0.50–2.23)		/	
	Additive	–	1.13 (0.83–1.54)	0.440	1.06 (0.79–1.41)	0.710	**0.54 (0.33–0.88)**	**0.009**
rs72703442	Allele	C	1		1		1	
		A	0.87 (0.60–1.25)	0.452	0.86 (0.60–1.22)	0.389	**0.39 (0.20–0.78)**	**0.005**
	Codominant	CC	1	0.530	1	0.290	1	**0.003**
		CA	0.81 (0.53–1.22)		0.77 (0.52–1.14)		**0.29 (0.13–0.64)**	
		AA	1.62 (0.18–14.68)		2.36 (0.26–21.50)		0.91 (0.09–9.16)	
	Dominant	CC	1	0.350	1	0.260	1	**0.001**
		CA/AA	0.82 (0.55–1.24)		0.80 (0.54–1.18)		0.32 (0.15–0.67)	
	Recessive	CC/CA	1	0.620	1	0.370	1	0.910
		AA	1.71 (0.19–15.47)		2.52 (0.28–22.91)		1.15 (0.11–11.52)	
	Additive	–	0.86 (0.58–1.26)	0.440	0.84 (0.58–1.22)	0.370	**0.37 (0.18–0.74)**	**0.002**
rs2025327	Allele	T	1		1		1	
		C	0.95 (0.63–1.44)	0.808	1.00 (0.67–1.47)	0.980	0.80 (0.43–1.50)	0.485
	Codominant	TT	1	0.630	1	0.890	1	0.540
		CT	1.05 (0.66–1.68)		1.05 (0.67–1.62)		0.86 (0.44–1.68)	
		CC	0.46 (0.09–2.31)		0.71 (0.14–3.56)		/	
	Dominant	TT	1	1.000	1	0.920	1	0.570
		CC/TC	1.00 (0.63–1.57)		1.02 (0.67–1.57)		0.83 (0.42–1.61)	
	Recessive	TT/TC	1	0.340	1	0.670	1	/
		CC	0.45 (0.09–2.28)		0.70 (0.14–3.52)		/	
	Additive	–	0.95 (0.63–1.44)	0.810	1.00 (0.67–1.48)	0.990	0.80 (0.42–1.53)	0.490
rs55683539	Allele	C	1		1		1	
		T	0.97 (0.71–1.33)	0.845	0.98 (0.73–1.31)	0.873	**0.59 (0.36–0.96)**	**0.033**
	Codominant	CC	1	0.900	1	0.320	1	0.061
		CT	0.92 (0.62–1.36)		0.83 (0.57–1.19)		0.50 (0.27–0.92)	
		TT	1.07 (0.40–2.86)		1.60 (0.60–4.26)		0.60 (0.17–2.16)	
	Dominant	CC	1	0.730	1	0.480	1	**0.019**
		TT/TC	0.94 (0.64–1.36)		0.88 (0.62–1.26)		**0.51 (0.29–0.91)**	
	Recessive	CC/TC	1	0.850	1	0.270	1	0.600

Table 5 (continued)

SNP	Model	Genotype	ER		PR		HER2	
			OR (95% CI)	*p*	OR (95% CI)	*p*	OR (95% CI)	*p*
		TT	1.10 (0.42–2.91)		1.70 (0.64–4.51)		0.72 (0.20–2.57)	
	Additive	–	0.96 (0.70–1.33)	0.820	0.96 (0.71–1.31)	0.810	**0.60 (0.37–0.98)**	**0.032**
rs2181559	Allele	T	1		1		1	
		A	0.94 (0.71–1.24)	0.661	0.97 (0.75–1.26)	0.835	**0.62 (0.41–0.94)**	**0.025**
	Codominant	TT	1	0.690	1	0.940	1	0.091
		TA	1.03 (0.70–1.51)		0.94 (0.65–1.35)		0.62 (0.36–1.07)	
		AA	0.78 (0.41–1.48)		0.97 (0.52–1.82)		0.44 (0.16–1.20)	
	Dominant	TT	1	0.910	1	0.740	1	**0.037**
		TA/AA	0.98 (0.68–1.42)		0.94 (0.66–1.33)		**0.58 (0.35–0.97)**	
	Recessive	TT/TA	1	0.400	1	0.990	1	0.180
		AA	0.77 (0.42–1.41)		1.01 (0.55–1.83)		0.53 (0.20–1.42)	
	Additive	–	0.93 (0.70–1.24)	0.640	0.97 (0.74–1.26)	0.800	**0.64 (0.43–0.97)**	**0.029**
rs10965059	Allele	C	1		1		1	
		T	0.95 (0.60–1.5)	0.825	1.20 (0.78–1.85)	0.417	0.94 (0.50–1.78)	0.851
	Codominant	CC	1	0.820	1	0.600	1	0.820
		CT	0.87 (0.52–1.46)		1.07 (0.66–1.75)		0.82 (0.38–1.79)	
		TT	1.31 (0.26–6.62)		2.17 (0.43–11.00)		1.33 (0.25–7.13)	
	Dominant	CC	1	0.690	1	0.600	1	0.730
		TT/TC	0.90 (0.55–1.48)		1.13 (0.71–1.82)		0.88 (0.43–1.81)	
	Recessive	CC/TC	1	0.720	1	0.330	1	0.720
		TT	1.34 (0.26–6.76)		2.14 (0.42–10.87)		1.37 (0.26–7.31)	
	Additive	–	0.95 (0.61–1.47)	0.810	1.17 (0.77–1.78)	0.460	0.95 (0.52–1.73)	0.860
rs10965064	Allele	C	1		1		1	
		G	0.89 (0.69–1.17)	0.410	0.88 (0.68–1.13)	0.306	1.28 (0.89–1.83)	0.184
	Codominant	CC	1	0.600	1	0.360	1	0.330
		CG	0.82 (0.55–1.22)		0.76 (0.52–1.11)		1.18 (0.66–2.11)	
		CC	0.84 (0.48–1.47)		0.82 (0.48–1.40)		1.73 (0.85–3.55)	
	Dominant	CC	1	0.320	1	0.160	1	0.310
		CG/GG	0.82 (0.56–1.20)		0.78 (0.54–1.11)		1.32 (0.77–2.26)	
	Recessive	CC/CG	1	0.800	1	0.860	1	0.170
		GG	0.94 (0.56–1.57)		0.96 (0.59–1.56)		1.57 (0.83–2.97)	
	Additive	–	0.89 (0.69–1.16)	0.410	0.87 (0.68–1.12)	0.280	1.30 (0.91–1.86)	0.150

OR: odds ratio; 95% CI: 95% confidence interval

"/" indicates that the data is not available. *P* was calculated by chi square test

Bold indicated that $p < 0.05$ meant the data was statistically significant

Discussion

In recent years, it has been recognized that most tumor formations are under the combined effects of environmental and genetic factors. According to research, the occurrence of tumors may be the result of the superposition of multiple microscopic susceptibility genes [30], which may affect the metabolism of carcinogens, repair DNA damages, regulate hormone levels, and protect the immune function. Although numerous studies have been published on genetic associations with BC and, genetic effects of *MIR31HG* on cancer, few studies are concerned with available whether *MIR31HG* could serve as a candidate gene for BC. To the best of our knowledge, this study is the basic and fundamental one to analyze the association between *MIR31HG* gene polymorphisms and BC risk in Chinese women.

Through studies on *MIR31HG* related the functions in a BC mouse model, Augoff et al. confirmed that the changes in the expression level of this gene will facilitate tumor invasion and eventually metastasis [24]. Shi et al. also proved that knocking of *MIR31HG* could inhibit tumor growth in vivo, which showed that the expression level of *MIR31HG* could be referred as a diagnostic and prognostic marker for BC [25]. In this study, we assessed

Table 6 Stratified analysis of age at menarche, number of births and menopausal status in case group

SNP	Model	Genotype	Age of menarche		Number of births		Menopausal status	
			OR (95% CI)	*p*	OR (95% CI)	*p*	OR (95% CI)	*p*
rs79988146	Allele	T	1		1		1	
		C	1.35 (0.72–2.55)	0.347	1.27 (0.71–2.27)	0.423	1.22 (0.65–2.30)	0.532
	Codominant	TT	1	0.230	1	0.460	1	0.970
		CT	1.63 (0.83–3.20)		1.03 (0.54–1.94)		1.03 (0.40–2.69)	
		CC	/		/		/	
	Dominant	TT	1	0.210	1	0.790	1	0.930
		CC/TC	1.54 (0.79–3.01)		1.09 (0.58–2.03)		1.04 (0.40–2.71)	
	Recessive	TT/TC	1	0.320	1	/	1	/
		CC	/		/		/	
	Additive	–	1.40 (0.75–2.64)	0.300	1.14 (0.63–2.07)	0.660	1.05 (0.41–2.70)	0.920
rs1332184	Allele	G	1		1		1	
		A	**1.50 (1.08–2.09)**	**0.017**	1.01 (0.75–1.35)	0.969	1.17 (0.85–1.61)	0.342
	Codominant	GG	1	**0.048**	1	0.880	1	0.430
		GA	1.41 (0.90–2.20)		0.93 (0.63–1.37)		1.46 (0.82–2.63)	
		AA	**2.58 (1.15–5.79)**		1.10 (0.49–2.47)		1.02 (0.33–3.15)	
	Dominant	GG	1	**0.044**	1	0.780	1	0.250
		GA/AA	**1.55 (1.01–2.37)**		0.95 (0.65–1.38)		1.38 (0.80–2.41)	
	Recessive	GG/GA	1	**0.049**	1	0.740	1	0.830
		AA	**2.24 (1.03–4.91)**		1.14 (0.52–2.51)		0.88 (0.29–2.67)	
	Additive	–	**1.52 (1.08–2.12)**	**0.016**	0.98 (0.72–1.34)	0.920	1.21 (0.78–1.89)	0.400
rs72703442	Allele	C	1		1		1	
		A	1.11 (0.73–1.69)	0.635	0.88 (0.61–1.27)	0.502	1.15 (0.77–1.72)	0.490
	Codominant	CC	1	0.790	1	0.420	1	0.570
		CA	1.08 (0.66–1.75)		0.93 (0.61–1.43)		1.20 (0.64–2.27)	
		AA	1.81 (0.30–11.08)		0.26 (0.03–2.49)		3.16 (0.28–35.51)	
	Dominant	CC	1	0.680	1	0.600	1	0.460
		CA/AA	1.11 (0.69–1.78)		0.89 (0.59–1.36)		1.26 (0.68–2.36)	
	Recessive	CC/CA	1	0.540	1	0.200	1	0.370
		AA	1.78 (0.29–10.83)		0.27 (0.03–2.53)		3.03 (0.27–33.76)	
	Additive	–	1.13 (0.73–1.75)	0.600	0.86 (0.58–1.27)	0.450	1.30 (0.73–2.32)	0.370
rs2025327	Allele	T	1		1		1	
		C	1.33 (0.84–2.11)	0.220	1.26 (0.84–1.90)	0.265	1.21 (0.77–1.91)	0.400
	Codominant	TT	1	0.230	1	0.550	1	0.530
		CT	1.15 (0.68–1.95)		1.24 (0.77–1.98)		1.31 (0.65–2.63)	
		CC	4.64 (0.76–28.32)		1.82 (0.32–10.40)		0.37 (0.03–4.30)	
	Dominant	TT	1	0.390	1	0.310	1	0.580
		CC/TC	1.26 (0.75–2.09)		1.27 (0.80–2.01)		1.21 (0.61–2.38)	
	Recessive	TT/TC	1	0.100	1	0.520	1	0.400
		CC	4.51 (0.74–27.46)		1.75 (0.31–9.96)		0.35 (0.03–4.07)	
	Additive	–	1.34 (0.84–2.13)	0.230	1.26 (0.83–1.92)	0.280	1.10 (0.59–2.04)	0.770
rs55683539	Allele	C	1		1		1	
		T	1.14 (0.80–1.63)	0.462	0.93 (0.69–1.28)	0.670	0.96 (0.69–1.33)	0.791
	Codominant	CC	1	0.640	1	**0.031**	1	0.310
		CT	1.05 (0.67–1.65)		1.26 (0.85–1.87)		0.90 (0.51–1.60)	
		TT	1.67 (0.58–4.81)		**0.32 (0.11–0.93)**		2.81 (0.67–11.74)	
	Dominant	CC	1	0.650	1	0.640	1	0.970
		TT/TC	1.10 (0.72–1.70)		1.09 (0.75–1.60)		1.01 (0.58–1.75)	
	Recessive	CC/TC	1	0.360	1	**0.017**	1	0.140

Table 6 (continued)

SNP	Model	Genotype	Age of menarche		Number of births		Menopausal status	
			OR (95% CI)	*p*	OR (95% CI)	*p*	OR (95% CI)	*p*
		TT	1.64 (0.58–4.67)		**0.29 (0.10–0.86)**		2.92 (0.71–12.00)	
	Additive	-	1.14 (0.79–1.65)	0.490	0.94 (0.68–1.30)	0.690	1.14 (0.71–1.83)	0.600
rs2181559	Allele	T	1		1		1	
		A	1.27 (0.93–1.74)	0.136	1.04 (0.79–1.36)	0.772	1.13 (0.84–1.51)	0.433
	Codominant	TT	1	0.140	1	0.750	1	0.500
		TA	1.05 (0.67–1.64)		1.14 (0.78–1.69)		1.23 (0.70–2.17)	
		AA	2.05 (1.00–4.18)		0.96 (0.49–1.86)		1.79 (0.63–5.10)	
	Dominant	TT	1	0.450	1	0.580	1	0.340
		TA/AA	1.18 (0.77–1.81)		1.11 (0.76–1.61)		1.30 (0.75–2.24)	
	Recessive	TT/TA	1	0.051	1	0.730	1	0.360
		AA	2.00 (1.02–3.93)		0.89 (0.47–1.69)		1.60 (0.58–4.36)	
	Additive	–	1.29 (0.93–1.78)	0.130	1.04 (0.78–1.39)	0.790	1.29 (0.84–1.99)	0.250
rs10965059	Allele	C	1		1		1	
		T	1.12 (0.68–1.87)	0.652	1.03 (0.66–1.61)	0.911	0.91 (0.57–1.47)	0.707
	Codominant	CC	1	0.890	1	0.032	1	0.076
		CT	1.15 (0.64–2.07)		1.43 (0.85–2.43)		0.68 (0.32–1.47)	
		TT	0.94 (0.18–4.81)		0.12 (0.01–1.11)		13.61 (0.69–269.91)	
	Dominant	CC	1	0.670	1	0.460	1	0.720
		TT/TC	1.13 (0.64–1.98)		1.21 (0.73–2.00)		0.87 (0.42–1.81)	
	Recessive	CC/TC	1	0.920	1	0.024	1	0.040
		TT	0.92 (0.18–4.68)		0.12 (0.01–1.05)		14.48 (0.73–288.35)	
	Additive	–	1.09 (0.67–1.76)	0.740	1.02 (0.65–1.60)	0.940	1.09 (0.58–2.05)	0.790
rs10965064	Allele	C	1		1		1	
		G	1.21 (0.89–1.64)	0.226	0.78 (0.60–1.01)	0.059	0.97 (0.73–1.29)	0.844
	Codominant	CC	1	0.410	1	0.050	1	0.220
		CG	1.34 (0.84–2.14)		1.04 (0.69–1.56)		0.77 (0.43–1.38)	
		CC	1.38 (0.73–2.61)		0.53 (0.30–0.94)		1.57 (0.68–3.59)	
	Dominant	CC	1	0.180	1	0.510	1	0.760
		CG/GG	1.35 (0.87–2.10)		0.88 (0.60–1.29)		0.92 (0.53–1.59)	
	Recessive	CC/CG	1	0.600	1	**0.015**	1	0.130
		GG	1.17 (0.66–2.10)		**0.52 (0.31–0.89)**		1.80 (0.83–3.89)	
	Additive	–	1.20 (0.89–1.63)	0.230	0.79 (0.61–1.04)	0.088	1.11 (0.76–1.64)	0.580

OR: odds ratio; 95% CI: 95% confidence interval

"/" indicates that the data is not available

P was calculated by chi square test

Bold indicated that $p < 0.05$ meant the data was statistically significant

the relationship between *MIR31HG* gene polymorphisms and the risk of BC, and found some related targets on evaluating BC risk. The results showed that rs72703442, rs55683539 and rs2181559 on *MIR31HG* were significantly associated with a reduced risk of BC in Chinese women. And, after age stratification, these three SNPs (rs72703442, rs55683539, and rs2181559) on *MIR31HG* were significantly associated with a reduced risk of BC in Chinese women ≥ 52 years old. However, no SNPs were found to be associated with BC risk in Chinese women in the < 52-year-old stratification.

By analyzing gene polymorphism, Xia et al. concluded that BC risk was evaluated according to the different ER and PR states [8]. Zhou et al. also believed that the status of ER and PR is still the key to determine the type of BC adjuvant therapy, because estrogen stimulates ER-mediated transcription to increase cell proliferation, thereby increasing the number of DNA replication errors [31]. Above conclusions proved that clinical indicators especially in the status of ER and PR, have a certain influence on the development of BC.

ER and PR stratification showed that rs79988146 on *MIR31HG* was positively correlated with ER and PR in

Table 7 Stratified analysis of LN metastasis, clinical stage, and tumor size in case group

SNP	Model	Genotype	LN metastasis		Clinical stage		Tumor size	
			OR (95% CI)	*p*	OR (95% CI)	*p*	OR (95% CI)	*p*
rs79988146	Allele	T	1		1		1	
		C	**1.79 (1.03–3.11)**	**0.038**	0.89 (0.48–1.65)	0.718	0.70 (0.38–1.28)	0.245
	Codominant	TT	1	0.092	1	0.660	1	0.120
		CT	1.92 (1.06–3.48)		0.79 (0.40–1.54)		0.57 (0.30–1.10)	
		CC	1.08 (0.07–17.65)		2.25 (0.14–36.89)		/	
	Dominant	TT	1	**0.032**	1	0.560	1	0.140
		CC/TC	**1.88 (1.05–3.36)**		0.83 (0.43–1.58)		0.62 (0.32–1.17)	
	Recessive	TT/TC	1	1	1	0.560	1	/
		CC	1.00 (0.06–16.32)		2.31 (0.14–37.95)		/	
	Additive	–	**1.76 (1.01–3.08)**	**0.042**	0.88 (0.47–1.61)	0.670	1	0.120
rs1332184	Allele	G	1		1		1	
		A	1.24 (0.94–1.63)	0.132	**1.43 (1.06–1.92)**	**0.020**	1.27 (0.90–1.78)	0.172
	Codominant	GG	1	0.200	1	**0.036**	1	0.380
		GA	1.09 (0.76–1.55)		1.19 (0.80–1.78)		1.19 (0.78–1.84)	
		AA	2.00 (0.93–4.33)		**2.69 (1.27–5.72)**		1.79 (0.70–4.60)	
	Dominant	GG	1	0.360	1	0.120	1	0.270
		GA/AA	1.17 (0.83–1.65)		1.35 (0.92–1.98)		1.26 (0.83–1.91)	
	Recessive	GG/GA	1	0.080	1	**0.015**	1	0.260
		AA	1.94 (0.91–4.14)		**2.50 (1.20–5.22)**		1.67 (0.66–4.23)	
	Additive	–	1.23 (0.93–1.62)	0.150	**1.42 (1.04–1.92)**	**0.026**	1.26 (0.90–1.77)	0.180
rs72703442	Allele	C	1		1		1	
		A	1.03 (0.73–1.46)	0.858	1.28 (0.88–1.86)	0.192	1.46 (0.94–2.26)	0.091
	Codominant	CC	1	0.900	1	0.260	1	0.100
		CA	1.05 (0.71–1.54)		1.40 (0.92–2.13)		1.68 (1.02–2.76)	
		AA	0.70 (0.12–4.26)		0.59 (0.07–5.40)		0.74 (0.12–4.53)	
	Dominant	CC	1	0.870	1	0.150	1	0.050
		CA/AA	1.03 (0.71–1.51)		1.36 (0.89–2.07)		1.61 (0.99–2.61)	
	Recessive	CC/CA	1	0.690	1	0.570	1	0.660
		AA	0.69 (0.11–4.20)		0.54 (0.06–4.93)		0.66 (0.11–4.01)	
	Additive	–	1.01 (0.71–1.45)	0.940	1.28 (0.87–1.90)	0.220	1.48 (0.94–2.34)	0.084
rs2025327	Allele	T	1		1		1	
		C	1.27 (0.86–1.87)	0.225	1.36 (0.90–2.05)	0.139	1.51 (0.91–2.49)	0.106
	Codominant	TT	1	0.410	1	0.300	1	0.083
		CT	1.34 (0.87–2.06)		1.32 (0.83–2.10)		1.35 (0.79–2.31)	
		CC	1.07 (0.21–5.40)		2.50 (0.49–12.61)		/	
	Dominant	TT	1	0.190	1	0.180	1	0.160
		CC/TC	1.32 (0.87–2.02)		1.37 (0.87–2.16)		1.46 (0.86–2.48)	
	Recessive	TT/TC	1	0.980	1	0.310	1	/
		CC	1.02 (0.20–5.10)		2.36 (0.47–11.87)		/	
	Additive	–	1.27 (0.86–1.87)	0.230	1.38 (0.91–2.08)	0.130	1.52 (0.92–2.52)	0.091
rs55683539	Allele	C	1		1		1	
		T	0.92 (0.69–1.23)	0.591	1.11 (0.81–1.53)	0.526	1.15 (0.81–1.65)	0.434
	Codominant	CC	1	0.400	1	0.510	1	0.280
		CT	1.04 (0.73–1.48)		1.24 (0.83–1.84)		1.37 (0.88–2.12)	
		TT	0.54 (0.21–1.39)		0.82 (0.29–2.35)		0.74 (0.25–2.15)	
	Dominant	CC	1	0.880	1	0.380	1	0.240
		TT/TC	0.97 (0.69–1.38)		1.19 (0.81–1.75)		1.29 (0.84–1.96)	
	Recessive	CC/TC	1	0.180	1	0.600	1	0.450

Table 7 (continued)

SNP	Model	Genotype	LN metastasis		Clinical stage		Tumor size	
			OR (95% CI)	*p*	OR (95% CI)	*p*	OR (95% CI)	*p*
		TT	0.54 (0.21–1.36)		0.76 (0.27–2.16)		0.66 (0.23–1.91)	
	Additive	–	0.92 (0.68–1.23)	0.560	1.10 (0.79–1.54)	0.560	1.16 (0.80–1.67)	0.440
rs2181559	Allele	T	1		1		1	
		A	1.03 (0.80–1.32)	0.839	1.27 (0.96–1.67)	0.099	1.36 (0.99–1.86)	0.058
	Codominant	TT	1	0.850	1	0.170	1	0.130
		TA	0.96 (0.68–1.37)		1.15 (0.77–1.72)		1.45 (0.95–2.22)	
		AA	1.15 (0.62–2.12)		1.87 (0.98–3.58)		1.80 (0.80–4.03)	
	Dominant	TT	1	0.970	1	0.250	1	0.051
		TA/AA	0.99 (0.71–1.39)		1.25 (0.85–1.84)		1.50 (1.00–2.26)	
	Recessive	TT/TA	1	0.600	1	0.080	1	0.300
		AA	1.17 (0.65–2.10)		1.74 (0.95–3.21)		1.50 (0.69–3.27)	
	Additive	–	1.03 (0.79–1.34)	0.840	1.29 (0.96–1.73)	0.089	**1.39 (1.00–1.94)**	**0.046**
rs10965059	Allele	C	1		1		1	
		T	0.71 (0.47–1.09)	0.114	0.85 (0.52–1.37)	0.492	0.87 (0.53–1.43)	0.577
	Codominant	CC	1	0.360	1	0.600	1	0.810
		CT	0.73 (0.45–1.17)		0.95 (0.56–1.63)		0.95 (0.53–1.69)	
		TT	0.64 (0.15–2.74)		0.37 (0.04–3.16)		0.60 (0.13–2.76)	
	Dominant	CC	1	0.150	1	0.680	1	0.730
		TT/TC	0.72 (0.45–1.14)		0.90 (0.53–1.51)		0.91 (0.52–1.57)	
	Recessive	CC/TC	1	0.590	1	0.320	1	0.530
		TT	0.67 (0.16–2.88)		0.38 (0.04–3.18)		0.61 (0.13–2.78)	
	Additive	–	0.75 (0.50–1.12)	0.160	0.86 (0.54–1.38)	0.530	0.89 (0.55–1.42)	0.620
rs10965064	Allele	C	1		1		1	
		G	1.21 (0.95–1.55)	0.130	1.25 (0.95–1.64)	0.114	0.83 (0.62–1.12)	0.224
	Codominant	CC	1	0.350	1	0.330	1	0.330
		CG	1.13 (0.78–1.63)		1.17 (0.77–1.78)		0.95 (0.61–1.48)	
		CC	1.46 (0.87–2.46)		1.53 (0.87–2.70)		0.64 (0.35–1.17)	
	Dominant	CC	1	0.290	1	0.250	1	0.470
		CG/GG	1.20 (0.85–1.70)		1.25 (0.85–1.85)		0.86 (0.56–1.30)	
	Recessive	CC/CG	1	0.200	1	0.200	1	0.140
		GG	1.37 (0.85–2.22)		1.41 (0.84–2.37)		0.66 (0.38–1.14)	
	Additive	–	1.19 (0.93–1.52)	0.160	1.22 (0.93–1.61)	0.150	0.83 (0.62–1.11)	0.200

OR: odds ratio; 95% CI: 95% confidence interval

"/" indicates that the data is not available

P was calculated by chi square test

Bold indicated that $p < 0.05$ meant the data was statistically significant

Chinese female BC patients. In addition, stratification by age at menarche only found that rs1332184 on *MIR31HG* was associated with age at menarche in BC patients. Number of births stratification showed that rs10965064 on *MIR31HG* was negatively correlated with BC patients' Number of births.

BC is a complex disease affected by the interaction of factors such as heredity. Multi-gene or SNP-SNP interaction studies may help to discover the risk factors of BC. Therefore, we performed MDR analysis to determine the potential SNP-SNP interactions among the eight SNPs in the *MIR31HG* gene polymorphisms. SNP-SNP interaction analysis indicated a strong interaction between SNPs on *MIR31HG* for BC sensitivity. In addition, in the multi-site model, the combination of rs79988146, rs1332184, rs72703442, rs2025327, rs55683539, rs2181559, rs10965059, and rs10965064 is the best multi-site model for predicting BC sensitivity.

However, there are also some limitations that cannot be neglected in our study. First of all, this is a hospital-based case and control study, which may have some inevitable sample selection bias and the absence of partial sample

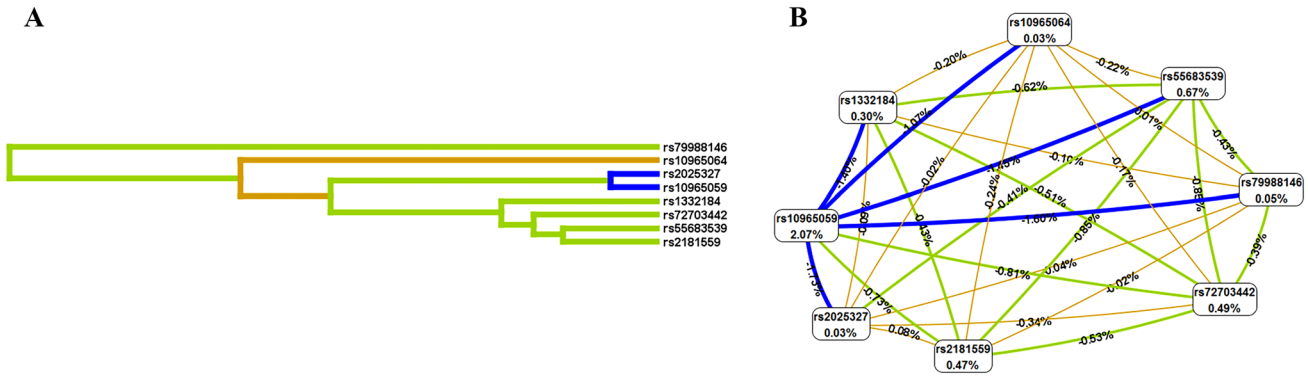

Fig. 1 The dendogram (**A**) and fruchterman Rheingold (**B**) of *MIR31HG* SNP-SNP interaction for BC risk. **A** Short connections among nodes represent stronger redundant interactions. **B** Negative percent entropy indicates redundancy

Table 8 SNP–SNP interaction models of the *MIR31HG* gene the predisposition of BC

Model	Testing Bal. Acc	CVC	*p*
rs55683539	0.5038	5/10	**0.020**
rs55683539,rs10965059	0.5509	9/10	**< 0.001**
rs55683539,rs10965059,rs10965064	0.5057	5/10	**< 0.001**
rs1332184,rs55683539,rs10965059,rs10965064	0.4858	4/10	**< 0.001**
rs1332184,rs2025327,rs55683539,rs10965059,rs10965064	0.5009	7/10	**< 0.001**
rs79988146,rs1332184,rs2025327,rs55683539,rs10965059,rs10965064	0.5236	5/10	**< 0.001**
rs79988146,rs1332184,rs72703442,rs2025327,rs55683539,rs10965059,rs10965064	0.5123	7/10	**< 0.001**
rs79988146,rs1332184,rs72703442,rs2025327,rs55683539,rs2181559,rs10965059,rs10965064	0.5179	10/10	**< 0.001**

P values were calculated using χ^2 tests

Bold indicated that $p < 0.05$ meant the data was statistically significant

MDR multifactor dimensionality reduction, *Bal. Acc.* balanced accuracy, *CVC* cross-validation consistency, *OR* odds ratio, *CI* confidence interval

information. Second, our study has a limited generalizability because all participants were Han Chinese. Therefore, further well-designed study with a larger population or other ethnic groups is needed to confirm our findings. Then, our sample size was too insufficient to support stratified analysis of tumor subtypes. Finally, due to other information was incomplete, we didn't analyze other risk factors for BC, such as lifestyle, family history, and other benign breast lesions. Therefore, population-based studies with a large amount of sample size and more complete information will be needed in the future to improve and enhance the accuracy of assessments and to explore the interaction between genetic variants and these factors.

Conclusion

In conclusion, this study firstly shows that *MIR31HG* gene polymorphisms are associated with a reduced risk of BC in Chinese women, and provides a theoretical basis for future explorations of the relationship between *MIR31HG* gene and BC risk in different populations. These findings can provide new biological insights for understanding the role of *MIR31HG* in the occurrence of BC.

Acknowledgements First of all, we thank all authors for their contributions and supports. Then, we are grateful to all participants for providing blood samples.

Author contributions XZ: conceived and designed the experiments; YW and XW: performed the experiments; CZ and ZZ: analyzed the data; YC: contributed reagents/materials/analysis tools; ZB: prepared the figures and/or tables; XZ, YW and XW: drafted the work or revised it critically for important content. All authors have read and approved the manuscript.

Funding No.

Availability of data and material All data obtained from the current study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors have declared that they have no competing interests.

Ethics approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Consent for publication Written informed consent was obtained from the patient for publication of this report.

References

1. Fan L, Strasser-Weippl K, Li J-J et al (2014) Breast cancer in China. Lancet Oncol 15(7):e279–e289
2. Huang BF, Tzeng HE, Chen PC et al (2018) HMGB1 genetic polymorphisms are biomarkers for the development and progression of breast cancer. Int J Med Sci 15(6):580–586
3. Kresovich JK, Gann PH, Erdal S et al (2018) Candidate gene DNA methylation associations with breast cancer characteristics and tumor progression. Epigenomics 10(4):367–378
4. Wang Y, Zhang H, Lin M et al (2018) Association of FGFR2 and PI3KCA genetic variants with the risk of breast cancer in a Chinese population. Cancer Manag Res 10:1305–1311
5. Yan Y, Zhang X (2017) The association between CD28 gene rs3116496 polymorphism and breast cancer risk in Chinese women. Biosci Rep 37(6):BSR20170884
6. Zhang M, Jin M, Yu Y et al (2012) Associations of miRNA polymorphisms and female physiological characteristics with breast cancer risk in Chinese population. Eur J Cancer Care (Engl) 21(2):274–280
7. Wu L, Shi W, Long J et al (2018) A transcriptome-wide association study of 229,000 women identifies new candidate susceptibility genes for breast cancer. Nat Genet 50(7):968–978
8. Xia P, Li B, Geng T et al (2015) FGFR2 gene polymorphisms are associated with breast cancer risk in the Han Chinese population. Am J Cancer Res 5(5):1854–1861
9. Tian T, Wang M, Zheng Y et al (2018) Association of two FOXP3 polymorphisms with breast cancer susceptibility in Chinese Han women. Cancer Manag Res 10:867–872
10. Zheng W, Zhang B, Cai Q et al (2013) Common genetic determinants of breast-cancer risk in East Asian women: a collaborative study of 23 637 breast cancer cases and 25 579 controls. Hum Mol Genet 22(12):2539–2550
11. de Moraes CL, Cruz EMN, Valoyes MAV et al (2021) AGR2 and AGR3 play an important role in the clinical characterization and prognosis of basal like breast cancer. Clin Breast Cancer 22(2):e242–e252
12. Zhu Y, Wang T, Tong Y et al (2021) 21-gene recurrence assay associated with favorable metabolic profiles in HR-positive, HER2-negative early-stage breast cancer patients. Front Endocrinol 12:725161
13. Invernizzi M, Lopez G, Michelotti A et al (2020) Integrating biological advances into the clinical management of breast cancer related lymphedema. Front Oncol 10:422
14. de Sire A, Losco L, Cisari C et al (2020) Axillary web syndrome in women after breast cancer surgery referred to an Oncological Rehabilitation Unit: which are the main risk factors? A retrospective case-control study. Eur Rev Med Pharmacol Sci 24(15):8028–8035
15. Suvanto M, Beesley J, Blomqvist C et al (2020) SNPs in lncRNA regions and breast cancer risk. Front Genet 11:550
16. Cui P, Zhao Y, Chu X et al (2018) SNP rs2071095 in LincRNA H19 is associated with breast cancer risk. Breast Cancer Res Treat 171(1):161–171
17. Qin J, Ning H, Zhou Y et al (2018) LncRNA *MIR31HG* overexpression serves as poor prognostic biomarker and promotes cells proliferation in lung adenocarcinoma. Biomed Pharmacother 99:363–368
18. Wang B, Jiang H, Wang L et al (2017) Increased *MIR31HG* lncRNA expression increases gefitinib resistance in non-small cell lung cancer cell lines through the EGFR/PI3K/AKT signaling pathway. Oncol Lett 13(5):3494–3500
19. Eide PW, Eilertsen IA, Sveen A et al (2019) Long noncoding RNA *MIR31HG* is a bona fide prognostic marker with colorectal cancer cell-intrinsic properties. Int J Cancer 144(11):2843–2853
20. He A, Chen Z, Mei H et al (2016) Decreased expression of LncRNA *MIR31HG* in human bladder cancer. Cancer Biomark 17(2):231–236
21. Shih J-W, Chiang W-F, Wu ATH et al (2017) Long noncoding RNA LncHIFCAR/*MIR31HG* is a HIF-1α co-activator driving oral cancer progression. Nat Commun 8(1):15874
22. Sun K, Zhao X, Wan J et al (2018) The diagnostic value of long non-coding RNA *MIR31HG* and its role in esophageal squamous cell carcinoma. Life Sci 202:124–130
23. Ren ZP, Chu XY, Xue ZQ et al (2017) Down-regulation of lncRNA *MIR31HG* correlated with aggressive clinicopathological features and unfavorable prognosis in esophageal squamous cell carcinoma. Eur Rev Med Pharmacol Sci 21(17):3866–3870
24. Augoff K, McCue B, Plow EF et al (2012) miR-31 and its host gene lncRNA LOC554202 are regulated by promoter hypermethylation in triple-negative breast cancer. Mol Cancer 11:5
25. Shi Y, Lu J, Zhou J et al (2014) Long non-coding RNA Loc554202 regulates proliferation and migration in breast cancer cells. Biochem Biophys Res Commun 446(2):448–453
26. Patil N, Berno AJ, Hinds DA et al (2001) Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science (New York, NY) 294(5547):1719–1723
27. Ren HT, Li YM, Wang XJ et al (2016) PD-1 rs2227982 polymorphism is associated with the decreased risk of breast cancer in northwest Chinese women: a hospital-based observational study. Medicine 95(21):e3760
28. Leem S, Park T (2017) An empirical fuzzy multifactor dimensionality reduction method for detecting gene-gene interactions. BMC Genom 18(Suppl 2):115
29. Shahin NN, Abd-Elwahab GT, Tawfiq AA et al (2020) Potential role of aryl hydrocarbon receptor signaling in childhood obesity. Biochim Biophys Acta 1865(8):158714
30. Curran JE, Weinstein SR, Griffiths LR (2000) Polymorphisms of glutathione S-transferase genes (GSTM1, GSTP1 and GSTT1) and breast cancer susceptibility. Cancer Lett 153(1–2):113–120
31. Zhou L, He N, Feng T et al (2015). Association of five single nucleotide polymorphisms at 6q25.1 with breast cancer risk in northwestern China. Am J Cancer Res 5(8):2467–2475