

RESEARCH

Open Access



# CYP4B1 polymorphisms and the risk of breast cancer in Chinese women: a case-control study

Yanhai Yin<sup>1,2†</sup>, Liangqian Tong<sup>3†</sup>, Zhenling Wan<sup>4†</sup>, Yanfang Sui<sup>5†</sup>, Fen Li<sup>2</sup>, Qian Huang<sup>2</sup> and Xinhan Zhao<sup>1\*</sup>

## Abstract

**Background** Breast cancer (BC) is one of the malignant diseases threatening the life and health of women worldwide. The *CYP4B1* gene was abnormally expressed in BC and was associated with the prognosis of BC patients. This study aimed to explore the relationship between *CYP4B1* single nucleotide polymorphisms (SNPs) and BC risk in Chinese women.

**Methods** A case-control study of 1,143 women (571 patients and 572 healthy individuals) was conducted. Rs2297813 G/T, rs12142787 G/A, and rs3766197 C/T in *CYP4B1* were selected and genotyped by MassARRAY system. The relationships between these SNPs and the risk of BC were assessed by logistic regression analysis. In addition, multi-factor dimensionality reduction (MDR) was used to analyze SNP-SNP interactions.

**Results** *CYP4B1* rs2297813 had a risk-increasing effect on BC in women with body mass index (BMI)  $\leq 24$  kg/m<sup>2</sup> (OR = 1.72,  $p = 0.026$ ). *CYP4B1* rs12142787 was associated with an increased BC risk in smokers (AA: OR = 1.32,  $p = 0.045$ ). Among non-drinkers, rs2297813 (OR = 1.69,  $p = 0.009$ ) and rs12142787 (OR = 1.51,  $p = 0.020$ ) were related to an increased incidence of BC. *CYP4B1* rs3766197 (OR = 1.61,  $p = 0.031$ ) was associated with a higher risk of advanced stages (III/IV stage) of BC. Besides, the contributions of *CYP4B1* rs2297813 (OR = 1.55,  $p = 0.021$ ) and rs12142787 (OR = 1.53,  $p = 0.033$ ) to BC risk might be associated with more than one birth in patients with BC. The three-locus model consisting of rs2297813, rs12142787, and rs3766197 was regarded as the best predictive model for BC risk.

**Conclusion** *CYP4B1* SNPs were associated with BC risk in Chinese women, especially in patients with BMI  $\leq 24$  kg/m<sup>2</sup>, smokers, non-drinkers, patients in advanced stages (III/IV stage), and patients who reproduced once. These findings shed light on the relationship between *CYP4B1* SNPs and BC risk in Chinese women.

**Keywords** Breast cancer, *CYP4B1*, Single nucleotide polymorphisms, Chinese women

<sup>†</sup>Yanhai Yin, Liangqian Tong, Zhenling Wan, and Yanfang Sui contributed equally to this work.

\*Correspondence:

Xinhan Zhao  
zhaoxinhan@mail.xjtu.edu.cn

<sup>1</sup>Department of Internal Medicine Oncology, The First Affiliated Hospital of Xi'an Jiaotong University School of Medicine, No.76, Yanta West Road, Xi'an City 710061, Shaanxi Province, China

<sup>2</sup>Department of Nuclear Medicine, Hainan General Hospital, Hainan Affiliated Hospital of Hainan Medical University, Haikou, Hainan 570208, China

<sup>3</sup>Department of Nuclear Medicine, Central South University Xiangya School Affiliated Haikou Hospital, Haikou, Hainan 570208, China

<sup>4</sup>Department of Pathology, Hainan Women and Children Medical Center, Haikou, Hainan 570208, China

<sup>5</sup>Department of Rehabilitation Medicine, Central South University Xiangya School Affiliated Haikou Hospital, Haikou, Hainan 570208, China



## Background

Breast cancer (BC), one of the most common malignant diseases in women worldwide, can be highly invasive and metastatic [1]. According to Global Cancer Statistics 2020, female breast cancer has surpassed lung cancer as the most commonly diagnosed cancer, with an estimated 2.3 million new cases (11.7%) [2]. Female breast cancer in China accounts for approximately 18% (0.42 million) of the global breast cancer deaths [3]. According to data published in 2022, an estimated 43,250 women died from BC in the United States, in other words, there are nearly 118 BC-related deaths in women per day [4]. And more than 90% of BC-related deaths are related to cancer cell metastasis [5]. Previous studies have shown that breast tumor cells can easily metastasize to some important organs, such as the lung, liver, bone, or brain through blood vessels and lymphatic vessels [1]. BC is an intricate disease characterized by genetic, epigenetic, and phenotypic changes [6]. At present, the carcinogenic mechanism of BC is not fully understood. Some studies have indicated that the occurrence of BC is influenced by genes, lifestyles, and other factors [7–9]. As revealed by investigations, age, family history, obesity, oral contraceptives, status menopausal, smoking, alcohol consumption, lifestyle, and genetics factors are significantly linked to the occurrence and development BC [10], especially genetics factors. Currently, some association studies have identified several susceptible single nucleotide polymorphisms (SNPs) correlated with BC risk [11–13].

Cytochrome P450 (CYP), a superfamily of cysteine heme monooxygenases, is responsible for bio-transforming a large number of endogenous and exogenous substances via oxidative reactions [14]. As one of the CYP isoforms, *CYP4B1*, mainly expressed in lung and bladder tissues can also be detected in BC tissues and cell lines (<https://gtexportal.org/home/gene/CYP4B1>). Studies have shown that the alteration of *CYP4B1* gene expression may be associated with certain cancers [15], suggesting that *CYP4B1* may work in the development of cancer. The high *CYP4B1* expression increased the chances of bioactivation of carcinogenic aromatic amines, thereby leading to a high risk of cancer growth [15–17]. Gene mutations are associated with protein expression levels, and gene polymorphisms may play a role in disease susceptibility by affecting gene expression and/or enzyme activity [18, 19]. Previous studies have reported that *CYP4B1* polymorphisms are significantly correlated with the risk of various cancers, such as gastric cancer [20], bladder cancer [21], and lung cancer [22]. Iscan et al. have pointed out that *CYP4B1* mRNA is expressed in both breast tumor and tumor-free tissues by reverse transcription-polymerase chain reaction (RT-PCR) [23]. However, to date, little is known about the association of *CYP4B1* SNPs with BC risk.

This case-control study of 1,143 Chinese women was conducted to evaluate the association between *CYP4B1* SNPs and the risk of BC in the Chinese Han population. Additionally, our study revealed the relationship between these SNPs and BC susceptibility by stratified analyses based on multiple factors (demographic and clinicopathological features), which provides a theoretical basis for the prevention and early detection of BC.

## Materials and methods

### Study population and clinical data

In this study, a total of 571 Chinese women diagnosed with BC by clinical examinations and Breast Imaging Reporting and Data System (BI-RADS) were randomly selected as the case group, and 572 healthy Chinese women with no personal or family history of cancer were chosen as the control group. There was no genetic relationship between cases and controls. Patients aged <20 years or >70 years with a prior history of cancer, metastatic cancer, radiotherapy, or chemotherapy were excluded. All the healthy volunteers were recruited from the health examination center in the same hospital during the same period as controls. The demographic information (age, weight, height, smoking and drinking habits, reproductive numbers, and menopause) and clinical information (estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor-2 (HER2), and cell proliferation marker (Ki67), lymph node metastasis (LNM) and staging) about all participants were collected from questionnaires and medical records, respectively. Each participant completed an informed consent form after learning about the purpose of the study. This study was approved by the Ethics Committee of the First Affiliated Hospital of Xi'an Jiaotong University School of Medicine, and all experimental methods were strictly in compliance with the 1964 Helsinki Declaration.

ER, PR, HER2, and cell proliferation marker (Ki67) in patient tissue samples were stained by Immune-histochemical (IHC) assay. According to Breast Cancer, version 3.2018 [24], we diagnosed patients as ER-positive (ER<sup>+</sup>) or ER-negative (ER<sup>-</sup>), PR-positive (PR<sup>+</sup>) or PR-negative (PR<sup>-</sup>), and HER2-positive (HER2<sup>+</sup>) or HER2-negative (HER2<sup>-</sup>). BC cells with more than 25% Ki67 staining were Ki67-positive (Ki67<sup>+</sup>) or otherwise Ki67-negative (Ki67<sup>-</sup>). Tumor staging and lymph node metastasis (LNM) were identified based on the 2010 American Joint Committee on Cancer (AJCC) criteria [25].

### DNA extraction and SNP genotyping

Fasting venous blood samples were collected from all participants in the morning using a vacutainer containing EDTA. Genomic DNA was extracted and purified according to the kit instructions (GoldMag Co., Ltd.,

Xi'an, China). DNA was then stored in a refrigerator at -80 °C before the experiments. Three SNPs (rs2297813 G/T, rs12142787 G/A, and rs3766197 C/T) of *CYP4B1* were selected based on (1) the variations (113,617) of *CYP4B1* through the e!GRCh37 ([http://asia.ensembl.org/Homo\\_sapiens/Info/Index](http://asia.ensembl.org/Homo_sapiens/Info/Index)) database, (2) the variations (923) with global minor allele frequency (MAF)>0.05, (3) the biallelic variations (402), in which some variations (30) were randomly selected, and (4) combined MassARRAY primer design software, Hardy-Weinberg equilibrium>0.05, MAF>0.05 and the call rate>95% in our

study population.,. Furthermore, the MassARRAY system (Agena, San Diego, CA, USA) was applied for genotyping, and all primers were designed by MassARRAY Assay Design software.

**Data analysis**

The expression analysis of the *CYP4B1* gene in BC was predicted based on the GEPIA database (<http://gepia.cancer-pku.cn/index.html>). Kaplan-Meier Plotter database (<https://kmplot.com/analysis/>) was utilized to assess the correlation between the expression of *CYP4B1* and survival in patients with BC. HaploReg v4.1 (<https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>), RegulomeDB (<https://regulome.stanford.edu/regulome-search/>), and GTEx Portal database (<https://gtex-portal.org/home/>) were applied to predict the potential functions of SNPs.

Data analysis was performed by SPSS software version 21.0 (SPSS, Chicago, IL, USA), and two-sided  $p < 0.05$  indicated statistical significance. The differences in demographic characteristics (age, smoking, and drinking) between BC patients and healthy controls were analyzed by  $t$ -test/ $\chi^2$ -test. HWE was tested by  $\chi^2$ -test to determine whether the subjects in our study reached genetic equilibrium. Logistic regression models adjusted by age, drinking, and smoking were introduced to calculate odds ratios (ORs) and 95% confidence intervals (CIs) with the SNPStats online software (<https://www.snpstats.net/start.htm?q=snpstats/start.htm>). ORs and CIs represented the effect of *CYP4B1* SNPs on the risk of BC (OR=1: no effect; OR<1: protective factor; OR>1: risk factor). The association of SNP-SNP interactions with BC risk was evaluated by multi-factor dimensionality reduction (MDR) 3.0.2 software.

**Results**

**The characteristics of study participants**

The basic characteristics of participants are summarized in Table 1. The average ages of cases and controls were 52.11±10.14 years and 51.88±9.78 years, respectively. And no significant difference in age ( $p=0.700$ ), smoking ( $p=0.190$ ), and drinking ( $p=0.260$ ) between the two groups was detected. Among 571 patients, the majority were menopausal (42.4%), and had more than one reproduction (42.0%). The proportions of early-stage (stage I, II) tumors and late-stage (stage III, IV) tumors were 60.6% and 25.2%, respectively. The results of IHC showed that most cases were diagnosed as ER+ (63.4%), PR+ (54.8%), and HER2+ (22.1%), and had a high level of Ki67 (62.9%). In the BC group, there were 194 patients (34.0%) with LNM and 241 patients (42.2%) without LNM.

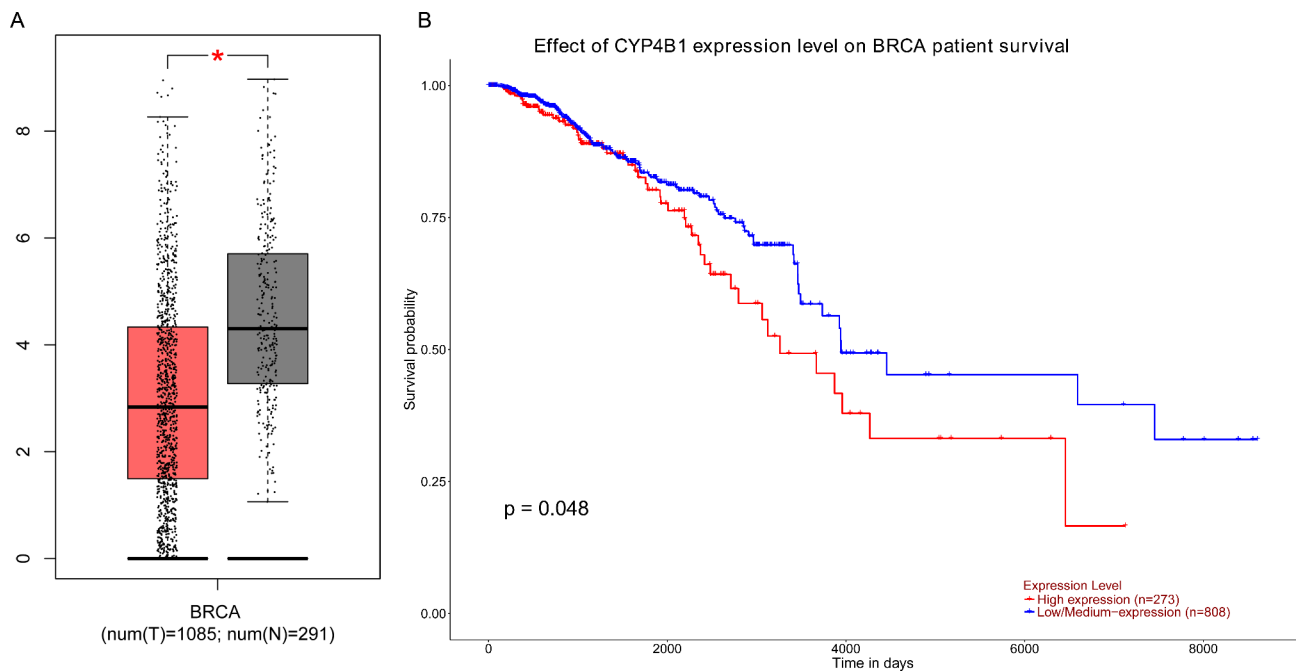
***CYP4B1* gene expression and its correlation with the prognosis of BC patients and SNP genotype *in silico*.**

**Table 1** Characteristics of patients with breast cancer and healthy individuals

Characteristics	Group	Case (N=571)	Control (N=572)	p
Age (years)	Mean ± SD	52.11 ± 10.14	51.88 ± 9.78	0.700 <sup>a</sup>
	> 52	270 (47.3%)	271 (47.4%)	
	≤ 52	301 (52.7%)	301 (52.6%)	
Smoking	Yes	292 (51.1%)	270 (47.2%)	0.193 <sup>b</sup>
	No	279 (48.9%)	302 (52.8%)	
Drinking	Yes	312 (54.6%)	293 (51.2%)	0.260 <sup>b</sup>
	No	259 (45.4%)	279 (48.8%)	
BMI (kg/m <sup>2</sup> )	> 24	95 (16.6%)	139 (24.3%)	
	≤ 24	202 (35.4%)	232 (40.6%)	
	Missing	274 (48.0%)	201 (35.1%)	
Tumor site	Left	210 (36.8%)		
	Right	198 (34.5%)		
Menopause	Yes	242 (42.4%)		
	No	135 (23.6%)		
	Missing	194 (34.0%)		
Reproductive number	= 1	211 (37.0%)		
	> 1	240 (42.0%)		
	Missing	120 (21.0%)		
ER	Positive	362 (63.4%)		
	Negative	161 (28.2%)		
	Missing	48 (8.4%)		
PR	Positive	313 (54.8%)		
	Negative	209 (36.6%)		
	Missing	49 (8.6%)		
Her-2	Positive	126 (22.1%)		
	Negative	194 (34.0%)		
	Missing	251 (44.0%)		
Ki67	> 25%	359 (62.9%)		
	≤ 25%	150 (26.3%)		
	Missing	62 (10.9%)		
LNM	Yes	194 (34.0%)		
	No	241 (42.2%)		
	Missing	136 (23.8%)		
Staging	III, IV	144 (25.2%)		
	I, II	346 (60.6%)		
	Missing	81 (14.2%)		

a: Student's t test (t test); b: Chi-square test ( $\chi^2$  test)

BMI: Body mass index; ER: Estrogen receptor; PR: Progesterone receptor; Her-2: Human epidermal growth factor receptor 2; LNM: Lymph node metastasis



**Fig. 1** The expression of *CYP4B1* mRNA in BC tissue (A) and its correlation with prognosis of BC patients (B). Data from GEPIA database (<http://gepia.cancer-pku.cn/index.html>) and Kaplan-Meier Plotter database (<https://kmplot.com/analysis/>). \* $P < 0.01$ . Red color means BC tissues and grey color means normal tissues

**Table 2** The basic information and HWE about the selected SNPs of *CYP4B1*

SNP ID	Chr: position	Alleles	Function	MAF		HWE (p value)	Haploreg4.1	RegulomeDB
				Cases	Controls			
rs2297813	1:46799052	G/T	5'-UTR	0.142	0.133	0.469	Promoter histone marks; Enhancer histone marks; DNase; Motifs changed; Selected eQTL hits	eQTL/caQTL + TF binding + any motif + Footprint + chromatin accessibility peak
rs12142787	1:46800918	G/A	intronic	0.255	0.234	0.295	Promoter histone marks; Enhancer histone marks; Motifs changed; GRASP QTL hits Selected eQTL hits	eQTL/caQTL + TF binding + any motif + Footprint + chromatin accessibility peak
rs3766197	1:46816155	C/T	intronic	0.167	0.166	0.762	Proteins bound; Motifs changed; GRASP QTL hits; Selected eQTL hits	eQTL/caQTL + TF binding / chromatin accessibility peak

$p < 0.05$ ;

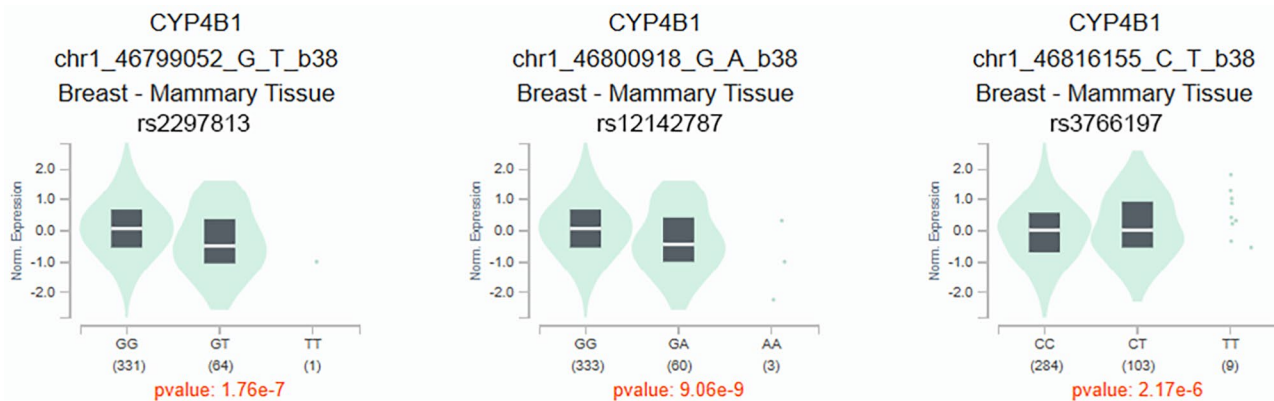
SNP: Single nucleotide polymorphism; Chr: Chromosome; MAF: Minor allele frequency; HWE: Hardy-Weinberg.

HaploReg v4.1 (<https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>), RegulomeDB (<https://regulome.stanford.edu/regulome-search/>)

The expression of the *CYP4B1* gene in normal and BC tissues was predicted by the GEPIA database indicating that the expression level of *CYP4B1* in BC tissues was significantly lower than that in normal tissues ( $p < 0.05$ , Fig. 1A). Moreover, the results of the Kaplan-Meier Plotter database analysis displayed that the higher *CYP4B1* expression, the worse survival in BC patients ( $p = 0.048$ , Fig. 1B).

In bioinformatics analysis, these SNPs might be associated with the regulation of promoter/ enhancer histone marks, motif changes, selected eQTL hits, GRASP QTL

hits, transcription factor (TF) binding, and chromatin accessibility peak (Table 2). Based on the GTEx Portal database, the genotypes of rs2297813 ( $p = 1.76e-7$ ), rs12142787 ( $p = 9.06e-9$ ), and rs3766197 ( $p = 2.17e-6$ ) were associated with the mRNA expression of *CYP4B1* in breast tissues (Fig. 2). Compared with RS2297813-GG and RS12142787-GG genotype, GT and TT genotypes of rs2297813 and GA and AA genotypes of rs12142787 may be associated with the lower expression of *CYP4B1* mRNA. In addition, the CT and TT genotypes of



**Fig. 2** The violin plot for the association between the genotypes of rs2297813, rs12142787, and rs3766197 and the mRNA expression of *CYP4B1* in breast tissues. Data from GTEx Portal database (<https://gtexportal.org/home/>)

rs3766197 may have higher mRNA expression of *CYP4B1* than CC genotype.

#### Information about candidate SNPs

Basic information about three candidate genetic loci of *CYP4B1* (rs2297813 G/T, rs12142787 G/A, and rs3766197 C/T) is demonstrated in Table 2. The MAFs of each candidate SNP were higher than 0.05 and in line with HWE ( $p > 0.05$ ).

#### Association between *CYP4B1* SNPs and BC risk

There was no significant correlation between *CYP4B1* SNPs and BC risk in the overall analysis (Table 3). Further, stratified analysis was conducted to explore the association between *CYP4B1* polymorphisms and BC risk in different subgroups based on age, smoking and drinking habits, BMI, and tumor location.

**BMI** (Table 4): *CYP4B1*-rs2297813 might be a risk factor for increasing the incidence of BC in participants with  $BMI \leq 24 \text{ kg/m}^2$  under the allele (OR=1.53, 95% CI 1.01–2.32,  $p=0.040$ ), dominant (OR=1.69, 95% CI 1.06–2.69,  $p=0.027$ ), over-dominant (OR=1.72, 95% CI 1.07–2.77,  $p=0.026$ ), and log-additive (OR=1.56, 95% CI 1.02–2.39,  $p=0.041$ ) models.

**Smoking and drinking** (Table 5): When stratified by smoking, rs12142787 was associated with AN increased incidence of BC in smokers under the allele model (OR=1.32, 95% CI 1.01–1.74,  $p=0.045$ ). Among non-drinkers rs2297813 (allele: OR=1.48, 95% CI 1.04–2.10,  $p=0.029$ ; heterozygote: OR=1.75, 95% CI 1.16–2.64,  $p=0.026$ ; dominant: OR=1.69, 95% CI 1.14–2.52,  $p=0.009$ ; over-dominant: OR=1.75, 95% CI 1.16–2.63,  $p=0.007$ ; log-additive: OR=1.53, 95% CI 1.06–2.19,  $p=0.021$ ) and rs12142787 (allele: OR=1.36, 95% CI 1.03–1.79,  $p=0.029$ ; dominant: OR=1.51, 95% CI 1.07–2.13,  $p=0.020$ ; over-dominant: OR=1.47, 95% CI 1.03–2.09,  $p=0.034$ ; log-additive: OR=1.34, 95% CI 1.02–1.78,

$p=0.037$ ) were studied to be related to an increased incidence of BC.

**Tumor stage** (Table 6): *CYP4B1* rs3766197 was associated with the higher risk of advanced stages (III/IV stage) of BC under the heterozygote (OR=1.57, 95% CI 1.04–2.38,  $p=0.035$ ) and over-dominant (OR=1.61, 95% CI 1.05–2.46,  $p=0.031$ ) models.

**Reproductive number** (Table 7): The contribution of *CYP4B1* rs2297813 (allele: OR=1.55, 95% CI 1.07–2.26,  $p=0.021$ ; dominant: OR=1.61, 95% CI 1.05–2.47,  $p=0.029$ ; over-dominant: OR=1.55, 95% CI 1.00–2.39,  $p=0.049$ ; log-additive: OR=1.58, 95% CI 1.06–2.35,  $p=0.024$ ) and rs12142787 (allele: OR=1.38, 95% CI 1.02–1.87,  $p=0.037$ ; dominant: OR=1.53, 95% CI 1.03–2.25,  $p=0.033$ ; log-additive: OR=1.38, 95% CI 1.01–1.89,  $p=0.044$ ) to BC risk might be associated with more than one birth in patients with BC.

**Other factors** After analyses stratified based on age and tumor location, no significant association was found (Table S1,  $p > 0.05$ ). Patients were grouped according to menstrual status, tumor biomarkers, and LNM to analyze the relationship between these SNPs and the risk of BC, however, there was no significant correlation between them (Table S2,  $p > 0.05$ ).

#### *CYP4B1* SNP-SNP interaction in BC risk

MDR was used to analyze and evaluate SNP-SNP interaction (Fig. 3). The blue line indicates that rs12142787 and rs3766197 have a redundant effect on BC risk. As shown in Table 8, the three-locus model (a combination of rs2297813, rs12142787, and rs3766197) had the highest cross-validation consistency (CVC, 10/10) and testing balanced accuracy (Testing Bal. Acc., 0.507), making it the best predictive model for BC risk.



**Table 3** Analysis of the association between susceptibility of BC and *CYP4B1* SNPs (Overall analysis)

SNP ID	Model	Genotype	Case	Control	Adjusted by age, smoke and drink	
					OR (95% CI)	p
rs2297813	Allele	G	980 (85.8%)	992 (86.7%)	1.00	0.530
		T	162 (14.2%)	152 (13.3%)	1.08 (0.85–1.37)	
	Codominant	G/G	418 (73.2%)	432 (75.5%)	1.00	0.430
		G/T	144 (25.2%)	128 (22.4%)	1.17 (0.89–1.54)	
		T/T	9 (1.6%)	12 (2.1%)	0.77 (0.32–1.86)	
		G/G	418 (73.2%)	432 (75.5%)	1.00	0.360
	Recessive	G/T-T/T	153 (26.8%)	140 (24.5%)	1.13 (0.87–1.48)	
		G/G-G/T	562 (98.4%)	560 (97.9%)	1.00	0.510
	Over-dominant	T/T	9 (1.6%)	12 (2.1%)	0.74 (0.31–1.78)	
		G/G-T/T	427 (74.8%)	444 (77.6%)	1.00	0.250
Log-additive	G/T	144 (25.2%)	128 (22.4%)	1.17 (0.89–1.54)		
				1.08 (0.85–1.37)	0.520	
rs12142787	Allele	G	846 (74.5%)	874 (76.5%)	1.00	0.250
		A	290 (25.5%)	268 (23.5%)	1.12 (0.92–1.35)	
	Codominant	G/G	316 (55.6%)	339 (59.4%)	1.00	0.410
		A/G	214 (37.7%)	196 (34.3%)	1.18 (0.92–1.51)	
		A/A	38 (6.7%)	36 (6.3%)	1.14 (0.70–1.85)	
	Dominant	G/G	316 (55.6%)	339 (59.4%)	1.00	0.190
		A/G-A/A	252 (44.4%)	232 (40.6%)	1.17 (0.93–1.48)	
	Recessive	G/G-A/G	530 (93.3%)	535 (93.7%)	1.00	0.780
		A/A	38 (6.7%)	36 (6.3%)	1.07 (0.67–1.72)	
	Over-dominant	G/G-A/A	354 (62.3%)	375 (65.7%)	1.00	0.220
A/G		214 (37.7%)	196 (34.3%)	1.16 (0.91–1.48)		
Log-additive				1.12 (0.93–1.35)	0.240	
rs3766197	Allele	C	951 (83.3%)	951 (83.4%)	1.00	0.930
		T	191 (16.7%)	189 (16.6%)	1.01 (0.81–1.26)	
	Codominant	C/C	399 (69.9%)	395 (69.3%)	1.00	0.580
		T/C	153 (26.8%)	161 (28.2%)	0.94 (0.72–1.22)	
		T/T	19 (3.3%)	14 (2.5%)	1.38 (0.68–2.79)	
		C/C	399 (69.9%)	395 (69.3%)	1.00	0.840
	Recessive	T/C-T/T	172 (30.1%)	175 (30.7%)	0.98 (0.76–1.26)	
		C/C-T/C	552 (96.7%)	556 (97.5%)	1.00	0.340
	Over-dominant	T/T	19 (3.3%)	14 (2.5%)	1.40 (0.69–2.83)	
		C/C-T/T	418 (73.2%)	409 (71.8%)	1.00	0.580
Log-additive	T/C	153 (26.8%)	161 (28.2%)	0.93 (0.72–1.21)		
				1.01 (0.81–1.26)	0.900	

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

### Discussion

BC is a heterogeneous disease, and BC patients with different demographic characteristics and tumor status may exhibit different clinical manifestations [7]. To facilitate diagnosis, prognosis, and treatment, patients are classified according to their clinical data, tumor morphology, genomics, proteomics, and other characteristics [8]. The World Health Organization has identified 21 histological types of BC [26]. Although the mortality of BC shows a trend of decline in recent years, it remains high in all female malignancies [4, 27]. Identifying reliable susceptibility genes associated with BC and investigating the

correlation of genetic variants with BC risk is crucial for the effective prevention and treatment of BC [28].

Cytochrome P450 4B1 (*CYP4B1*) exerts an influence on exogenous biotransformation, and its major extrahepatic expression is associated with some tissue-specific toxicity [16]. *CYP4B1* can hydroxylate common endobiotic substrates such as fatty acids, as well as activate xenobiotics such as 4-ipomeanol (4-IPO), perilla ketone (PK) or valproic acid (VPA) [29]. The contribution of *CYP4B1* in cancer is of particular interest as *CYP4B1* gene expression has been found to be altered and presumably associated with certain cancers [30]. In bioinformatics analysis, *CYP4B1* expression in BC tissues was lower than that in

**Table 4** The SNPs of *CYP4B1* associated with susceptibility of BC in participants (BMI)

SNP ID	Model	Genotype	Case	Control	OR (95% CI)	p	Case	Control	OR (95% CI)	p
BMI			> 24 kg/m <sup>2</sup>				≤ 24 kg/m <sup>2</sup>			
rs2297813	Allele	G	74 (78.7%)	98 (72.1%)	1.00	0.440	347 (85.9%)	419 (90.3%)	1.00	<b>0.040*</b>
		T	20 (21.3%)	38 (27.9%)	0.73 (0.39–1.37)		57 (14.1%)	45 (9.7%)	1.53 (1.01–2.32)	
	Codominant	G/G	1 (1.1%)	3 (2.2%)	0.37 (0.04–3.90)	0.250	148 (73.3%)	190 (81.9%)	1.00	0.080
		G/T	74 (77.9%)	98 (70.5%)	1.00		51 (25.2%)	39 (16.8%)	1.72 (1.07–2.78)	
	Dominant	T/T	21 (22.1%)	41 (29.5%)	0.70 (0.38–1.30)		3 (1.5%)	3 (1.3%)	1.24 (0.24–6.38)	
		G/G	94 (99.0%)	136 (97.8%)	1.00	0.430	148 (73.3%)	190 (81.9%)	1.00	<b>0.027*</b>
	Recessive	G/T-T/T	1 (1.1%)	3 (2.2%)	0.41 (0.04–4.23)		54 (26.7%)	42 (18.1%)	1.69 (1.06–2.69)	
		G/G-G/T	75 (79.0%)	101 (72.7%)	1.00	0.350	199 (98.5%)	229 (98.7%)	1.00	0.910
	Over-dominant	T/T	20 (21.1%)	38 (27.3%)	0.74 (0.40–1.39)		3 (1.5%)	3 (1.3%)	1.10 (0.21–5.66)	
		G/G-T/T			0.70 (0.40–1.23)	0.210	151 (74.8%)	193 (83.2%)	1.00	<b>0.026*</b>
Log-additive	G/T	74 (77.9%)	98 (70.5%)	1.00		51 (25.2%)	39 (16.8%)	1.72 (1.07–2.77)		
		20 (21.1%)	38 (27.3%)	0.73 (0.39–1.37)	0.440			1.56 (1.02–2.39)	<b>0.041*</b>	
rs12142787	Allele	G	149 (78.4%)	215 (77.3%)	1.00	0.782	301 (74.9%)	365 (79.0%)	1.00	0.150
		A	41 (21.6%)	63 (22.7%)	0.94 (0.60–1.47)		101 (25.1%)	97 (21.0%)	1.26 (0.92–1.74)	
	Codominant	G/G	59 (62.1%)	85 (61.1%)	1.00	0.910	111 (55.2%)	145 (62.8%)	1.00	0.160
		A/G	31 (32.6%)	45 (32.4%)	1.04 (0.59–1.85)		79 (39.3%)	75 (32.5%)	1.49 (0.98–2.24)	
	Dominant	A/A	5 (5.3%)	9 (6.5%)	0.79 (0.24–2.56)		11 (5.5%)	11 (4.8%)	1.34 (0.55–3.25)	
		G/G	59 (62.1%)	85 (61.1%)	1.00	1.000	111 (55.2%)	145 (62.8%)	1.00	0.056
	Recessive	A/G-A/A	36 (37.9%)	54 (38.9%)	1.00 (0.58–1.72)		90 (44.8%)	86 (37.2%)	1.47 (0.99–2.18)	
		G/G-A/G	90 (94.7%)	130 (93.5%)	1.00	0.670	190 (94.5%)	220 (95.2%)	1.00	0.740
	Over-dominant	A/A	5 (5.3%)	9 (6.5%)	0.78 (0.25–2.48)		11 (5.5%)	11 (4.8%)	1.16 (0.48–2.77)	
		G/G-A/A	64 (67.4%)	94 (67.6%)	1.00	0.840	122 (60.7%)	156 (67.5%)	1.00	0.070
Log-additive	A/G	31 (32.6%)	45 (32.4%)	1.06 (0.60–1.87)		79 (39.3%)	75 (32.5%)	1.45 (0.97–2.18)		
				0.96 (0.62–1.50)	0.870			1.33 (0.96–1.84)	0.089	
rs3766197	Allele	C	163 (85.8%)	226 (81.3%)	1.00	0.202	332 (82.2%)	386 (83.2%)	1.00	0.690
		T	27 (14.2%)	52 (18.7%)	0.72 (0.43–1.20)		72 (17.8%)	78 (16.8%)	1.07 (0.75–1.53)	
	Codominant	C/C	68 (71.6%)	91 (65.5%)	1.00	0.150	137 (67.8%)	162 (69.8%)	1.00	0.980
		T/C	27 (28.4%)	44 (31.6%)	0.83 (0.46–1.49)		58 (28.7%)	62 (26.7%)	1.05 (0.68–1.62)	
	Dominant	T/T	0 (0%)	4 (2.9%)	0.00 (0.00-NA)		7 (3.5%)	8 (3.5%)	0.99 (0.34–2.83)	
		C/C	68 (71.6%)	91 (65.5%)	1.00	0.370	137 (67.8%)	162 (69.8%)	1.00	0.850
	Recessive	T/C-T/T	27 (28.4%)	48 (34.5%)	0.77 (0.43–1.37)		65 (32.2%)	70 (30.2%)	1.04 (0.69–1.58)	
		C/C-T/C	95 (100%)	135 (97.1%)	1.00	0.065	195 (96.5%)	224 (96.5%)	1.00	0.960
	Over-dominant	T/T	0 (0%)	4 (2.9%)	0.00 (0.00-NA)		7 (3.5%)	8 (3.5%)	0.97 (0.34–2.77)	
		C/C-T/T	68 (71.6%)	95 (68.3%)	1.00	0.610	144 (71.3%)	170 (73.3%)	1.00	0.830
Log-additive	T/C	153 (26.8%)	44 (31.6%)	0.93 (0.72–1.21)		58 (28.7%)	62 (26.7%)	1.05 (0.68–1.61)		
				1.01 (0.81–1.26)	0.900			1.03 (0.72–1.46)	0.880	

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

p < 0.05: Bold text and “\*” represent statistical significance

normal tissues, and the higher *CYP4B1* expression was correlated with worse survival. Of note, the relationship between *CYP4B1* variants and BC risk has not been reported. Our results displayed that *CYP4B1* rs2297813 was linked to an increased risk of BC in women with BMI ≤ 24 kg/m<sup>2</sup>, and non-drinkers. And *CYP4B1* rs12142787 had the same effect on BC risk in smokers, and non-drinkers. Moreover, *CYP4B1* rs3766197 was related to the tumor stage, and the association of rs2297813 and rs12142787 with BC risk was related to reproductive number. Bioinformatics analysis showed that these SNPs might be associated with the regulation of promoter/ enhancer histone marks, motif changes,

selected eQTL hits, GRASP QTL hits, transcription factor (TF) binding, and chromatin accessibility peak. Based on the GTEx Portal database, the genotypes rs2297813, rs12142787, and rs3766197 were associated with the mRNA expression of *CYP4B1* in breast tissues. These results might suggest that these SNPs may be involved in BC etiology by regulating the expression or function of *CYP4B1*, which provides a theoretical basis for subsequent mechanistic studies.

Because of the clear correlation between increased BMI and increased incidence of BC, obese women have a higher risk of dying from BC than non-obese women [31, 32]. Therefore, the risk of developing BC is strongly





**Table 5** (continued)

SNP ID	Model	Genotype	Case	Control	OR	p	Case	Control	OR	p
rs12142787	Allele	G	479 (77.3%)	446 (76.1%)	1.00	0.637	367 (71.1%)	428 (77.0%)	1.00	<b>0.029*</b>
		A	141 (22.7%)	140 (23.9%)	0.94 (0.72–1.23)		149 (28.9%)	128 (23.0%)	1.36 (1.03–1.79)	
	Codominant	G/G	188 (60.6%)	172 (58.7%)	1.00	0.880	128 (49.6%)	167 (60.1%)	1.00	0.064
		A/G	103 (33.2%)	102 (34.8%)	0.93 (0.66–1.31)		111 (43.0%)	94 (33.8%)	1.53 (1.06–2.19)	
		A/A	19 (6.1%)	19 (6.5%)	0.89 (0.46–1.74)		19 (7.4%)	17 (6.1%)	1.42 (0.71–2.87)	
	Dominant	G/G	188 (60.6%)	172 (58.7%)	1.00	0.620	128 (49.6%)	167 (60.1%)	1.00	<b>0.020*</b>
		A/G-A/A	122 (39.4%)	121 (41.3%)	0.92 (0.66–1.27)		130 (50.4%)	111 (39.9%)	1.51 (1.07–2.13)	
	Recessive	G/G-A/G	291 (93.9%)	274 (93.5%)	1.00	0.790	239 (92.6%)	261 (93.9%)	1.00	0.61
		A/A	19 (6.1%)	19 (6.5%)	0.92 (0.47–1.77)		19 (7.4%)	17 (6.1%)	1.20 (0.60–2.37)	
	Over-dominant	G/G-A/A	207 (66.8%)	191 (65.2%)	1.00	0.700	147 (57.0%)	184 (66.2%)	1.00	<b>0.034*</b>
A/G		103 (33.2%)	102 (34.8%)	0.94 (0.67–1.31)		111 (43.0%)	94 (33.8%)	1.47 (1.03–2.09)		
	Log-additive				0.93 (0.72–1.21)	0.610			1.34 (1.02–1.78)	<b>0.037*</b>
rs3766197	Allele	C	515 (82.5%)	485 (83.3%)	1.00	0.712	436 (84.2%)	466 (83.5%)	1.00	0.770
		T	109 (17.5%)	97 (16.7%)	1.06 (0.78–1.43)		82 (15.8%)	92 (16.5%)	0.95 (0.69–1.32)	
	Codominant	C/C	214 (68.6%)	200 (68.7%)	1.00	0.600	185 (71.4%)	195 (69.9%)	1.00	0.920
		T/C	87 (27.9%)	85 (29.2%)	0.95 (0.67–1.36)		66 (25.5%)	76 (27.2%)	0.93 (0.63–1.38)	
		T/T	11 (3.5%)	6 (2.1%)	1.62 (0.58–4.47)		8 (3.1%)	8 (2.9%)	1.10 (0.40–3.02)	
	Dominant	C/C	214 (68.6%)	200 (68.7%)	1.00	0.980	185 (71.4%)	195 (69.9%)	1.00	0.780
		T/C-T/T	98 (31.4%)	91 (31.3%)	1.00 (0.70–1.41)		74 (28.6%)	84 (30.1%)	0.95 (0.65–1.38)	
	Recessive	C/C-T/C	301 (96.5%)	285 (97.9%)	1.00	0.330	251 (96.9%)	271 (97.1%)	1.00	0.830
		T/T	11 (3.5%)	6 (2.1%)	1.64 (0.60–4.51)		8 (3.1%)	8 (2.9%)	1.12 (0.41–3.07)	
	Over-dominant	C/C-T/T	225 (72.1%)	206 (70.8%)	1.00	0.700	193 (74.5%)	203 (72.8%)	1.00	0.710
T/C		87 (27.9%)	85 (29.2%)	0.93 (0.65–1.33)		66 (25.5%)	76 (27.2%)	0.93 (0.63–1.37)		
	Log-additive				1.04 (0.77–1.41)	0.780			0.97 (0.70–1.35)	0.870

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

p < 0.05: Bold text and “\*” represent statistical significance

affected by BMI. Wang et al. have reported that the SNPs of *NBS1* (rs1805812, rs2735385, and rs6999227), *BRIP1* (rs7220719), and *PTEN* (rs2299941) are associated with BC risk in subjects with BMI < 25 kg/m<sup>2</sup> [27]. Additionally, *TP53* rs12951053 and *BRIP1* rs16945628 are related to BC risk in people with BMI ≥ 25 kg/m<sup>2</sup> [27]. Our data revealed the association between *CYP4B1* SNPs and BC risk in patients with BMI ≤ 24 kg/m<sup>2</sup>.

It has been reported that compounds such as polycyclic hydrocarbons contained in tobacco smoke can induce BC [33]. Jones ME et al. have shown that smoking significantly increases the risk of BC in women [34]. In a study by Terry PD et al., the relationship between smoking and BC risk is affected by certain health index of smokers and tumor types, such as the BMI of smokers, and BC patients with ER+/- or PR+/- [33]. Studies over the past few decades have shown positive, negative, or no association between smoking and BC risk [33–35]. Hsieh et al. have reported that rs73229797 can increase the activity of the *CHRNA9* gene promoter and affect the risk of BC in smoking exposure (both smoking and passive smoking) [36]. Furthermore, Naif et al. have also reported that the C and G alleles of *CYP1A* SNPs are associated with the risk of BC in Iraq smokers [37]. Our research is the first to indicate that *CYP4B1* rs12142787 is a risk factor for BC in Chinese female smokers in the allele model.

Studies have revealed that smoking status is associated with BC risk [38, 39]. However, the association between rs12142787 and BC risk needs to be further confirmed in subgroup studies based on participants’ smoking status (active or passive).

Reproductive number and tumor stage are closely associated with the progression and metastasis of tumor cells [40, 41]. A previous study has shown that rs2735385 and rs6999227 in *NBS1* are significantly associated with BC in stage III [27]. LincRNA *H19* rs2071095 is associated with BC in stage I [42]. Our results showed that *CYP4B1* rs3766197 was also associated with BC risk according to stratified analysis by tumor stage. *CYP4B1* rs2297813 and rs12142787 were risk factors in the stratified analysis by reproductive number. Yan et al. have reported a remarkable interaction between lincRNA *HOTAIR* and reproductive factors influenced by various factors, such as ER status, menopause, and family history [40, 43]. Our data revealed that *CYP4B1* SNPs were correlated with BC risk after analysis stratified by reproductive numbers.

However, there still are limitations. Firstly, the molecular mechanisms of *CYP4B1* in the occurrence of BC has not been analyzed. In subsequent studies, it is necessary to understand the exact molecular pathway of *CYP4B1* in cells and to investigate the effect of SNPs on *CYP4B1* gene expression. Secondly, the role of these

**Table 6** The SNPs of *CYP4B1* associated with susceptibility of breast cancer in patients (Tumor staging)

SNP ID	Model	Genotype	Case III, IV	Control I, II	OR (95% CI)	p
rs2297813	Allele	G	238 (82.6%)	600 (86.7%)	1.00	0.099
		T	50 (17.4%)	92 (13.3%)	1.37 (0.94-2.00)	
	Codominant	G/G	108 (75.0%)	250 (72.2%)	1.00	0.130
		G/T	36 (25.0%)	90 (26.0%)	0.92 (0.59-1.45)	
		T/T	0 (0%)	6 (1.7%)		
	Dominant	G/G	108 (75.0%)	250 (72.2%)	1.00	0.520
		G/T-T/T	36 (25.0%)	96 (27.8%)	0.87 (0.55-1.36)	
	Recessive	G/G-G/T	144 (100%)	340 (98.3%)	1.00	0.046
		T/T	0 (0%)	6 (1.7%)	0.00 (0.00-NA)	
	Over-dominant	G/G-T/T	108 (75.0%)	256 (74.0%)	1.00	0.800
G/T		36 (25.0%)	90 (26.0%)	0.94 (0.60-1.48)		
rs12142787	Allele	G	210 (73.4%)	514 (74.7%)	1.00	0.676
		A	76 (26.6%)	174 (25.3%)	1.07 (0.78-1.46)	
	Codominant	G/G	85 (59.0%)	185 (53.9%)	1.00	0.340
		A/G	53 (36.8%)	133 (38.8%)	0.87 (0.58-1.33)	
		A/A	6 (4.2%)	25 (7.3%)	0.53 (0.21-1.34)	
	Dominant	G/G	85 (59.0%)	185 (53.9%)	1.00	0.330
		A/G-A/A	59 (41.0%)	158 (46.1%)	0.82 (0.55-1.22)	
	Recessive	G/G-A/G	138 (95.8%)	318 (92.7%)	1.00	0.190
		A/A	6 (4.2%)	25 (7.3%)	0.56 (0.22-1.39)	
	Over-dominant	G/G-A/A	91 (63.2%)	210 (61.2%)	1.00	0.720
A/G		53 (36.8%)	133 (38.8%)	0.93 (0.62-1.40)		
rs3766197	Allele	C	238 (82.6%)	576 (83.2%)	1.00	0.820
		T	50 (17.4%)	116 (16.8%)	1.04 (0.72-1.50)	
	Codominant	C/C	90 (62.5%)	249 (72.0%)	1.00	0.093
		T/C	49 (34.0%)	85 (24.6%)	1.62 (1.05-2.50)	
		T/T	5 (3.5%)	12 (3.5%)	1.20 (0.41-3.52)	
	Dominant	C/C	90 (62.5%)	249 (72.0%)	1.00	<b>0.035*</b>
		T/C-T/T	54 (37.5%)	97 (28.0%)	1.57 (1.04-2.38)	
	Recessive	C/C-T/C	139 (96.5%)	334 (96.5%)	1.00	0.960
		T/T	5 (3.5%)	12 (3.5%)	1.03 (0.35-2.99)	
	Over-dominant	C/C-T/T	95 (66.0%)	261 (75.4%)	1.00	<b>0.031*</b>
T/C		49 (34.0%)	85 (24.6%)	1.61 (1.05-2.46)		
	Allele	C	238 (82.6%)	576 (83.2%)	1.00	0.820
		T	50 (17.4%)	116 (16.8%)	1.04 (0.72-1.50)	
	Codominant	C/C	90 (62.5%)	249 (72.0%)	1.00	0.093
		T/C	49 (34.0%)	85 (24.6%)	1.62 (1.05-2.50)	
		T/T	5 (3.5%)	12 (3.5%)	1.20 (0.41-3.52)	
	Dominant	C/C	90 (62.5%)	249 (72.0%)	1.00	<b>0.035*</b>
		T/C-T/T	54 (37.5%)	97 (28.0%)	1.57 (1.04-2.38)	
	Recessive	C/C-T/C	139 (96.5%)	334 (96.5%)	1.00	0.960
		T/T	5 (3.5%)	12 (3.5%)	1.03 (0.35-2.99)	
	Over-dominant	C/C-T/T	95 (66.0%)	261 (75.4%)	1.00	<b>0.031*</b>
T/C		49 (34.0%)	85 (24.6%)	1.61 (1.05-2.46)		
	Allele	C	238 (82.6%)	576 (83.2%)	1.00	0.820
		T	50 (17.4%)	116 (16.8%)	1.04 (0.72-1.50)	
	Codominant	C/C	90 (62.5%)	249 (72.0%)	1.00	0.093
		T/C	49 (34.0%)	85 (24.6%)	1.62 (1.05-2.50)	
		T/T	5 (3.5%)	12 (3.5%)	1.20 (0.41-3.52)	
	Dominant	C/C	90 (62.5%)	249 (72.0%)	1.00	<b>0.035*</b>
		T/C-T/T	54 (37.5%)	97 (28.0%)	1.57 (1.04-2.38)	
	Recessive	C/C-T/C	139 (96.5%)	334 (96.5%)	1.00	0.960
		T/T	5 (3.5%)	12 (3.5%)	1.03 (0.35-2.99)	
	Over-dominant	C/C-T/T	95 (66.0%)	261 (75.4%)	1.00	<b>0.031*</b>
T/C		49 (34.0%)	85 (24.6%)	1.61 (1.05-2.46)		
	Allele	C	238 (82.6%)	576 (83.2%)	1.00	0.820
		T	50 (17.4%)	116 (16.8%)	1.04 (0.72-1.50)	
	Codominant	C/C	90 (62.5%)	249 (72.0%)	1.00	0.093
		T/C	49 (34.0%)	85 (24.6%)	1.62 (1.05-2.50)	
		T/T	5 (3.5%)	12 (3.5%)	1.20 (0.41-3.52)	
	Dominant	C/C	90 (62.5%)	249 (72.0%)	1.00	<b>0.035*</b>
		T/C-T/T	54 (37.5%)	97 (28.0%)	1.57 (1.04-2.38)	
	Recessive	C/C-T/C	139 (96.5%)	334 (96.5%)	1.00	0.960
		T/T	5 (3.5%)	12 (3.5%)	1.03 (0.35-2.99)	
	Over-dominant	C/C-T/T	95 (66.0%)	261 (75.4%)	1.00	<b>0.031*</b>
T/C		49 (34.0%)	85 (24.6%)	1.61 (1.05-2.46)		
	Allele	C	238 (82.6%)	576 (83.2%)	1.00	0.820
		T	50 (17.4%)	116 (16.8%)	1.04 (0.72-1.50)	
	Codominant	C/C	90 (62.5%)	249 (72.0%)	1.00	0.093
		T/C	49 (34.0%)	85 (24.6%)	1.62 (1.05-2.50)	
		T/T	5 (3.5%)	12 (3.5%)	1.20 (0.41-3.52)	
	Dominant	C/C	90 (62.5%)	249 (72.0%)	1.00	<b>0.035*</b>
		T/C-T/T	54 (37.5%)	97 (28.0%)	1.57 (1.04-2.38)	
	Recessive	C/C-T/C	139 (96.5%)	334 (96.5%)	1.00	0.960
		T/T	5 (3.5%)	12 (3.5%)	1.03 (0.35-2.99)	
	Over-dominant	C/C-T/T	95 (66.0%)	261 (75.4%)	1.00	<b>0.031*</b>
T/C		49 (34.0%)	85 (24.6%)	1.61 (1.05-2.46)		

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

p<0.05: Bold text and “\*” represent statistical significance

polymorphisms in treatment response and relapse has not been explored, and we will continue to collect samples and refine the experimental design to further explore the role of these polymorphisms in BC treatment response and recurrence. Thirdly, missing data in this study led to errors in some experimental results. Our findings need to be further confirmed in future studies with high-quality and large samples. Nevertheless, to our knowledge, our research remains the first to demonstrate that *CYP4B1* polymorphisms are related to BC risk in different subgroups of Chinese women.

### Conclusion

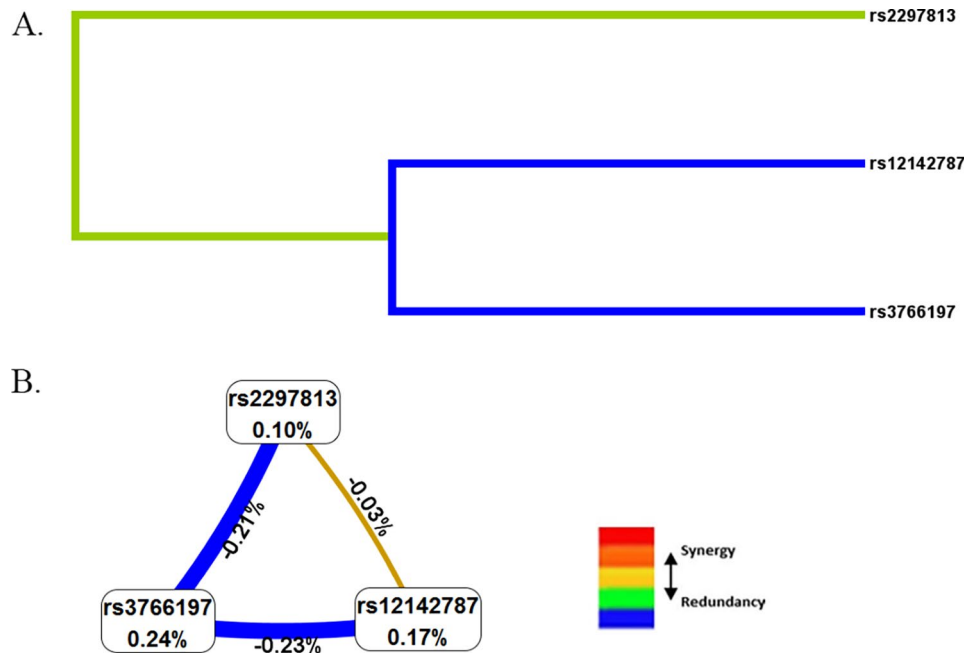
In summary, *CYP4B1* SNPs could be risk factors for BC in Chinese women, especially in patients with BMI≤24 kg/m<sup>2</sup>, smokers, non-drinkers, patients in advanced stages (III/IV stage), and patients who reproduced once. These findings shed light on the relationship between *CYP4B1* SNPs and BC risk in Chinese women and provide a theoretical and molecular foundation for the prevention of BC in the future.

**Table 7** The SNPs of *CYP4B1* associated with susceptibility of breast cancer in patients (Reproductive number)

SNP ID	Model	Genotype	Case = 1	Control > 1	OR (95% CI)	p
rs2297813	Allele	G	349 (82.7%)	423 (88.1%)	1.00	<b>0.021*</b>
		T	73 (17.3%)	57 (11.9%)	1.55 (1.07–2.26)	
	Codominant	G/G	142 (67.3%)	185 (77.1%)	1.00	0.080
		G/T	65 (30.8%)	53 (22.1%)	1.57 (1.02–2.43)	
		T/T	4 (1.9%)	2 (0.8%)	2.55 (0.44–14.66)	
		G/G	142 (67.3%)	185 (77.1%)	1.00	<b>0.029*</b>
	Dominant	G/T-T/T	69 (32.7%)	55 (22.9%)	1.61 (1.05–2.47)	
		G/G	142 (67.3%)	185 (77.1%)	1.00	<b>0.029*</b>
	Recessive	G/G-G/T	207 (98.1%)	238 (99.2%)	1.00	0.340
		T/T	4 (1.9%)	2 (0.8%)	2.28 (0.40–13.03)	
Over-dominant	G/G-T/T	146 (69.2%)	187 (77.9%)	1.00	<b>0.049*</b>	
	G/T	65 (30.8%)	53 (22.1%)	1.55 (1.00–2.39)		
Log-additive					1.58 (1.06–2.35)	<b>0.024*</b>
						<b>0.037*</b>
rs12142787	Allele	G	300 (71.8%)	372 (77.8%)	1.00	<b>0.037*</b>
		A	118 (28.2%)	106 (22.2%)	1.38 (1.02–1.87)	
	Codominant	G/G	106 (50.7%)	147 (61.5%)	1.00	0.100
		A/G	88 (42.1%)	78 (32.6%)	1.51 (1.01–2.28)	
		A/A	15 (7.2%)	14 (5.9%)	1.59 (0.72–3.53)	
	Dominant	G/G	106 (50.7%)	147 (61.5%)	1.00	<b>0.033*</b>
		A/G-A/A	103 (49.3%)	92 (38.5%)	1.53 (1.03–2.25)	
	Recessive	G/G-A/G	194 (92.8%)	225 (94.1%)	1.00	0.450
		A/A	15 (7.2%)	14 (5.9%)	1.35 (0.62–2.94)	
	Over-dominant	G/G-A/A	121 (57.9%)	161 (67.4%)	1.00	0.071
A/G		88 (42.1%)	78 (32.6%)	1.44 (0.97–2.15)		
Log-additive					1.38 (1.01–1.89)	<b>0.044*</b>
						0.362
rs3766197	Allele	C	356 (84.4%)	394 (82.1%)	1.00	0.362
		T	66 (15.6%)	86 (17.9%)	0.85 (0.60–1.21)	
	Codominant	C/C	150 (71.1%)	165 (68.8%)	1.00	0.220
		T/C	56 (26.5%)	64 (26.7%)	0.90 (0.58–1.38)	
		T/T	5 (2.4%)	11 (4.6%)	0.39 (0.13–1.19)	
		C/C	150 (71.1%)	165 (68.8%)	1.00	0.350
	Dominant	T/C-T/T	61 (28.9%)	75 (31.2%)	0.82 (0.54–1.24)	
		C/C	150 (71.1%)	165 (68.8%)	1.00	0.350
	Recessive	C/C-T/C	206 (97.6%)	229 (95.4%)	1.00	0.096
		T/T	5 (2.4%)	11 (4.6%)	0.41 (0.13–1.22)	
Over-dominant	C/C-T/T	155 (73.5%)	176 (73.3%)	1.00	0.770	
	T/C	56 (26.5%)	64 (26.7%)	0.94 (0.61–1.44)		
Log-additive					0.78 (0.55–1.12)	0.170

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

p < 0.05: Bold text and \*\* represent statistical significance



**Fig. 3** Analysis of SNP-SNP interactions (A. Dendrogram; B. Fruchterman-Reingold. Green and blue represent redundancy or correlation. Values in nodes represent the information gains of individual attribute (main effects). Values between nodes are information gains of each pair of attributes (interaction effects)

**Table 8** SNP - SNP interaction models analyzed by the MDR method

Model	Bal.Acc.CV Training	Bal.Acc.CV Testing	OR (95% CI)	p value	CV Consistency
rs12142787	0.5201	0.5117	1.18 (0.93–1.49)	0.169	8/10
rs2297813, rs12142787	0.5289	0.5048	1.27 (1.00–1.61)	<b>0.048*</b>	8/10
rs2297813, rs12142787, rs3766197	0.5362	0.5073	1.34 (1.05–1.69)	<b>0.016*</b>	10/10

MDR: Multifactor dimensionality reduction; Bal. Acc.: Balanced accuracy; CVC: Cross-validation consistency; OR: Odds ratio; 95% CI: 95% confidence interval  $p < 0.05$ : Bold text and \* represent statistical significance

**List of abbreviations**

- BC Breast cancer
- CYP Cytochrome P450
- RT-PCR reverse transcription-polymerase chain reaction
- SNPs single nucleotide polymorphisms
- BI-RADS Breast Imaging Reporting and Data System
- IHC Immune-histochemical
- ER estrogen receptor
- PR progesterone receptor
- HER2 human epidermal growth factor receptor-2
- Ki67 cell proliferation marker
- ER<sup>+</sup> ER-positive
- ER<sup>-</sup> ER-negative
- PR<sup>+</sup> PR-positive
- PR<sup>-</sup> PR-negative
- HER2<sup>+</sup> HER2-positive
- HER2<sup>-</sup> HER2-negative
- FISH fluorescence in situ hybridization

- Ki67<sup>+</sup> Ki67-positive
- Ki67<sup>-</sup> Ki67-negative
- LNM lymph node metastasis
- AJCC American Joint Committee on Cancer
- ORs odds ratios
- Cis confidence intervals
- MDR multi-factor dimensionality reduction

**Supplementary Information**

The online version contains supplementary material available at <https://doi.org/10.1186/s12885-023-11477-y>.

Supplementary Material 1

**Acknowledgements**

Not applicable.

**Authors' contributions**

Xinhan Zhao conceived and designed the study. Yanhai Yin and Liangqian Tong analyzed data and wrote the manuscript. Zhenling Wan and Yanfang Sui collected data and made charts. Fen Li and Qian Huang conceived of the study, participated in the design and coordination of the study. Xinhan Zhao revised the manuscript. All authors approved the final version of the manuscript.

**Funding**

This work was supported by Hainan Health Industry Scientific Research Project (20A200178).

**Data Availability**

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

## Declarations

### Ethics approval and consent to participate

Each participant completed an informed consent form after learning about the study objectives. The Ethics Committee of Tangdu Hospital approved this study, and all experimental methods were strictly in compliance with the 1964 Helsinki Declaration.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

Received: 2 February 2023 / Accepted: 3 October 2023

Published online: 01 December 2023

## References

- Medeiros B, Allan AL. Molecular mechanisms of breast Cancer metastasis to the lung: clinical and experimental perspectives. *Int J Mol Sci* 2019, 20(9).
- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and Mortality Worldwide for 36 cancers in 185 countries. *Cancer J Clin*. 2021;71(3):209–49.
- Cao W, Chen HD, Yu YW, Li N, Chen WQ. Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020. *Chin Med J*. 2021;134(7):783–91.
- Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin*. 2022;72(1):7–33.
- Chaffer CL, Weinberg RA. A perspective on cancer cell metastasis. *Sci (New York NY)*. 2011;331(6024):1559–64.
- Rao MY, Wu JB. Invasive ductal breast cancer with extensive subcutaneous metastases in trunk: a case report. *Eur Rev Med Pharmacol Sci*. 2015;19(21):4101–4.
- Dogan L, Kalaylioglu Z, Karaman N, Ozaslan C, Atalay C, Altinok M. Relationships between epidemiological features and tumor characteristics of breast cancer. *Asian Pac J cancer Prevention: APJCP*. 2011;12(12):3375–80.
- Mueller C, Haymond A, Davis JB, Williams A, Espina V. Protein biomarkers for subtyping breast cancer and implications for future research. *Expert Rev Proteomics*. 2018;15(2):131–52.
- Gao C, Zhuang J, Zhou C, Li H, Liu C, Li H, Feng F, Liu R, Sun C. SNP mutation-related genes in breast cancer for monitoring and prognosis of patients: a study based on the TCGA database. *Cancer Med*. 2019;8(5):2303–12.
- Fakhri N, Chad MA, Lahkim M, Houari A, Dehbi H, Belmouden A, El Kadmiri N. Risk factors for breast cancer in women: an update review. *Med Oncol (Northwood Lond Engl)*. 2022;39(12):197.
- Ho WK, Tai MC, Dennis J, Shu X, Li J, Ho PJ, Millwood IY, Lin K, Jee YH, Lee SH, et al. Polygenic risk scores for prediction of breast cancer risk in asian populations. *Genet Medicine: Official J Am Coll Med Genet*. 2022;24(3):586–600.
- Mathias C, Marin AM, Kohler AF, Sanchuki HBS, Sukow N, Beltrame MH, Baal SCS, Sebastião APM, de Fonseca Ribeiro S, Gradia EM. DF : LncRNA-SNPs in a brazilian breast Cancer cohort: a case-control study. *Genes* 2023, 14(5).
- He Y, Liu H, Chen Q, Shao Y, Luo S. Relationships between SNPs and prognosis of breast cancer and pathogenic mechanism. *Mol Genet Genom Med*. 2019;7(9):e871.
- Jarrar YB, Lee SJ. Molecular functionality of cytochrome P450 4 (CYP4) genetic polymorphisms and their clinical implications. *Int J Mol Sci* 2019, 20(17).
- Lim S, Alshagga M, Ong CE, Chieng JY, Pan Y. Cytochrome P450 4B1 (CYP4B1) as a target in cancer treatment. *Hum Exp Toxicol*. 2020;39(6):785–96.
- Wiek C, Schmidt EM, Roellecke K, Freund M, Nakano M, Kelly EJ, Kaisers W, Yarov-Yarovsky V, Kramm CM, Rettie AE, et al. Identification of amino acid determinants in CYP4B1 for optimal catalytic processing of 4-ipomeanol. *Biochem J*. 2015;465(1):103–14.
- Liu X, Jia Y, Shi C, Kong D, Wu Y, Zhang T, Wei A, Wang D. CYP4B1 is a prognostic biomarker and potential therapeutic target in lung adenocarcinoma. *PLoS ONE*. 2021;16(2):e0247020.
- Huang L, Liao J, Chen Y, Mo Z. Association between rs2188380 and the risk of breast cancer in southwest chinese population. *J Clin Lab Anal*. 2019;33(5):e22889.
- Murphy ME, Liu S, Yao S, Huo D, Liu Q, Dolfi SC, Hirshfield KM, Hong CC, Hu Q, Olshan AF, et al. A functionally significant SNP in TP53 and breast cancer risk in african-american women. *NPJ Breast Cancer*. 2017;3:5.
- Yu S, Chen Z, Cheng J, Shi X, Liu J, Zhong P, Song J. Case-control study on CYP4B1 gene polymorphism and susceptibility to gastric cancer in the chinese Han population. *BMC Med Genom*. 2022;15(1):223.
- Sasaki T, Horikawa M, Orikasa K, Sato M, Arai Y, Mitachi Y, Mizugaki M, Ishikawa M, Hiratsuka M. Possible relationship between the risk of japanese bladder cancer cases and the CYP4B1 genotype. *Jpn J Clin Oncol*. 2008;38(9):634–40.
- Tamaki Y, Arai T, Sugimura H, Sasaki T, Honda M, Muroi Y, Matsubara Y, Kanno S, Ishikawa M, Hirasawa N, et al. Association between cancer risk and drug-metabolizing enzyme gene (CYP2A6, CYP2A13, CYP4B1, SULT1A1, GSTM1, and GSTT1) polymorphisms in cases of lung cancer in Japan. *Drug Metab Pharmacokin*. 2011;26(5):516–22.
- Iscan M, KLaavuniemi T, Coban T, Kapucuoglu N, Pelkonen O, Raunio H. The expression of cytochrome P450 enzymes in human breast tumours and normal breast tissue. *Breast Cancer Res Treat*. 2001;70(1):47–54.
- Beyers TB, Helvie M, Bonaccio E, Calhoun KE, Daly MB, Farrar WB, Garber JE, Gray R, Greenberg CC, Greenup R, et al. Breast Cancer screening and diagnosis, Version 3.2018, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Cancer Network: JNCCN*. 2018;16(11):1362–89.
- Edge SB, Compton CC. The american Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol*. 2010;17(6):1471–4.
- Dieci MV, Orvieto E, Dominici M, Conte P, Guarneri V. Rare breast cancer subtypes: histological, molecular, and clinical peculiarities. *Oncologist*. 2014;19(8):805–13.
- Wang S, Zhang K, Tang L, Yang Y, Wang H, Zhou Z, Pang J, Chen F. Association between single-nucleotide polymorphisms in breast Cancer susceptibility genes and clinicopathological characteristics. *Clin Epidemiol*. 2021;13:103–12.
- Liang H. More reliable breast cancer risk assessment for every woman. *Cancer Cell*. 2021;39(4):457–9.
- Thesseling FA, Hutter MC, Wiek C, Kowalski JP, Rettie AE, Girhard M. Novel insights into oxidation of fatty acids and fatty alcohols by cytochrome P450 monooxygenase CYP4B1. *Arch Biochem Biophys*. 2020;679:108216.
- Eun HS, Cho SY, Lee BS, Seong IO, Kim KH. Profiling cytochrome P450 family 4 gene expression in human hepatocellular carcinoma. *Mol Med Rep*. 2018;18(6):4865–76.
- Engin A. Obesity-associated breast Cancer: analysis of risk factors. *Adv Exp Med Biol*. 2017;960:571–606.
- Picon-Ruiz M, Morata-Tarifa C, Valle-Goffin JJ, Friedman ER, Slingerland JM. Obesity and adverse breast cancer risk and outcome: mechanistic insights and strategies for intervention. *CA Cancer J Clin*. 2017;67(5):378–97.
- Terry PD, Rohan TE. Cigarette smoking and the risk of breast cancer in women: a review of the literature. *Cancer Epidemiol Biomarkers Prevention: Publication Am Association Cancer Res Cosponsored Am Soc Prev Oncol*. 2002;11(10 Pt 1):953–71.
- Jones ME, Schoemaker MJ, Wright LB, Ashworth A, Swerdlow AJ. Smoking and risk of breast cancer in the generations study cohort. *Breast cancer Research: BCR*. 2017;19(1):118.
- Wu B, Xing T, Li H, He W, Li B. Association between gene polymorphism of GRK5 and breast cancer risk in chinese population. *J BUON: Official J Balkan Union Oncol*. 2021;26(3):741–6.
- Hsieh YC, Lee CH, Tu SH, Wu CH, Hung CS, Hsieh MC, Chuang CW, Ho YS, Chiou HY. CHRNA9 polymorphisms and smoking exposure synergize to increase the risk of breast cancer in Taiwan. *Carcinogenesis*. 2014;35(11):2520–5.
- Naif HM, Al-Obaide MAI, Hassani HH, Hamdan AS, Kalaf ZS. Association of Cytochrome CYP1A1 gene polymorphisms and Tobacco Smoking with the risk of breast Cancer in women from Iraq. *Front Public Health*. 2018;6:96.
- Lash TL, Bradbury BD, Wilk JB, Aschengrau A. A case-only analysis of the interaction between N-acetyltransferase 2 haplotypes and tobacco smoke in breast cancer etiology. *Breast cancer Research: BCR*. 2005;7(3):R385–393.
- Morabia A, Bernstein M, Héritier S, Khatchatrian N. Relation of breast cancer with passive and active exposure to tobacco smoke. *Am J Epidemiol*. 1996;143(9):918–28.
- Sun YS, Zhao Z, Yang ZN, Xu F, Lu HJ, Zhu ZY, Shi W, Jiang J, Yao PP, Zhu HP. Risk factors and preventions of breast Cancer. *Int J Biol Sci*. 2017;13(11):1387–97.



41. Nathanson SD, Detmar M, Padera TP, Yates LR, Welch DR, Beadnell TC, Scheid AD, Wrenn ED, Cheung K. Mechanisms of breast cancer metastasis. *Clin Exp Metastasis*. 2022;39(1):117–37.
42. Cui P, Zhao Y, Chu X, He N, Zheng H, Han J, Song F, Chen K. SNP rs2071095 in LincRNA H19 is associated with breast cancer risk. *Breast Cancer Res Treat*. 2018;171(1):161–71.
43. Yan R, Cao J, Song C, Chen Y, Wu Z, Wang K, Dai L. Polymorphisms in lncRNA HOTAIR and susceptibility to breast cancer in a Chinese population. *Cancer Epidemiol*. 2015;39(6):978–85.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.