Received: 1 April 2024 | Revised: 23 August 2024 | Accepted: 8 September 2024

DOI: 10.1111/cas.16353

Cancer Science WILEY

ORIGINAL ARTICLE

Identification of sequence polymorphism in the D-loop region of mitochondrial DNA as a risk factor for breast cancer

Qian Nie[1,2,3] | Wenzhe Zhang[1,2,3] | Songping Lin[4] | Meng Huang[5] | Yan Li[1,2,3] | Yibin Qiu[1,2,3] | Jing Li[1,2,3] | Xiaobin Chen[1,2,3] | Yali Wang[1,2,3] | Xin Tong[1,2,3] | Jinqiao Wu[1,2,3] | Peng He[1,2,3] | Qindong Cai[1,2,3] | Lili Chen[1,2,3] | Minyan Chen[1,2,3] | Wenhui Guo[1,2,3] | Yuxiang Lin[1,2,3] | Liuwen Yu[1,2,3] | Jialin Hou[1,2,3] | Weifeng Cai[1,2,3] | Hanxi Chen[1,2,3] | Chuan Wang[1,2,3] | Fangmeng Fu[1,2,3]

[1]Department of Breast Surgery, Fujian Medical University Union Hospital, Fuzhou, Fujian, China

[2]Department of General Surgery, Fujian Medical University Union Hospital, Fuzhou, Fujian, China

[3]Breast Cancer Institute, Fujian Medical University, Fuzhou, Fujian, China

[4]Department of Breast Surgery, Affiliated Quanzhou First Hospital of Fujian Medical University, Quanzhou, Fujian, China

[5]Fujian Center for Disease Control and Prevention, Fuzhou, Fujian, China

Correspondence
Chuan Wang and Fangmeng Fu, Department of Breast Surgery, Fujian Medical University Union Hospital, No. 29, Xin Quan Road, Gulou, Fuzhou 350001, Fujian, China.
Email: dr_chuanwang@fjmu.edu.cn and ffm@fjmu.edu.cn

Funding information
Joint Funds for the Innovation of Science and Technology, Fujian Province, Grant/Award Number: 2018Y9055; Sailing Foundation of Fujian Medical University, Grant/Award Number: 2022QH2022; Pilot project of Fujian Province, Grant/Award Number: 2022Y0018

Abstract

Mitochondrial DNA (mtDNA) variations affect the efficiency of the electron transport chain and production of reactive oxygen species, contributing to carcinogenesis. The D-loop region of mtDNA has emerged as a variation hotspot region in human neoplasia; however, the potential contribution of these variations in breast cancer risk prediction remains unknown. We investigated the relationship between germline single nucleotide polymorphisms (SNPs) in the entire D-loop region and breast cancer risk in Chinese women. Peripheral blood-isolated mtDNA from 2329 patients with breast cancer and 2328 cancer-free controls was examined for SNPs. In the combined cohort, we used traditional risk factors, susceptibility germline polymorphisms, and logistic regression analysis to evaluate the predictive value of susceptibility variants for breast cancer risk. We calculated the area under the receiver operating characteristic curve (AUC) as a measure. We also measured the content of 8-hydroxy-2′-deoxyguanosine (8-OHdG). Individual polymorphisms SNP573 were significantly associated with breast cancer risk in both the discovery and validation cohorts. In the combined cohort, the AUC of the traditional risk factors was 64.3%; after adding susceptibility variants, the AUC was 64.9% (DeLong test, $p = 0.007$). 8-OHdG levels were significantly higher in patients with breast cancer than in controls and higher in individuals with SNP573 than in those negative for this variation. Overall, oxidative stress might be associated with the risk of breast cancer, and SNP573 might be associated with oxidative stress. Our results indicate the risk potential of polymorphisms in the D-loop region in breast cancer in Southern China.

KEYWORDS
breast cancer, D-loop region, mitochondrial DNA, risk, single nucleotide polymorphism

Abbreviations: AUC, the area under the curve; BC, breast cancer; CI, confidence interval; ER, estrogen receptor; HER2, human epidermal growth factor-2; HR, hormone receptor; MAF, minor allele frequency; mtDNA, mitochondrial DNA; OHdG, 8-hydroxy-2′-deoxyguanosine; OR, odds ratios; PR, progesterone receptor; ROC, receiver operating characteristic; ROS, reactive oxygen species; SNPs, single nucleotide polymorphisms.

1 | INTRODUCTION

Breast cancer (BC), a multifactorial and multistep disease, is one of the most common types of malignant neoplasm in women worldwide.[1] Genetic effects, environmental exposure, and gene–environment interactions contribute to the risk of breast cancer.[2–5] In addition to highly penetrant rare variants, such as BRCA1 and BRCA2, low-penetrant but common breast cancer susceptibility single nucleotide polymorphisms (SNPs) are important predictors of disease risk.[6–8]

Cancer development involves the accumulation of various genetic alterations present in both mitochondrial and nuclear genomes. Mitochondria play vital roles in energy production and oxidative phosphorylation.[9] Accordingly, defects in mitochondrial function are thought to contribute to the development and progression of cancer.[10] The human mitochondrial genome is a multicopy closed-circular duplex molecule of 16.569 kb in length, containing 37 genes that encode 13 essential polypeptide subunits of the oxidative phosphorylation system, 2 ribosomal RNAs, and 22 transfer RNAs.[11] Mitochondrial DNA (mtDNA) is believed to be more susceptible to DNA damage and mutates at a higher rate than nuclear DNA because of the lack of protection from histones and chromatin structures, limited DNA repair mechanisms, and exposure to high levels of reactive oxygen species (ROS).[12–14] Increased ROS production and the resulting damage to both mtDNA and nuclear DNA have long been thought to play crucial roles in carcinogenesis.[15,16] Variations in mtDNA are closely associated with degenerative diseases, aging, and cancer.[17] Studies have suggested that mtDNA polymorphisms are associated with various disorders, including neuroblastoma,[18] colorectal cancer,[19] hepatocellular carcinoma,[20] pancreatic cancer,[21] and breast cancer.[22]

The D-loop is the main regulatory region of mitochondrial replication and transcription that encompasses the initial site of heavy chain replication and promoters for heavy and light chain transcription.[12] The D-loop accumulates variations at a higher frequency than other regions of the mitochondrial genome.[23] SNPs in this region might affect mtDNA replication and lead to alterations in the electron transport chain, which is responsible for the release of ROS, thus contributing to cancer initiation and promotion.[24] Moreover, respiratory chain alterations might cause mitochondria-induced apoptotic dysfunction.[25] Specific SNPs in the mtDNA D-loop region have been reported to be substantially associated with many types of tumors, including pancreatic cancer,[26] hepatocellular carcinoma,[27] cervical cancer,[28] renal cell carcinoma,[29] non-Hodgkin lymphoma,[30] epithelial ovarian cancer,[31] and colorectal cancer.[32] Polymorphisms in the D-loop have also been associated with breast cancer risk. Through genotyping a partial sequence of the D-loop in European and American populations, Bai et al.[33] reported that T16519C increased the risk of women developing breast cancer. Tipirisetti et al.[34] screened the entire mitochondrial D-loop region in the South Indian population and found that the frequencies of 310'C' insertion ($P = 0.018$), T16189C ($P = 0.0019$) variants and 310'C'ins/16189C ($P = 0.00019$) haplotype were significantly higher in cases than in controls. However, all of the above studies had common limitations, such as limited sample size, lack of validation studies, and mitochondrial functional analyses (e.g., evaluation of oxidative stress-associated biomarkers).

In this study, we aimed to completely investigate the relationship between germline SNPs in the entire D-loop region of the mitochondrial genome and breast cancer risk in Chinese women. Direct DNA sequencing methods were used to genotype the entire D-loop region of mtDNA in a large sample and an independent validation cohort. Studies that analyze germline variations in the mtDNA D-loop region might provide additional insights into the role of mtDNA variations in breast cancer risk.

2 | MATERIALS AND METHODS

2.1 | Study participants

All study participants were Chinese women from the Fujian Province. A total of 2329 patients with breast cancer and 2328 healthy controls were enrolled in this hospital-based two-stage case–control study, including 720 patients and 724 controls in the training cohort from 2003 to 2013 and 1609 patients and 1604 controls in the validation cohort from 2014 to 2021. Patient eligibility criteria were as follows: (i) histopathological confirmation of invasive BC; and (ii) availability of complete data on potential breast cancer risk factors. Patients were enrolled in the Fujian Medical University Union Hospital, Fujian, China, and each case was histopathologically confirmed by at least two oncologists. After excluding participants with genotyping failures and missing raw data, the final analyses included 642 patients and 630 controls in the training cohort and 1519 patients and 1580 controls in the validation cohort.

Estrogen receptor (ER)/progesterone receptor (PR) positivity was determined by immunohistochemical analysis of the number of positively stained nuclei (≥10%), while HR positivity was defined as being either ER+ or PR+ or both. Tumors were considered human epidermal growth factor-2 (HER2)-positive if cells exhibited strong membrane staining (3+). Expression of 2+ required further in situ hybridization testing for HER2 amplification, whereas an expression of 0 or 1+ was regarded as negative. The subtypes were categorized as previously established[35]: luminal A (ER+, PR+ >20%, HER2−, Ki67 <14%, or grade I when Ki67 was unavailable), luminal B (HR+, HER2−, Ki67 >14%, or grade II/III when Ki67 was unavailable, or HR+, HER2+); HER2 enriched (HR−, HER2+), and triple-negative (HR− and HER2−). Healthy controls were selected from among patients who underwent routine health examinations at the same hospital during the corresponding period. Controls were age-matched (±3 years) healthy individuals without breast or other cancers.

After written informed consent was obtained, each study participant was personally interviewed face to face by trained interviewers

for collecting information, including demographic data, menstrual, reproductive, and breastfeeding histories, as well as family history of breast cancer. Subsequently, a peripheral blood sample (3 mL) was collected from each participant. The study protocol was approved by the ethics committee of Fujian Medical University Union Hospital (2019KJTZD003), and written informed consent was obtained from all individuals before their participation. All the procedures performed adhered to the guidelines set by the World Medical Association (Declaration of Helsinki).

2.2 | Single nucleotide polymorphism selection and genotyping

Genomic DNA was extracted from leukocytes isolated from ethylenediamine tetraacetic acid anti-coagulated whole blood using a Whole-Blood DNA Extraction Kit (Bioteke, Beijing, China) according to the manufacturer's protocol. DNA concentration was quantified using an Epoch Microplate Spectrophotometer (BioTek Instruments, Winooski, VT, USA), while DNA quality was determined via agarose gel electrophoresis. All samples were stored at −20°C. Qualified DNA samples were genotyped using Sanger sequencing.

DNA samples (1 μL) were subjected to 1% agarose electrophoresis for quality check and concentration estimation, and then each sample was diluted to a working concentration of 5–10 ng/μL based on the estimated concentration. Samples without obvious DNA banding on the agarose gel were not diluted. PCR primers were D-Loop-F, 5′-GCCGCAGACCTCCTCATTCT-3′ and D-Loop-R, 5′-GGCTGGCACGAAATTGACCA-3′. The PCR reaction mixture was 20 μL in volume, including 1 × HotStarTaq buffer, 2.0 mM Mg^{2+}, 0.2 mM dNTP, 0.2 μM for each primer, 1 U HotStarTaq polymerase (Qiagen Inc), and 1 μL template DNA. The cycling conditions were: 95°C for 2 min, 11 cycles at 94°C for 20 s, 66°C for 40 s, and 72°C for 2 min; 24 cycles at 94°C for 20 s, 60°C for 30 s, and 72°C for 2 s; 72°C for 2 min; and 4°C hold. PCR purification was performed by adding 0.5 U shrimp alkaline phosphatase and 4 U Exonuclease I to 8 μL of PCR products. The mixture was incubated at 37°C for 60 min and then at 75°C for 15 min. The primers used for sequencing were: D-Loop-R, 5′-GGCTGGCACGAAATTGACCA-3′; D-Loop-SEQF1, 5′-TCCAAGGACAAATCAGAGAAAAAG-3′; D-Loop-SEQF2, 5′-TGAACTGTATCCGACATCTGGTTC-3′; D-Loop-SEQF3, 5′-TGATGTGAGCCCGTCTAAACA-3′; and D-Loop-SEQF4, 5′-GGGATGCTTGCATGTGTAAT. The reaction mixture consisted of 3 μL BigDye 3.1, 2 μL primers (1 μM), and 1–2 μL purified PCR products. The cycling conditions were: 96°C for 1 min; 28 cycles at 96°C for 10 s, 50°C for 5 s, and 60°C for 4 min; and 4°C hold.

The products were sequenced using an ABI3730XL sequencer. The sequencing files were analyzed using the CodonCode Aligner 3.0.1 (CodonCode Corporation, Centerville, VA, USA) and the results were collated after manual proofreading. We sequenced the entire D-loop region, which was 1122 bp long (nucleotides 16,024–16 569 and 1–576) and referred to the revised Cambridge Reference Sequence. In the training cohort, a total of 344 variants were identified in the D-loop of mtDNA in patients with breast cancer and healthy controls.

2.3 | Measurement of 8-OHdG

We randomly selected 132 patients and 150 controls from the entire population. The levels of leukocyte 8-OHdG, a biomarker of oxidative DNA damage, was detected in this subset. Absolute levels of 8-OHdG were measured using the EpiQuik 8-OHdG DNA Damage Quantification Direct Kit (colorimetric) (no. P-6003; Epigentek, Farmingdale, NY, USA) according to the manufacturer's instructions. The suggested amount (300 ng) of DNA was used for each sample. Samples were plated in duplicate, and a recommended standard curve was used for the measurements.

2.4 | Statistical analysis

Differences between case and control participants in selected demographic characteristics and traditional risk factors were compared using Mann–Whitney *U*-tests (for nonnormally distributed continuous variables) or χ^2-tests (for categorical variables). The association between polymorphisms and breast cancer risk was evaluated using logistic regression analysis, estimating the odds ratios (ORs) and 95% confidence intervals (CIs) with and without adjustment for age, body-mass index (BMI), education level, age at menarche, menopausal status, age at first live birth, number of live births, duration of breastfeeding, and family history of breast cancer. To account for multiple testing, Bonferroni correction was used. We also evaluated the relationships between statistically significant SNPs and subgroups of demographic and clinicopathological characteristics using stratification analyses. We evaluated the predictive value of susceptibility variants for breast cancer risk using logistic regression analysis in the combined cohort. For comparison of 8-OHdG contents, the Mann–Whitney *U* test was used. All the above tests were conducted using SPSS 25.0 (IBM, New York, NY, USA). Scatter plots, box plots, and violin plots were generated using GraphPad Prism 8.0.1 (Graph Pad Software, San Diego, CA, USA). Heterogeneity among subgroups was assessed using Stata 12.0 (StataCorp, College Station, TX, USA). All statistical tests were two-sided, and $p < 0.05$ was considered statistically significant.

3 | RESULTS

A total of 2161 patients with pathologically confirmed as having breast cancer, and 2210 tumor-free controls were included in the training and validation cohorts. The case and control participants were comparable with respect to age because of the age-and frequency-matched design of this study (Table 1). In the training cohort, validation cohort and combined cohort, the breast cancer case

TABLE 1 Distribution of demographic characteristics and known breast cancer risk factors for cases and control participants.

Variables	Training cohort			Validation cohort			Combined cohort		
	Cases ($n=642$) no. (%)	Controls ($n=630$) no. (%)	p^*	Cases ($n=1519$) no. (%)	Controls ($n=1580$) no. (%)	p^*	Cases ($n=2161$) no. (%)	Controls ($n=2210$) no. (%)	p^*
Age, year, median (IQR)	46.0 (38–54)	46.5 (39–53)	0.598	47 (39–55)	47 (39–55)	0.130	47 (39–55)	47 (39–54)	0.331
BMI, kg/m^2, median (IQR)	22.432 (20.57–24.34)	22.621 (21.08–24.36)	0.091	22.432 (20.69–24.44)	22.432 (20.57–24.50)	0.978	22.432 (20.69–24.44)	22.491 (20.70–24.44)	0.352
Education level, no. (%)									
<High school	434 (67.6)	228 (36.2)	<0.001	735 (48.4)	612 (38.7)	<0.001	1169 (54.1)	840 (38.0)	<0.001
≥High school	208 (32.4)	402 (63.8)		784 (51.6)	968 (61.3)		992 (45.9)	1370 (62.0)	
Age at menarche, year, median (IQR)	15 (14–16)	15 (14–16)	0.304	15 (14–16)	15 (14–16)	0.031	15 (14–16)	15 (14–16)	0.017
Age at first live birth, year, no. (%)									
<20	27 (4.2)	26 (4.1)	0.294	48 (3.2)	64 (4.1)	0.366	75 (3.5)	90 (4.1)	0.536
20–25	329 (51.2)	350 (55.6)		777 (51.2)	787 (49.8)		1106 (51.2)	1137 (51.4)	
>25 or nulliparous	286 (44.5)	254 (40.3)		694 (45.7)	729 (46.1)		980 (45.3)	983 (44.5)	
Number of live births, median (IQR)	1 (1–2)	2 (1–2)	0.012	1 (1–2)	2 (1–2)	<0.001	1 (1–2)	2 (1–2)	<0.001
Menopausal status, no. (%)									
Premenopausal	427 (66.5)	412 (65.4)	0.675	949 (62.5)	995 (63.0)	0.774	1376 (63.7)	1407 (63.7)	0.995
Postmenopausal	215 (33.5)	218 (34.6)		570 (37.5)	585 (37.0)		785 (36.3)	803 (36.3)	
Duration of breastfeeding, month, median (IQR)	12 (8–24)	12 (6–24)	0.071	12 (7–24)	12 (6–24)	0.091	12 (8–24)	12 (6–24)	0.647
Family history of breast cancer, no. (%)									
No	608 (94.7)	630 (100.0)	<0.001	1381 (90.9)	1531 (96.9)	<0.001	1989 (92.0)	2161 (97.8)	<0.001
Yes	34 (5.3)	0 (0)		138 (9.1)	49 (3.1)		172 (8.0)	49 (2.2)	
ER status, no. (%)									
Positive	413 (64.3)			1073 (70.6)			1486 (68.8)		
Negative	222 (34.6)			435 (28.6)			657 (30.4)		
Unknown	7 (1.1)			11 (0.7)			18 (0.8)		
PR status, no. (%)									
Positive	358 (55.8)			908 (59.8)			1266 (58.6)		
Negative	277 (43.1)			599 (39.4)			876 (40.5)		

(Continues)

TABLE 1 (Continued)

Variables	Training cohort			Validation cohort			Combined cohort		
	Cases ($n=642$) no. (%)	Controls ($n=630$) no. (%)	p^*	Cases ($n=1519$) no. (%)	Controls ($n=1580$) no. (%)	p^*	Cases ($n=2161$) no. (%)	Controls ($n=2210$) no. (%)	p^*
Unknown	7 (1.1)			12 (0.8)			19 (0.9)		
HER2 status, no. (%)									
Positive	186 (29.0)			399 (26.3)			585 (27.1)		
Negative	448 (69.8)			1056 (69.5)			1504 (69.6)		
Unknown	8 (1.2)			64 (4.2)			72 (3.3)		
Subtype, no. (%)									
Luminal A	123 (19.2)			318 (20.9)			441 (20.4)		
Luminal B	289 (45.0)			721 (47.5)			1010 (46.7)		
HER2+	96 (15.0)			205 (13.5)			301 (13.9)		
Triple negative	125 (19.5)			208 (13.7)			333 (15.4)		
Unknown	9 (1.4)			67 (4.4)			76 (3.5)		

Abbreviations: ER, estrogen receptor; HER2, human epidermal growth factor-2; IQR, interquartile range; PR, progesterone receptor.
*p values (two-sided) were derived from Mann–Whitney U-tests (for continuous variables) and χ^2-tests (for categorical variables).

group had lower education levels, a lower number of live births, and a higher proportion of family history of breast cancer than the control group. The distribution of age at menarche was statistically different in the validation cohort and the combined cohort (Figure S1). All the differences were statistically significant ($p<0.05$).

The genotype distribution of 344 variants in the D-loop region was determined in the training cohort. Associations between all 344 variants genotypes and breast cancer risk with and without adjustment for environmental factors in the training cohort are shown in Table S1, although none of these variants reached the Bonferroni-corrected significance level ($\alpha=0.000145$ from 0.05/344). We excluded variants with a minor allele frequency (MAF) less than 0.01. Polymorphisms significantly associated with the breast cancer risk were SNP573, SNP16217, and SNP16335 ($p=0.003$; $p=0.007$; and $p=0.024$) (Table 2). Interestingly, SNP573 was identified to be an insertion variation based on the detection of an insertion of poly C at nts573. SNP16217 and SNP16335 were identified to be substitution variations. Of note, SNP573 identified in the training cohort was replicated in an independent set, with the genotype C2-8 in SNP573 being associated with significantly increased breast cancer risk after adjusting for age, BMI, education level, age at menarche, menopausal status, age at first live birth, number of live births, duration of breastfeeding, and family history of breast cancer (OR, 1.741; 95% CI, 1.232–2.459; Table 3), the remaining SNPs were not statistically significant (Table S2).

We combined the data from the training and validation cohorts and found that the genotype C2-8 in SNP573 was significantly associated with increased breast cancer risk (OR, 1.910; 95% CI, 1.406–2.595) even after adjusting for age, BMI, education level, age at menarche, menopausal status, age at first live birth, number of live births, duration of breastfeeding, and family history of breast cancer (OR, 1.955; 95% CI, 1.425–2.684; Table 3).

We further analyzed the effects of the SNP573 genotypes on the risk of breast cancer among different subgroups based on demographic and clinicopathological characteristics. For SNP573 (Figure 1), a significantly increased risk of breast cancer was consistently observed for different ages, BMI, education levels, menopausal statuses, number of live births, HR statuses, HER2 statuses, and subtypes. However, in the age at menarche subgroups, a significantly increased risk of breast cancer was found in late menarche individuals (OR, 2.052; 95% CI, 1.431–2.943) but not in early menarche individuals (OR, 1.715; 95% CI, 0.881–3.34). In the age at first live birth subgroups, compared with the subgroup between 20 and 25 years old and the subgroup over 25 years old or nullipara, breast cancer risk was not significantly increased in the subgroup under 20 years old (OR, 1.647; 95% CI, 0.35–7.754). In the subgroups of family history of breast cancer, a significantly increased risk of breast cancer was found in individuals without a family history (OR, 2.011; 95% CI, 1.454–2.781) but not in individuals with a family history (OR, 1.014; 95% CI, 0.246–4.185). Considering the width of the confidence intervals for the non-significant categories observed in age at menarche, age at first live birth, and family history, the lack of significance might be due to small numbers. No significant heterogeneity was observed in any of the subgroups, indicating that SNP573

TABLE 2 Associations between SNP genotypes and breast cancer risk in the training cohort.

SNPs	Genotype	Cases ($n=642$) no. (%)	Controls ($n=630$) no. (%)	OR (95% CI)	*p*
573	C	615 (95.8)	621 (98.6)	1.000	
	C2–8	27 (4.2)	9 (1.4)	3.443 (1.512–7.840)	0.003
16,217	T	585 (91.1)	540 (85.7)	1.000	
	C	57 (8.9)	90 (14.3)	0.587 (0.397–0.867)	0.007
16,335	A	627 (97.7)	599 (95.1)	1.000	
	G	15 (2.3)	31 (4.9)	0.459 (0.233–0.902)	0.024

Abbreviations: CI, confidence interval; OR, odds ratio; SNP, single nucleotide polymorphism.

TABLE 3 Associations between SNP573 genotypes and breast cancer risk in the validation and combined cohorts.

SNP573	Genotype	Cases no. (%)	Controls no. (%)	OR (95% CI)	*p*
Validation cohort		$N=1519$	$N=1580$		
	C	1426 (93.9)	1523 (96.4)	1.000	
	C2–8	93 (6.1)	57 (3.6)	1.741 (1.232–2.459)	0.002
Combined cohort		$N=2161$	$N=2210$		
	C	2041 (94.4)	2144 (97.0)	1.000	
	C2–8	120 (5.6)	66 (3.0)	1.955 (1.425–2.684)	0.000033

Abbreviations: CI, confidence interval; OR, odds ratio.

has consistent effects on different stratifications of different demographic and clinicopathological characteristics.

The levels of leukocyte 8-OHdG were quantified in a subset of 132 patients and 150 controls to reflect the overall status of oxidative stress. The levels of 8-OHdG in DNA from patients ($n=132$) and controls ($n=150$) were estimated to be 0.0058 (0.0054, 0.0064) and 0.0049 (0.0045, 0.0055), respectively. Of note, the levels of 8-OHdG were significantly higher in patients with breast cancer than in controls ($p<0.001$; Figure 2).

In addition, the levels of 8-OHdG in DNA from individuals with and without the SNP573 variation were estimated to be 0.0060 (0.0057, 0.0065) and 0.0054 (0.0049, 0.0060), respectively. Interestingly, the levels of 8-OHdG in individuals with SNP573 were higher than those in individuals negative for this variation ($p=0.037$; Figure 3).

In the combined cohort, we used traditional risk factors, susceptibility germline polymorphisms (SNP573), and logistic regression analysis to evaluate the predictive value of susceptibility variants for breast cancer risk (Figure 4). We calculated the area under the receiver operating characteristic curve (AUC) as a measure. The AUC of the traditional risk factors was 0.643 (95% CI, 0.627–0.660); after adding SNP573, the AUC was 64.9% (95% CI, 0.633–0.665) (DeLong test, $p=0.007$). Our data indicated that the risk value of SNP573 in the D-loop region in breast cancer in southern China.

4 | DISCUSSION

In this two-stage case-control study, we investigated the relationship between germline SNPs in the entire D-loop region and the risk of breast cancer in Chinese women. We provided evidence that SNP573 in the D-loop region is a risk factor for breast cancer. Our data showed that oxidative stress might be associated with the risk of breast cancer, and SNP573 might be associated with oxidative stress. Our data also indicated the risk value of SNP573 in the D-loop region in breast cancer in southern China.

The polymorphism at 568–573 (SNP573) in the mtDNA D-loop was first identified by Torroni et al.[36] in 1994. They found that the insertion, which incorporated two to six additional Cs, occurred within a stretch of six Cs between np 568 and np 573, suggesting that mtDNAs with different numbers of Cs are very unstable and can be rapidly fixed at the somatic and germline levels. SNP573 has been detected in individual samples of certain cancers, such as gastric cancer,[37] oral cavity squamous cell carcinoma,[38] ovarian cancer,[39] hepatocellular carcinoma,[40] and thyroid tumors[41]; however, no association with disease risk has been reported. In a study conducted using stored blood samples from 36 patients with breast cancer and 20 controls, Tommasi et al.[42] described length variations at 514–523 nt and 568–573 nt with a higher frequency than that in healthy individuals, at 16.7% and 5% versus 1% and 0%, respectively. This finding suggested that length variation at 568–573 (SNP573) in the mtDNA D-loop region might be involved in breast carcinogenesis.

In our study population, SNP573 was significantly associated with breast cancer risk. We included a larger sample size and an independent validation cohort, which provided more evidence for the crucial role of this polymorphism in breast cancer. To gain further insight into the relationship between SNP573 and oxidative damage, we measured the levels of the DNA damage marker 8-OHdG. In nuclear DNA and mtDNA, 8-OHdG is one of the predominant forms of free radical-induced oxidative lesions and has, therefore, been widely used as a biomarker for oxidative stress and carcinogenesis.[43,44]

SNP573

Characteristics	OR (95% CI)a	*P*-value.	Pb
Age			0.691
<50	2.094 (1.371 - 3.198)	0.001	
≥50	1.832 (1.134 - 2.959)	0.013	
BMI			0.842
<23	1.891 (1.236 - 2.893)	0.003	
≥23	2.023 (1.258 - 3.253)	0.004	
Education			0.479
<High school	1.722 (1.048 - 2.830)	0.032	
≥High school	2.185 (1.449 - 3.293)	0.00019	
Age at menarche			0.647
≤13	1.715 (0.881 - 3.34)	0.113	
>13	2.052 (1.431 - 2.943)	9.4e-05	
Menopausal status			0.674
Premenopausal	1.84 (1.218 - 2.78)	0.004	
Postmenopausal	2.13 (1.297 - 3.498)	0.003	
Age at first live birth			0.946
<20	1.647 (0.35 - 7.754)	0.528	
20-25	1.9 (1.222 - 2.954)	0.004	
>25 or nulliparous	2.096 (1.298 - 3.384)	0.002	
Number of live births			0.89
≤1	1.921 (1.24 - 2.976)	0.003	
≥2	2.012 (1.27 - 3.188)	0.003	
Family history of breast cancer			0.347
No	2.011 (1.454 - 2.781)	2.4e-05	
Yes	1.014 (0.246 -4.185)	0.985	
ER status			0.667
Positive	1.982 (1.407 - 2.793)	9.2e-05	
Negative	1.752 (1.133 - 2.708)	0.012	
PR status			0.658
Positive	1.782 (1.238 - 2.564)	0.002	
Negative	2.017 (1.368 - 2.974)	0.000397	
HR status			0.765
Positive	1.961 (1.392 - 2.762)	0.000118	
Negative	1.799 (1.162 - 2.784)	0.008	
HER2 status			0.781
Positive	1.744 (1.106 - 2.751)	0.017	
Negative	1.894 (1.341 - 2.674)	0.000284	
Subtype			0.987
LuminalA	1.966 (1.221 - 3.167)	0.005	
Luminal B	1.8 (1.217 - 2.661)	0.003	
HER2+	1.796 (1.017 - 3.171)	0.044	
Triple negative	1.705 (0.969 - 3.000)	0.064	

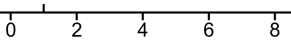

FIGURE 1 Associations between SNP573 and the risk of breast cancer among different subgroups based on demographic and clinicopathological characteristics. ORs and their 95% CIs for age, BMI, education level, age at menarche, menopausal status, age at first live birth, number of live births, duration of breastfeeding, and family history of breast cancer. *P*b for heterogeneity test. BMI, body mass index; CI, confidence interval; OR, odds ratio.

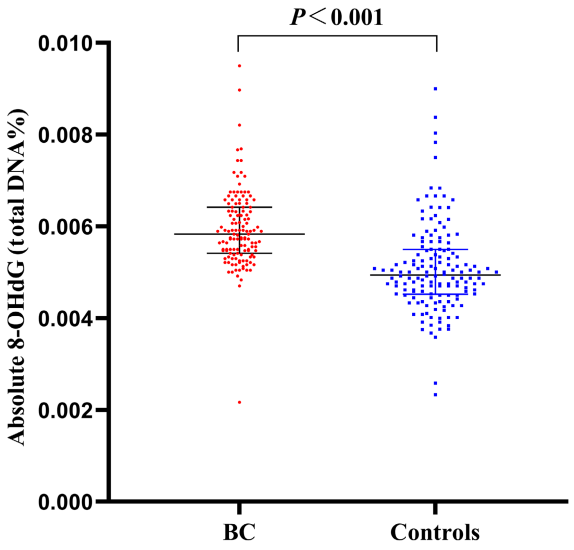

FIGURE 2 The difference in 8-OHdG between patients with breast cancer and healthy controls. The middle horizontal line in the scatter plot represents the median, and the upper and lower horizontal lines represent the upper and lower quartiles, respectively. *p* for Mann–Whitney *U*-test. BC, breast cancer.

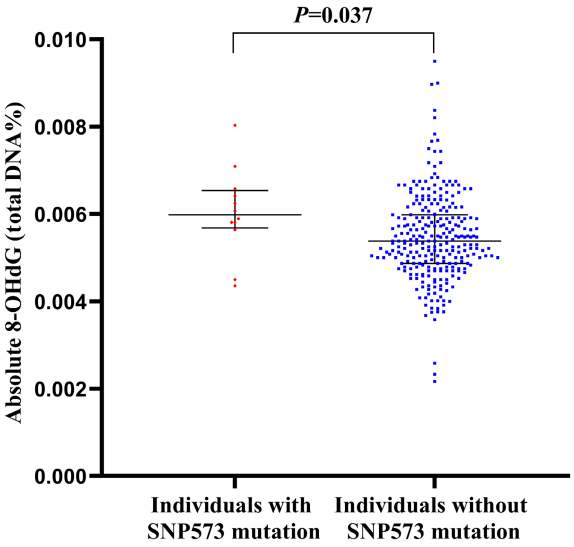

FIGURE 3 The difference in 8-OHdG between individuals with SNP573 and individuals without SNP573. The individuals including both cases and controls. The middle horizontal line in the scatter plot represents the median, and the upper and lower horizontal lines represent the upper and lower quartiles respectively. *p* for Mann–Whitney *U*-test.

The accumulation of 8-OHdG in DNA has predictive significance in breast cancer risk assessment and is conceivably a major contributor to the development of breast neoplasia.[45,46] In our study, the higher levels of 8-OHdG in patients with breast cancer compared with those in healthy controls strongly supported the role of oxidative stress in breast cancer. Higher levels of oxidative stress markers in individuals carrying the SNP573 pathogenic variation further implied an association between this variation and oxidative stress.

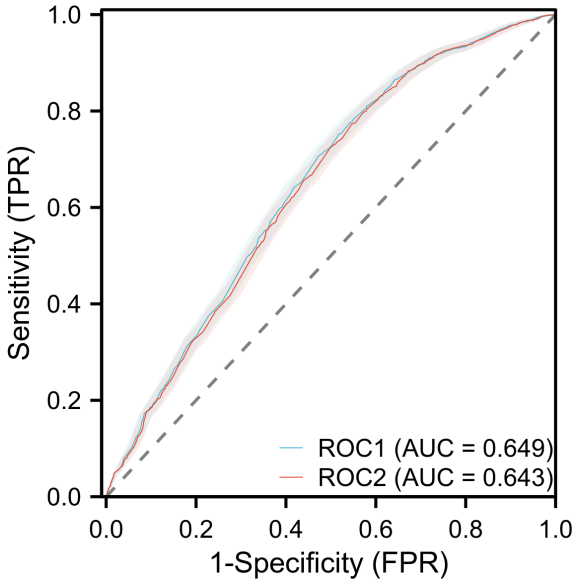

FIGURE 4 The predictive value of traditional risk factors and SNP573 for the risk of breast cancer in combined cohort. The blue curve included traditional risk factors and SNP573; the red curve only included traditional risk factors. Traditional risk factors included age, body-mass index (BMI), education level, age at menarche, menopausal status, age at first live birth, number of live births, duration of breastfeeding, and family history of breast cancer. The shaded part indicates the 95% confidence interval for the AUC. *p* for Delong test. AUC, area under curve; ROC, receiver operating characteristic.

Overall, our data also showed that oxidative stress might be associated with the risk of breast cancer, and SNP573 might be associated with oxidative stress. Extensive studies exploring the role of mtDNA damage and mitochondrial dysfunction in these disorders might provide molecular targets for the development of pharmacological agents for treating these diseases.

Previous studies on mtDNA D-loop variations have focused on SNPs, and only a few studies have evaluated the association between germline mtDNA variations in the D-loop region and cancer. Ye et al.[47] used data from a case-control study conducted among Chinese women in Shanghai that included 1058 cases and 1129 age frequency-matched community controls and found no association between breast cancer risk and mtDNA D-loop polymorphisms. Given that these studies only sequenced partial fragments of the D-loop region, the number of mtDNA sequence variants that are potentially related to breast cancer susceptibility might far exceed that observed. By genotyping the partial sequence of the D-loop in European and American women, Bai et al.[33] suggested that SNP T16519C in the D-loop region increases breast cancer risk (OR, 1.98; 95% CI, 1.25–3.12; $p=0.0030$, adjusted $p=0.0366$). Tipirisetti et al.[34] screened the entire mitochondrial D-loop region in a South Indian population and found that the frequencies of 310"C" insertion and T16189C variants were substantially higher in patients compared with those in controls. This discrepancy might be due to genetic and ethnic variability among the populations studied. Notably, all of the above studies had common

limitations, such as limited sample size and lack of validation studies and mitochondrial functional analyses.

To conclude, we found that the germline SNP573 in the D-loop region of mtDNA is an independent risk factor for breast cancer in southern Chinese population, and oxidative stress might be associated with the risk of breast cancer, and SNP573 might be associated with oxidative stress. Our results indicate the risk potential of polymorphisms in the D-loop region in breast cancer in southern China, and SNP573 should be included in a panel of molecular biomarkers for breast cancer susceptibility. Future studies with larger sample sizes and more mitochondrial function indicators are required to explore the potential functional mechanisms and validate the predictive values of the polymorphisms identified in this study. The utility of these polymorphisms in predicting breast cancer risk is promising for breast cancer prevention.

AUTHOR CONTRIBUTIONS

Qian Nie: Conceptualization; data curation; formal analysis; writing – original draft. **Wenzhe Zhang:** Conceptualization; data curation; formal analysis. **Songping Lin:** Conceptualization; data curation; formal analysis. **Meng Huang:** Conceptualization; data curation; formal analysis. **Yan Li:** Conceptualization; data curation; formal analysis. **Yibin Qiu:** Conceptualization; data curation; formal analysis. **Jing Li:** Conceptualization; data curation; formal analysis. **Xiaobin Chen:** Conceptualization; data curation; formal analysis. **Yali Wang:** Conceptualization; data curation; formal analysis. **Xin Tong:** Conceptualization; data curation; formal analysis. **Jinqiao Wu:** Conceptualization; data curation; formal analysis. **Peng He:** Conceptualization; data curation; formal analysis. **Qindong Cai:** Conceptualization; data curation; formal analysis. **Lili Chen:** Conceptualization; data curation; formal analysis. **Minyan Chen:** Conceptualization; data curation; formal analysis. **Wenhui Guo:** Conceptualization; data curation; formal analysis. **Yuxiang Lin:** Conceptualization; data curation; formal analysis. **Liuwen Yu:** Conceptualization; data curation; formal analysis. **Jialin Hou:** Conceptualization; data curation; formal analysis. **Weifeng Cai:** Conceptualization; data curation; formal analysis. **Hanxi Chen:** Conceptualization; data curation; formal analysis. **Chuan Wang:** Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; resources; software; supervision; validation; visualization; writing – original draft; writing – review and editing. **Fangmeng Fu:** Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; resources; software; supervision; validation; visualization; writing – original draft; writing – review and editing.

ACKNOWLEDGMENTS

We acknowledge the individuals who enrolled in this study and made this project possible. This study was supported by Joint Funds for the Innovation of Science and Technology, Fujian Province [grant no. 2018Y9055]; Pilot project of Fujian Province [grant no. 2022Y0018]; Sailing Foundation of Fujian Medical University [grant no. 2022QH2022].

FUNDING INFORMATION

This study was supported by Joint Funds for the Innovation of Science and Technology, Fujian Province [grant no. 2018Y9055]; Pilot project of Fujian Province [grant no.2022Y0018]; Sailing Foundation of Fujian Medical University [grant no. 2022QH2022].

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interests.

ETHICS STATEMENT

Approval of the research protocol by an Institutional Review Board: The study protocol was approved by the ethics committee of Fujian Medical University Union Hospital (2019KJTZD003).

Written informed consent was obtained from all subjects before their participation. All the procedures performed in studies involving human participants adhere to the World Medical Association (Declaration of Helsinki).

Registry and the Registration No. of the study/trial: N/A.

Animal Studies: N/A.

ORCID

Qian Nie https://orcid.org/0009-0002-7851-7088
Chuan Wang https://orcid.org/0000-0002-4451-0788
Fangmeng Fu https://orcid.org/0000-0002-3065-5708

REFERENCES

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. *CA Cancer J Clin*. 2024;74(1):12-49.
2. McPherson K, Steel CM, Dixon JM. ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. *BMJ*. 2000;321(7261):624-628.
3. Collaborative Group on Hormonal Factors in Breast Cancer. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118,964 women with breast cancer from 117 epidemiological studies. *Lancet Oncol*. 2012;13(11):1141-1151.
4. Yuan X, Yi F, Hou C, et al. Induced abortion, birth control methods, and breast cancer risk: a case-control study in China. *J Epidemiol*. 2019;29(5):173-179.
5. Chan DSM, Abar L, Cariolou M, et al. World Cancer Research Fund international: continuous update project-systematic literature review and meta-analysis of observational cohort studies on physical activity, sedentary behavior, adiposity, and weight change and breast cancer risk. *Cancer Causes Control*. 2019;30(11):1183-1200.
6. Mavaddat N, Pharoah PD, Michailidou K, et al. Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst*. 2015;107(5):djv036.
7. Mavaddat N, Michailidou K, Dennis J, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am J Hum Genet*. 2019;104(1):21-34.
8. Van Veen EM, Brentnall AR, Byers H, et al. Use of single-nucleotide polymorphisms and mammographic density plus classic risk factors for breast cancer risk prediction. *JAMA Oncol*. 2018;4(4):476-482.
9. Wallace DC. Mitochondrial diseases in man and mouse. *Science*. 1999;283(5407):1482-1488.
10. Carew JS, Huang P. Mitochondrial defects in cancer. *Mol Cancer*. 2002;19:1-12.
11. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet*. 1999;23(2):147.

12. Taylor RW, Turnbull DM. Mitochondrial DNA variations in human disease. *Nat Rev Genet*. 2005;6(5):389-402.
13. Beal MF. Mitochondria, free radicals, and neurodegeneration. *Curr Opin Neurobiol*. 1996;6(5):661-666.
14. Lightowlers RN, Chinnery PF, Turnbull DM, Howell N. Mammalian mitochondrial genetics: heredity, heteroplasmy and disease. *Trends Genet*. 1997;13(11):450-455.
15. Penta JS, Johnson FM, Wachsman JT, Copeland WC. Mitochondrial DNA in human malignancy. *Mutat Res*. 2001;488(2):119-133.
16. Luo Y, Ma J, Lu W. The significance of mitochondrial dysfunction in cancer. *Int J Mol Sci*. 2020;21(16):5598.
17. Stewart JB, Chinnery PF. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nat Rev Genet*. 2015;16(9):530-542.
18. Chang X, Liu Y, Glessner J, et al. Identification of mitochondrial DNA variants associated with risk of neuroblastoma. *J Natl Cancer Inst*. 2022;114(6):910-913.
19. Yuan Q, Su L, Wang T, et al. Mitochondrial DNA haplogroup M7 confers a reduced risk of colorectal cancer in a Han population from northern China. *J Cell Mol Med*. 2021;25(15):7538-7544.
20. Jin Y, Yu Q, Zhou D, et al. The mitochondrial DNA 9-bp deletion polymorphism is a risk factor for hepatocellular carcinoma in the Chinese population. *Genet Test Mol Biomarkers*. 2012;16(5):330-334.
21. Lam ET, Bracci PM, Holly EA, et al. Mitochondrial DNA sequence variation and risk of pancreatic cancer. *Cancer Res*. 2012;72(3):686-695.
22. Canter JA, Kallianpur AR, Parl FF, Millikan RC. Mitochondrial DNA G10398A polymorphism and invasive breast cancer in African-American women. *Cancer Res*. 2005;65(17):8028-8033.
23. Michikawa Y, Mazzucchelli F, Bresolin N, Scarlato G, Attardi G. Aging-dependent large accumulation of point variations in the human mtDNA control region for replication. *Science*. 1999;286(5440):774-779.
24. Bandy B, Davison AJ. Mitochondrial variations may increase oxidative stress: implications for carcinogenesis and aging? *Free Radic Biol Med*. 1990;8(6):523-539.
25. Zamzami N, Kroemer G. The mitochondrion in apoptosis: how Pandora's box opens. *Nat Rev Mol Cell Biol*. 2001;2(1):67-71.
26. Wang L, Bamlet WR, de Andrade M, et al. Mitochondrial genetic polymorphisms and pancreatic cancer risk. *Cancer Epidemiol Biomarkers Prev*. 2007;16(7):1455-1459.
27. Zhang R, Zhang F, Wang C, Wang S, Shiao YH, Guo Z. Identification of sequence polymorphism in the D-loop region of mitochondrial DNA as a risk factor for hepatocellular carcinoma with distinct etiology. *J Exp Clin Cancer Res*. 2010;29(1):130.
28. Zhai K, Chang L, Zhang Q, Liu B, Wu Y. Mitochondrial C150T polymorphism increases the risk of cervical cancer and HPV infection. *Mitochondrion*. 2011;11(4):559-563.
29. Xu J, Guo Z, Zhang J, et al. Single nucleotide polymorphisms in the mitochondrial displacement loop and age-at-onset of renal cell carcinoma. *Sci Rep*. 2013;3:2408.
30. Fan H, Wang C, Guo Z. Single nucleotide polymorphisms in the mitochondrial displacement loop and age at onset of non-Hodgkin lymphoma. *Onco Targets Ther*. 2013;6:61041-61045.
31. Liu S, Shi S, Li Y, Kong D. Identification of sequence nucleotide polymorphisms in the D-loop region of mitochondrial DNA as a risk factor for epithelial ovarian cancer. *Mitochondrial DNA A DNA Mapp Seq Anal*. 2016;27(1):9-11.
32. Guo Z, Zhao S, Fan H, du Y, Zhao Y, Wang G. Identification of sequence polymorphisms in the D-loop region of mitochondrial DNA as a risk factor for colon cancer. *Mitochondrial DNA A DNA Mapp Seq Anal*. 2016;27(6):4244-4245.
33. Bai RK, Leal SM, Covarrubias D, Liu A, Wong LJC. Mitochondrial genetic background modifies breast cancer risk. *Cancer Res*. 2007;67(10):4687-4694.
34. Tipirisetti NR, Govatati S, Pullari P, et al. Mitochondrial control region alterations and breast cancer risk: a study in South Indian population. *PLoS One*. 2014;9(1):e85363.
35. Goldhirsch A, Winer EP, Coates AS, et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen international expert consensus on the primary therapy of early breast cancer 2013. *Ann Oncol*. 2013;24(9):2206-2223.
36. Torroni A, Lott MT, Cabell MF, Chen YS, Lavergne L, Wallace DC. MtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. *Am J Hum Genet*. 1994;55(4):760-776.
37. Wu CW, Yin PH, Hung WY, et al. Mitochondrial DNA variations and mitochondrial DNA depletion in gastric cancer. *Genes Chromosomes Cancer*. 2005;44(1):19-28.
38. Lai CH, Huang SF, Liao CT, Chen IH, Wang HM, Hsieh LL. Clinical significance in oral cavity squamous cell carcinoma of pathogenic somatic mitochondrial variations. *PLoS One*. 2013;8(6):e65578.
39. Van Trappen PO, Cullup T, Troke R, et al. Somatic mitochondrial DNA variations in primary and metastatic ovarian cancer. *Gynecol Oncol*. 2007;104(1):129-133.
40. Qiao L, Ru G, Mao Z, et al. Mitochondrial DNA depletion, mitochondrial variations and high TFAM expression in hepatocellular carcinoma. *Oncotarget*. 2017;8(48):84373-84383.
41. Maximo V, Soares P, Lima J, et al. Mitochondrial DNA somatic variations (point variations and large deletions) and mitochondrial DNA variants in human thyroid pathology: a study with emphasis on Hurthle cell tumors. *Am J Pathol*. 2002;160(5):1857-1865.
42. Tommasi S, Favia P, Weigl S, et al. Mitochondrial DNA variants and risk of familial breast cancer: an exploratory study. *Int J Oncol*. 2014;44(5):1691-1698.
43. Wu D, Liu B, Yin J, et al. Detection of 8-hydroxydeoxyguanosine (8-OHdG) as a biomarker of oxidative damage in peripheral leukocyte DNA by UHPLC-MS/MS. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2017;1064:10641-10646.
44. Karihtala P, Kauppila S, Puistola U, Jukkola-Vuorinen A. Divergent behaviour of oxidative stress markers 8-hydroxydeoxyguanosine (8-OHdG) and 4-hydroxy-2-nonenal (HNE) in breast carcinogenesis. *Histopathology*. 2011;58(6):854-862.
45. Musarrat J, Arezina-Wilson J, Wani AA. Prognostic and aetiological relevance of 8-hydroxyguanosine in human breast carcinogenesis. *Eur J Cancer*. 1996;32A(7):1209-1214.
46. Pande D, Negi R, Karki K, Khanna S, Khanna RS, Khanna HD. Oxidative damage markers as possible discriminatory biomarkers in breast carcinoma. *Transl Res*. 2012;160(6):411-418.
47. Ye C, Gao YT, Wen W, et al. Association of mitochondrial DNA displacement loop (CA)n dinucleotide repeat polymorphism with breast cancer risk and survival among Chinese women. *Cancer Epidemiol Biomarkers Prev*. 2008;17(8):2117-2122.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Nie Q, Zhang W, Lin S, et al. Identification of sequence polymorphism in the D-loop region of mitochondrial DNA as a risk factor for breast cancer. *Cancer Sci*. 2024;115:4064-4073. doi:10.1111/cas.16353