Received: 3 January 2024 | Accepted: 6 June 2024

DOI: 10.1002/cai2.142

CANCER INNOVATION

ORIGINAL ARTICLE

Identification of significant single-nucleotide polymorphisms associated with breast cancer recurrence and metastasis using GWAS

Shujuan Sun | Sha Yin | Jie Huang | Dongdong Zhou | Qiaorui Tan | Xiaochu Man | Wen Wang | Jiale Zhang | Huihui Li

Department of Breast Medical Oncology, Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan, Shandong, China

Correspondence
Huihui Li, Department of Breast Medical Oncology, Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, 440 Jiyan Rd, Jinan 250117, Shandong, China.
Email: huihuili82@163.com

Funding information
Beijing Science and Technology Innovation Fund, Grant/Award Number: KC2021-ZZ-0010-1; Shandong Provincial Natural Science Foundation Cancer Prevention Joint Fund, Grant/Award Number: ZR2022LZL004; Shanghai Cancer Prevention and Treatment Foundation, Grant/Award Number: CYBER-2022-003

Abstract

Background: Identification of risk genes and loci associated with the recurrence and metastasis of breast cancer (BC) is of utmost importance. Genome-wide association studies (GWASs) represent valuable tools for identifying the disease risk associated with a given single-nucleotide polymorphism (SNP); they offer significant insights into the disease progression mechanism by analyzing SNP information of the entire genome. Though GWAS has already identified several genetic susceptibility SNPs for BC, their significance in the recurrence and metastasis of this cancer remains unclear. Here, we used a GWAS approach to identify SNPs specifically associated with the risk of BC recurrence and metastasis.

Methods: This study adopted a two-stage GWAS approach. In the first stage, 97 pairs of BC patients with or without recurrence and metastasis, treated at the Shandong Cancer Hospital and Institute from November 2013 to April 2014, were identified using propensity score matching. DNA extracted from the patient peripheral blood was then subjected to Illumina ASA chip analysis for genome-wide SNP detection. In the second stage, the findings were verified in a validation set of 854 BC patients recruited at the same hospital from May 2014 to June 2015. SNP genotyping was performed using time-of-flight mass spectrometry. The SNP loci and their corresponding genes and pathways were analyzed using the DAVID (https://david.ncifcrf.gov/) online enrichment analysis tool.

Abbreviations: BC, breast cancer; *CASC16*, cancer susceptibility candidate 16; CI, confidence interval; DFS, disease-free survival; EAS, East Asian Population; GWAS, genome-wide association study; HER2, human epidermal growth factor receptor 2; HR, hazard ratio; IQR, interquartile range; MAF, minor allele frequency; *NAMPT*, nicotinamide phosphoribosyl transferase; *PVT1*, plasmacytoma variant translocation 1; SNP, single-nucleotide polymorphism.

Shujuan Sun, Sha Yin, and Jie Huang contributed equally to this study.

Results: Based on the GWAS results, 191 SNP-related genes significantly associated with BC recurrence and metastasis were identified as expression quantitivative trait loci ($p < 0.001$). Functional and pathway enrichment analyses subsequently revealed the potential involvement of glutamatergic synaptic transmission, calcium signaling, and insulin secretion pathways in BC recurrence and metastasis. Based on genotype correlation and database expression levels, rs10108514, rs12920540, rs4273077, and rs4730155 were found to be significantly associated with the risk of BC recurrence and metastasis.

Conclusion: Our study suggests that the SNPs rs10108514, rs12920540, rs4273077, and rs4730155 are correlated with the risk of BC recurrence and metastasis, potentially by being implicated in glutamatergic synaptic transmission, calcium signaling, and insulin secretion pathways.

KEYWORDS

breast cancer, metastasis, recurrence, single-nucleotide polymorphisms

1 | INTRODUCTION

Breast cancer (BC) is the most common malignancy affecting women worldwide, with an increasing incidence and a low average age of onset [1]. Recent advances in early screening and treatment strategies have significantly improved the outcomes for BC patients. However, a significant proportion (20%–30%) of patients with early-stage BC still face the challenge of disease progression to advanced stages, characterized by local recurrence or distant metastasis [2–4]. Despite improved accessibility to various treatment options, based on cancer stage and molecular profiles, some patients with BC are still at a significant risk of disease progression [5]. Hence, the primary objective of current treatment strategies for BC is to prevent recurrence and metastasis by ensuring the early identification of patients at high risk of BC progression. Although some progress in delineating the molecular mechanisms underlying the recurrence and metastasis of BC has been made [6], the heterogeneity of this disease limits the generalizability of these findings. Thus, the development of individualized treatment approaches is becoming increasingly important.

Predicting tumor recurrence and metastasis for the clinical monitoring and treatment of patients with BC is challenging [7]. For instance, an individual's genetic characteristics have a substantial impact on BC tumor initiation and progression [8]. Genome-wide association studies (GWASs) enable the exploration of the relationship between genes and diseases or traits by comparing large-scale genetic variations. GWAS are highly effective at identifying genes that render individuals more susceptible to diseases such as BC. Single-nucleotide polymorphisms (SNPs), the most common form of genetic variation in the human genome, can affect gene expression or protein function, thus contributing to the development of diseases, including cancer [9]. Numerous preliminary studies [10–12] have identified SNPs that are associated with BC risk and prognosis. For example, Miedl and colleagues [13] reported that the SNPs targeting estrogen receptor 1, specifically rs2046210 and rs9383590, were correlated with BC risk. Furthermore, Cui et al. [14] demonstrated that the SNP rs2071095 might influence the susceptibility of Chinese women to BC by affecting the expression of the *H19* gene.

Though this type of research has identified certain SNPs that are associated with BC outcomes, the use of candidate SNP research strategies is limited by the restricted number of genetic variations and the reliance on existing knowledge of cancer biology. Conversely, GWAS offers comprehensive gene coverage, enabling a complete genomic investigation, which can reveal greater numbers of disease or phenotype risk loci [15]. This allows for precise genome mapping and in-depth investigations of diseases or phenotypes. The identified risk SNPs, together with other prognostic factors, may provide valuable information to enable the effective prediction, prevention, and treatment of BC.

Here, we performed GWAS of patients with and without BC recurrence by adopting a multi-stage, case–control study design and using Illumina ASA chip technology. After identifying SNPs associated with the risk of BC recurrence, we selected a subset of these loci for further validation in an expanded BC cohort to identify the true SNP variants associated with BC recurrence. This research aims to provide reliable molecular markers that can be used to predict the rates of BC progression in clinical practice and offer theoretical insights into the molecular mechanisms underlying BC recurrence and metastasis.

2 | METHODS

2.1 | Study design

This study used a two-phase GWAS research strategy (Figure 1). Using propensity score matching, 97 pairs of BC patients, treated at the Shandong Cancer Hospital and Institute between November 2013 and April 2014, with or without recurrence and metastasis at the first stage, were selected for this study. The DNA was extracted from the patient peripheral blood samples and used for genome-wide detection on Illumina ASA chips. The SNP loci associated with BC recurrence and metastasis were identified using a logistic regression analysis of GWAS data. Between May 2014 and June 2015, a total of 854 BC patients from the same hospital were enrolled in the second phase of the study. SNP genotyping was performed using time-of-flight mass spectrometry to confirm the association with BC recurrence and metastasis.

2.2 | Study subjects

The study population consisted of patients with BC who visited the Shandong Cancer Hospital and Institute between November 1, 2013, and June 30, 2015. The inclusion criteria were as follows: (1) women with primary BC, who had a histopathologically confirmed diagnosis of BC, (2) local surgical treatment of the breast, and (3) blood and DNA samples that met the quality control standards. The exclusion criteria were as follows: (1) primary stage IV BC, (2) double primary BC, (3) a history of other malignancies, (4) incomplete clinicopathological data, (5) <1 month of disease-free survival (DFS), or (6) unknown loss to follow-up. Based on these criteria, 1048 female BC patients were enrolled in the study. None of these individuals had any familial connection to the Han Chinese ethnic group. The Ethics Committee of Shandong Cancer Hospital and Institute approved this study (approval number: SDTHEC2022009018). Patients signed a consent form, and 2 mL of peripheral blood was collected from the elbow vein. The samples were stored at −80°C until use.

2.3 | DNA extraction and Illumina ASA microarray whole-genome assay

Peripheral venous blood (2 mL) was collected from patients into EDTA-containing tubes. DNA was extracted

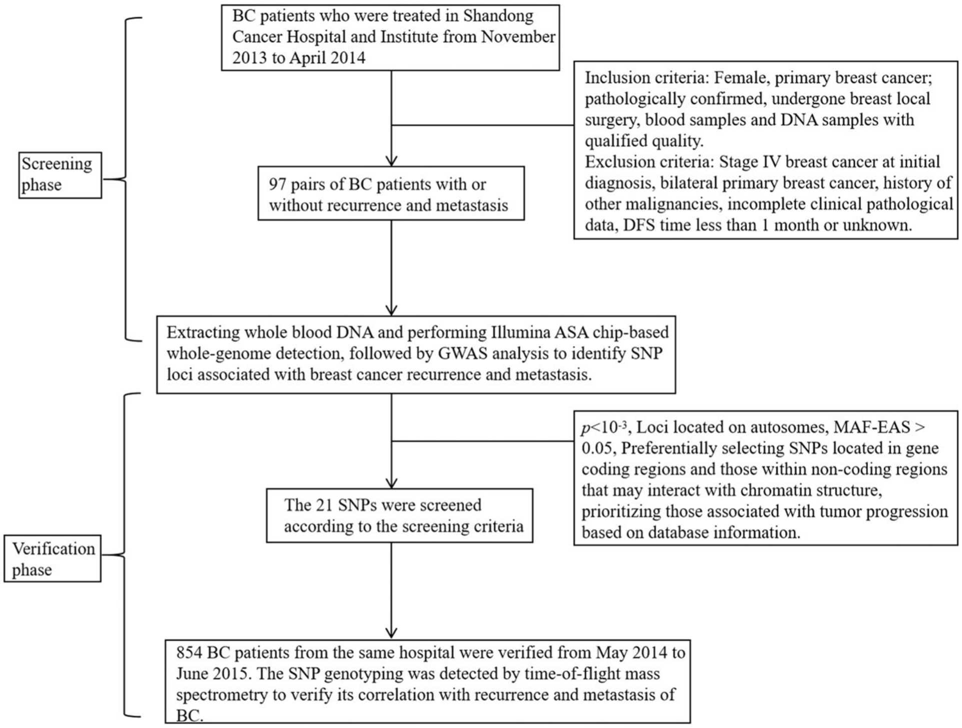
Screening phase
BC patients who were treated in Shandong Cancer Hospital and Institute from November 2013 to April 2014
Inclusion criteria: Female, primary breast cancer; pathologically confirmed, undergone breast local surgery, blood samples and DNA samples with qualified quality.
Exclusion criteria: Stage IV breast cancer at initial diagnosis, bilateral primary breast cancer, history of other malignancies, incomplete clinical pathological data, DFS time less than 1 month or unknown.
97 pairs of BC patients with or without recurrence and metastasis
Extracting whole blood DNA and performing Illumina ASA chip-based whole-genome detection, followed by GWAS analysis to identify SNP loci associated with breast cancer recurrence and metastasis.
Verification phase
p<10-3, Loci located on autosomes, MAF-EAS > 0.05, Preferentially selecting SNPs located in gene coding regions and those within non-coding regions that may interact with chromatin structure, prioritizing those associated with tumor progression based on database information.
The 21 SNPs were screened according to the screening criteria
854 BC patients from the same hospital were verified from May 2014 to June 2015. The SNP genotyping was detected by time-of-flight mass spectrometry to verify its correlation with recurrence and metastasis of BC.

FIGURE 1 Study design flow diagram. BC, breast cancer; DFS, disease-free survival; EAS, East Asian population; GWAS, genome-wide association studies; MAF, minor allele frequency; SNP, single-nucleotide polymorphism.

from the blood samples using a DNA extraction kit and stored at −80°C. The concentration and purity of the DNA were determined, and then each sample was subjected to experimental quality control using the Illumina ASA high-throughput genotyping chip. The samples were chosen based on a detection rate >90%, a minor allele frequency (MAF) >0.01 at each locus, and a Hardy–Weinberg equilibrium p value <0.000001.

2.4 | SNP selection and validation

After analyzing the GWAS data, SNP loci that met the following criteria: p value $<10^{-3}$, loci located on autosomes, and an MAF >0.05 in the East Asian population (EAS) were selected. The dbSNP database (https://www.ncbi.nlm.nih.gov/) was searched to find the genes associated with the SNP loci mentioned above and their positions on the DNA strands of a given gene. A search was performed in PubMed (http://www.ncbi.nlm.nih.gov/pubmed). The three-dimensional (3D) SNP database (http://cbportal.org/3dsnp/) was searched to determine the current status of the SNP loci, focusing specifically on malignant tumors and breast tumors, and retrieved the chromatin spatial interactions of SNP loci. The priority was to identify SNP loci in genes related to tumor evolution, which were present in gene-coding regions, and that may interact with chromatin spatial structures in the noncoding regions. This led to the selection of 21 SNPs and their inclusion in the subsequent validation phase. Primer sequences were designed and synthesized based on the SNP loci identified through screening. The SNP genotypes were detected using time-of-flight mass spectrometry.

2.5 | Bioinformatics analysis

This study analyzed gene expression data obtained from both BC and normal tissues. A combination of two databases, namely, The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/) and the Genotype-Tissue Expression (GTEx) database (https://commonfund.nih.gov/GTEx/), were used to examine differential gene expression. Based on the results of SNP loci in the 3D SNP database (http://cbportal.org/3dsnp/), information on the regulation mediated by the noncoding SNP loci and chromatin loops was obtained. This approach provided a new perspective for studying the influence of SNPs on gene regulation. Furthermore, the DAVID online enrichment analysis website (https://david.ncifcrf.gov/) was used to investigate the enrichment of relevant pathways, enabling an exploration of correlations between SNPs and enriched pathways.

2.6 | Statistical analysis

The statistical analysis was performed using SPSS 22.0 software. Study subject data with a normal distribution were conveyed as the mean ± standard deviation, whereas the median with interquartile range (IQR) was used to describe data with a skewed distribution. The t test was used to determine the statistical significance of differences between two groups with normal distributions, whereas the Wilcoxon rank-sum test was used to compare groups whose distributions were not normal. Statistical data were expressed as the number of patients and percentages, whereas the chi-square test or Fisher's exact test was used to compare the two groups. The Kaplan–Meier method was used to calculate survival time and plot survival curves, whereas the log-rank test was employed to compare the differences in DFS of patients bearing different SNP loci. Cox regression analysis was used to adjust for patient age, hormone receptor status, HER-2 status, and TNM staging. The hazard ratio (HR) and 95% confidence intervals (CIs) for recurrent metastasis were then calculated. Statistical tests were performed as two-sided probability tests, with a p value <0.05 considered as a measure of statistically significant differences.

3 | RESULTS

3.1 | Basic information about the study subjects

The study participants were divided into the screening and validation cohorts (Table 1). The screening cohort consisted of 194 participants, with a median age of 46 (IQR = 40–51) years. The validation cohort included 854 participants, with a median age of 47.5 (IQR = 41–55) years. There were no statistically significant differences between the two groups regarding BC laterality, pathological type, histological grading, TNM staging, and family history ($p > 0.05$). The predominant pathological BC type observed was invasive ductal carcinoma, comprising 80% of cases. Approximately 35% of cases were considered as Grade III based on the histological grading, whereas 45% were classified as Grades I–II. Regarding TNM staging, Stage II was the most prevalent, representing approximately 41% of cases, followed by Stage III (~35%) and Stage I (~24%). Approximately 25% of the patients had a family history (immediate relatives) of cancer.

3.2 | Results of the genome-wide association analysis

After processing the raw Illumina ASA chip data of 194 patients with BC, we conducted a systematic quality

TABLE 1 Patient characteristics.

Variable	Screening stage ($n = 194$, %)	Validation stage ($n = 854$)	p value
Age, median (inner distance)	46 (40–51)	47.5 (41–55)	0.013
Laterality			0.925
Left	100 (51.55)	437 (51.17)	
Right	94 (48.45)	417 (48.83)	
Pathological type			0.063
Infiltrating ductal carcinoma	170 (87.63)	701 (82.08)	
Others	24 (12.37)	153 (17.92)	
Histological grading			0.475
Grades I–II	90 (46.39)	384 (44.96)	
Grade III	71 (36.60)	292 (34.19)	
Unknown	33 (17.01)	178 (20.84)	
Breast subtype			0.002
Luminal A	47 (24.23)	206 (24.12)	
Luminal B HER2-negative	51 (26.29)	252 (29.51)	
Luminal B HER2-positive	15 (7.73)	122 (14.29)	
HER2+	39 (20.10)	102 (11.94)	
Triple-negative	38 (19.59)	128 (14.99)	
Unknown	4 (2.06)	44 (5.15)	
TNM staging			0.956
Phase I	45 (23.20)	206 (24.12)	
Phase II	81 (41.75)	356 (41.69)	
Phase III	68 (35.05)	292 (34.19)	
Family history			0.355
Yes	52 (26.80)	202 (23.65)	
No	142 (73.20)	652 (76.35)	

Note: Tumor-node-metastasis (TNM) staging was defined according to the *2017 American Joint Commission for Cancer (AJCC)* (8th edition) guidelines; Molecular staging was defined according to the 2021 NCCN guidelines.

Abbreviation: HER2, human epidermal growth factor receptor 2.

control assessment of the genotyping data, which confirmed that the results passed the quality control criteria. Population stratification issues in the samples were identified and adjusted using principal component analysis. After removing outliers, the genetic variation between the samples was found to be minimal, indicating a good level of consistency (Figure 2a–c). Subsequently, an association analysis was performed on the 510,152 high-quality genotyped loci. An additive model was used to investigate the association between the genetic loci and the recurrence and metastasis of BC. Covariates such as patient age, hormone receptor status, HER-2 status, and TNM stage were adjusted using logistic regression to identify SNPs associated with BC recurrence and metastasis. A Manhattan plot was used to illustrate the distribution of genetic variants associated with BC recurrence and metastasis across chromosomes and the whole genome (Figure 2d). Furthermore, the Quantile–Quantile (*Q*–*Q*) plot was used to demonstrate the differences between the observed and predicted values of the genetic variants (Figure 2e).

In this study, 20,132 SNP sites associated with BC recurrence and metastasis were identified based on the additive model ($p < 0.05$); of these, 3368 SNP sites had a $p < 0.01$ and 191 SNP sites had a $p < 0.001$. The most significant SNP site had a p value of 1.88×10^{-6}; however, no SNP sites in this study met the significance threshold of the statistical analysis ($p < 5.00 \times 10^{-8}$). Based on the results of the GWAS analysis, we followed the SNP screening strategy described in the Methods section. This resulted in the identification of 21 SNP loci (Tables 2 and 3), which could be subjected to subsequent validation.

3.3 | Association between the genotypes of rs10108514, rs12920540, rs4273077, and rs4730155 and the DFS of patients with BC

The genotypes of the 21 SNP loci mentioned above and their corresponding recurrence and metastasis events were examined using the log-rank test in the verified population. The survival curves revealed that, among the different genotypes of the four significant SNPs loci ($p < 0.05$), only rs10108514 (A>G), rs12920540 (C>A), rs4273077 (A>G), and rs4730155 (T>C) significantly impacted DFS (Figure 3a–d). The identified SNPs were stratified to evaluate their specific impact on tumor metastasis or recurrence. However, no significant SNPs associated with tumor metastasis were observed. Nonetheless, we did identify three SNPs, namely, rs1799801 (T>C), rs2227251 (C>A), and rs4730155 (T>C), which were significantly associated with tumor recurrence (Supporting Information S1: Figure S1).

Subsequent Cox regression analysis further delineated the differences in the risk of recurrent metastasis across these genotypes (Figure 3e). Compared with homozygous AA, carriers of rs10108514 (AG/GG) had a 27% reduced risk of BC recurrence and metastasis (HR = 0.73, 95% CI = 0.54–0.98), consistent with the GWAS result (OR = 2.825, 95% CI = 1.667–4.788) for the

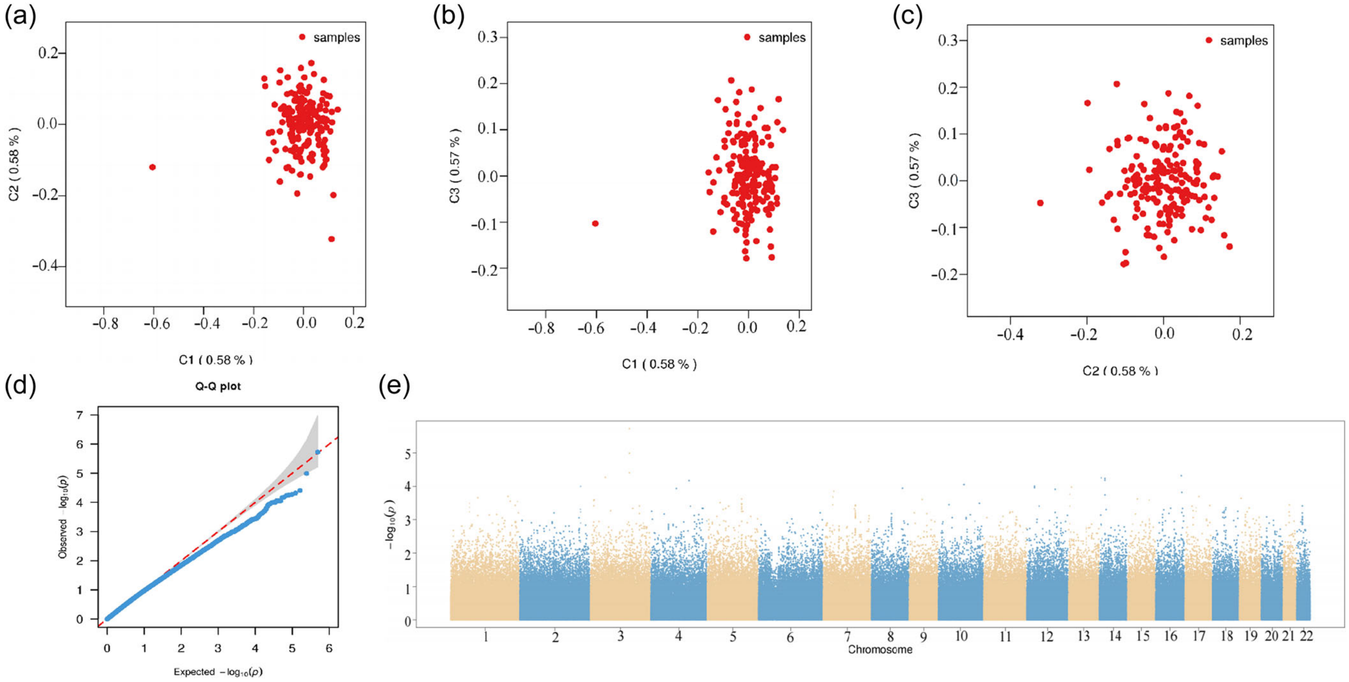

FIGURE 2 Population hierarchical assessment and GWAS visualizations for breast cancer (BC) recurrence and metastasis. (a–c) Sample population hierarchical assessment chart. (d) Manhattan plot for the GWAS of breast cancer recurrence and metastasis. (e) *Q–Q* map for the GWAS of BC recurrence and metastasis. GWAS, Genome-Wide Association Study.

A allele being a risk allele. Compared with homozygous CC, carriers of rs12920540 (CA/AA) had a 52% increased risk of BC recurrence and metastasis (HR = 1.52, 95% CI = 1.06–2.18), consistent with the GWAS result (OR = 2.472, 95% CI = 1.469–4.159) for the C allele being a risk allele. Carriers of rs4273077 (AG/GG) had a 26% reduced risk of BC recurrence and metastasis compared with homozygous AA (HR = 0.74, 95% CI = 0.54–1.01), consistent with the GWAS result for the A allele being a risk allele (OR = 2.788, 95% CI = 1.56–4.983). Conversely, carriers of rs4730155 (CT/TT) had a 35% reduced risk of BC recurrence and metastasis compared with homozygous CC (HR = 0.65, 95% CI = 0.45–0.93), consistent with the GWAS result for the T allele being a protective allele (OR = 0.182, 95% CI = 0.067–0.491). Of note, all four SNPs are located in intronic regions of the gene.

3.4 | 3D SNP spatial role information

We used the 3D SNP database (http://cbportal.org/3dsnp) to determine the distribution and spatial interactions of the four SNPs (rs10108514, rs12920540, rs4273077, and rs4730155) in the genome. Additionally, Circos plots were used to visually represent the chromosomal interactions among the noncoding mutations, distal regulatory elements, and promoters. The expression quantitative trait loci analysis revealed that rs10108514 was significantly associated with plasmacytoma variant translocation 1 (*PVT1*) expression in the thyroid. The rs10108514 is located in the intron region of the *PVT1* gene (on chromosome 8q24) [16], which encodes a long noncoding RNA [16, 17]. The functional annotation of the rs10108514 mutation revealed its primary involvement in three functional categories: transcription factor binding sites (TFBS, score: 64.66), enhancers (score: 50.57), and promoters (score: 20.95). Specifically, in pancreatic PANC-1 cells, rs10108514 resided at the binding site of *POLR2A* with high DNA accessibility (852/1000). In addition, it resided at the binding sites of *EP300*, *ESR1*, and *FOXA1* with high DNA accessibility (265/1000) in BC T-47D cells. In vascular tissue HUVEC cells, the binding sites for JUN, MYC, and *FOXA1* exhibited high DNA accessibility (223/1000) (Figure 4a). The total score of the rs12920540 mutation was derived from its enhancer (0.25) function. The gene located 2 kb upstream or downstream of the SNP was identified as cancer susceptibility candidate 16 (*CASC16*) (Figure 4b). The total score of the rs4273077 mutation was derived from its role as a TFBS (17.49) and an enhancer (5.00). The gene located 2 kb upstream or downstream of rs4273077 was identified as tumor necrosis factor receptor superfamily member 13B (*TNFRSF13B*). In the TFBS fraction, rs4273077 was located at the binding site of *CHD2, CTCF, RAD21, YY1*, and *ZNF143* in H1-hESC cells of esophageal squamous epithelial tissue with DNA accessibility (310/1000) (Figure 4c). Finally, the rs4730155 mutation predominantly scored in the promoter category (100.00) and TFBS (40.93).

TABLE 2 Primer sequences for the 21 SNPs identified in this study.

Loci	Sequence 1 (5′–3′)	Sequence 2 (5′–3′)	Amplification length (bp)
rs1003533	TGGATGGAGAGTTGAATGCCAGCCAC	TGGATGGAACACATGTGATTTACACC	139
rs10108514	TGGATGATTAGCTTGAGTGCCTGGTG	TGGATGGACAGTGAGGGCCAGATTC	119
rs1043996	TGGATGGGGTCACAGTCATTGATGTC	TGGATGATGGTATCTGCACCAACCTG	117
rs10752609	TGGATGGCAGCTTCAGTTGTCTGGTG	TGGATGTGCAGTGTTCAGTCCCTCAG	125
rs1109866	TGGATGTTCTTCTTCACGCTCGTGCC	TGGATGCCAAGACAGCGAATCAGCAC	134
rs111770568	TGGATGCTCGACAGAATGGTACCATC	TGGATGCAAGAGCAGTCCAAGAGTAT	121
rs1164760	TGGATGAAAACCTCAGCGAATGCACC	TGGATGTCTAGGTCTTTCCACTGTCC	96
rs12920540	TGGATGCGCTTCCATAAAGTGGTGAG	TGGATGAGGGATTATATGGCCTGCTC	110
rs1799801	TGGATGTTATACTTCTCTGACTCGGG	TGGATGGAGCTGAAACAAAGCAAGCC	112
rs2227251	TGGATGTGTGTCTGCCGGGATGATG	TGGATGTGAAGACTGAGTCCAGGTGC	153
rs249820	TGGATGAAACCAGCTCCCTTAAGTGC	TGGATGCTTCATGGTGGGCATTAAGG	91
rs2561530	TGGATGGAATTAGTAACCTGGCTGTC	TGGATGAGTTGCCCACCAGCAATTTC	100
rs3746069	TGGATGTGCTTCAGACACTGCCGTAG	TGGATGACAGGGACACAGCACCCAA	113
rs3814811	TGGATGCCTGCCCCCTTAAAACAGAG	TGGATGGGATCTTGATGCCAAGTGTG	107
rs4273077	TGGATGTTCACAGTGCCAGCGGATTC	TGGATGTCACCATGGCTACAGGTTTC	116
rs4730155	TGGATGTTGAAATCGAGCCAAGATCC	TGGATGTCATACTTGATAGTGTTCGG	105
rs74860409	TGGATGTTATCCACTCCCATTTCAAG	TGGATGCTTTAGTCTCCCCACCATTC	111
rs7719624	TGGATGGTTGGCTTCTGAGTTCCATC	TGGATGTCCTCAATGAGATCCTGGTG	100
rs8736	TGGATGTCCCATGGCTTCCATCTGAG	TGGATGGATTTACACACGGTGACCTG	114
rs921943	TGGATGCCCGTGTAGAGCTTCTAACT	TGGATGTCTTCCAAAAGAGACCCTGC	128
rs9830253	TGGATGAAAGCAGCTGTTAACCTCCG	TGGATGGGGTGGGATGCTATCTTTTC	118

Note: The amplification length encompasses the sequence tagged and the longest SNP sequence.

Abbreviation: SNP, single-nucleotide polymorphism.

The gene located 2 kb upstream or downstream of this SNP was identified as nicotinamide phosphoribosyl transferase (*NAMPT*). In the expression quantitative trait loci fraction, rs4730155 was significantly associated with the expression level of *NAMPT* in cells that have been transformed into fibroblasts and left ventricular cells. In the TFBS fraction, rs4730155 was located at the binding site of *POLR2A* in esophageal squamous epithelial tissue H1-hESC cells with high DNA accessibility (652/1000) (Figure 4d). Overall, these findings provide valuable insights into the functional roles of these mutations and their association with gene expression and DNA binding.

3.5 | Analysis of differential gene expression

To understand the expression of the four SNPs in breast tumor tissues compared with normal or paracancerous tissues, the following SNPs were investigated: *PVT1*: rs10108514 (A>G), *CASC16*: rs12920540 (C>A), *TNFRSF13B*: rs4273077 (A>G), and *NAMPT*: rs4730155 (T>C). For this study, we obtained paired breast invasive carcinoma and nearby tissue samples from the BRCA project in TCGA. We also acquired normal tissue data from the GTEx database for comparison. The results showed a significant increase in the expression of the *PVT1* gene corresponding to rs10108514 in breast tumor tissues when compared with normal tissues and paired adjacent tissues ($p < 0.001$) (Figure 5a). Similarly, the expression of the *CASC16* gene corresponding to rs12920540 was substantially increased in breast tumor tissues compared with normal tissues and paired adjacent tissues ($p < 0.001$) (Figure 5b). However, the expression of the *TNFRSF13B* gene corresponding to rs4273077 was not significantly different among BC tumor tissues, normal tissues, or paired adjacent

TABLE 3 Basic information for the 21 verified SNPs.

Gene	SNP ID	Chromosome location	Alleles gene	MAF-EAS	OR (95% CI)
IRF1-AS1	rs1003533	chr5:131755650	C>T	0.331	2.459 (1.484–4.075)
PVT1	rs10108514	chr8:128822708	A>G	0.309	2.825 (1.667–4.788)
NOTCH3	rsl043996	chr19:15295133	G>A	0.593	2.549 (1.483–4.379)
KCNN3	rsl0752609	chrl:154791127	G>A	0.800	3.429 (1.679–7.004)
ABCB6	rsl109866	chr2:220083279	C>T	0.138	0.208 (0.094–0.459)
POLN	rsll1770568	chr4:2167331	G>A	0.083	3.426 (1.702–6.897)
CHST11	rs1164760	chr12:105063253	C>T	0.280	0.340 (0.181–0.642)
CASC16	rs12920540	chr16:52625612	C>A	0.430	2.472 (1.469–4.159)
ERCC4	rsl799801	chr16:14041957	T>C	0.259	0.334 (0.175–0.637)
CTDSP1	rs2227251	chr2:219266422	C>A	0.071	4.09 (1.853–9.030)
LINCO2453	rs249820	chr12:98897472	A>G	0.105	10.47 (3.157–34.690)
COQ8B, ITPKC	rs2561530	chr19:41224313	G>A	0.714	2.516 (1.456–4.346)
GNA11	rs3746069	chr19:3114863	C>T	0.252	3.309 (1.751–6.253)
NYNRIN	rs3814811	chr14:24887461	T>C	0.784	0.207 (0.096–0.445)
TNFRSF13B	rs4273077	chr17:16849138	A>G	0.479	2.788 (1.560–4.983)
NAMPT	rs4730155	chr7:105924142	T>C	0.876	0.182 (0.067–0.491)
ZBTB38	rs74860409	chr3:141144140	C>T	0.082	5.194 (2.066–13.05)
TGFBI	rs7719624	chr5:135377565	C>T	0.581	2.593 (1.509–4.457)
MBOAT7, TMC4	rs8736	chr19:54677188	T>C	0.218	2.894 (1.550–5.404)
DMGDH	rs921943	chr5:78316475	C>T	0.139	0.288 (0.145–0.573)
COL6A6	rs9830253	chr3:130284283	G>A	0.444	3.186 (1.834–5.535)

Abbreviations: CI, confidence interval; EAS, East Asian Population; MAF, minor allele frequency; OR, odds ratio; SNP, single-nucleotide polymorphism.

tissues ($p \geq 0.05$) (Figure 5c). Finally, the expression of the *NAMPT* gene corresponding to rs4730155 was significantly higher in breast tumor tissues compared with normal tissues and paired adjacent tissues ($p < 0.05$) (Figure 5d). This study demonstrated distinct gene expression patterns in breast tumor tissues compared with normal and paracancerous tissues for the investigated SNPs.

3.6 | Significant association between high *TNFRSF13B* expression and BC prognosis

To investigate the association between the genes corresponding to the four identified SNP loci and BC prognosis, we next obtained survival and prognosis information from TCGA. We categorized gene expression into high and low based on a 50% cutoff. The results revealed that high *TNFRSF13B* expression was significantly associated with a more favorable prognosis, as indicated by the prolonged overall survival and progress-free interval of BC patients ($p < 0.05$). Moreover, patients with high *TNFRSF13B* expression tended to have a longer disease-specific survival than those with low *TNFRSF13B* expression ($p = 0.077$). Conversely, *PVT1*, *CASC16*, and *NAMPT* expression did not significantly impact BC prognosis (Figure 6a–c).

3.7 | GO/KEGG enrichment analysis

Using the online DAVID tool (https://david.ncifcrf.gov/), we performed gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses on genes corresponding to SNP loci substantially associated with BC recurrence and metastasis ($p < 0.001$). The results revealed that the enriched target genes were involved in biological processes such as glutamatergic synaptic transmission and KEGG signaling pathways such as the calcium signaling system and the insulin secretion route (Figure 6d).

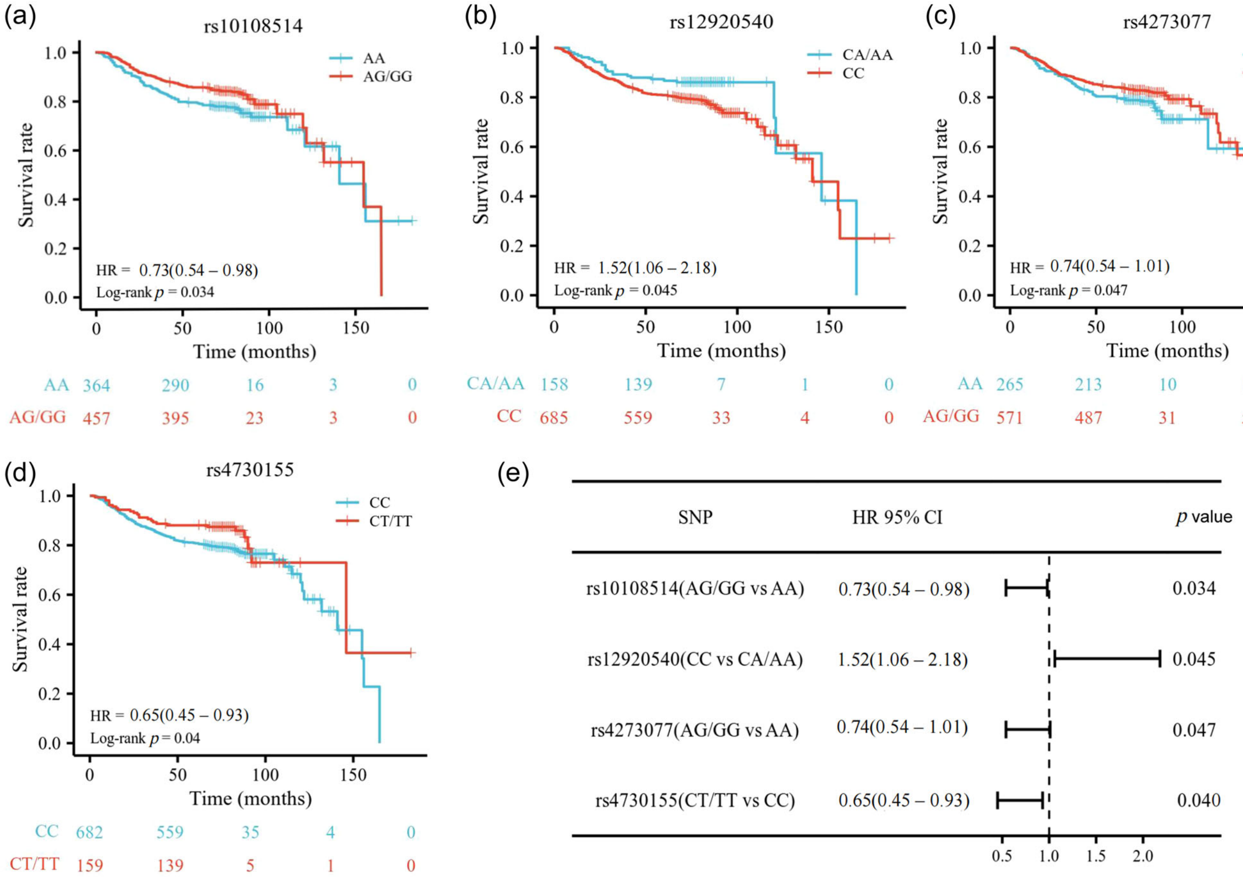

FIGURE 3 Log-rank and Cox regression tests of differential DFS in breast cancer (BC) patients by genotype for four SNPs. Log-rank test of DFS in BC patients with different genotypes of rs10108514 (a), rs12920540 (b), rs4273077 (c), or rs4730155 (d). (e) The Cox regression test was used to compare the contributions of the different genotypes of rs10108514, rs12920540, rs4273077, or rs4730155 to the risk of BC recurrence and metastasis. CI, confidence interval; DFS, disease-free survival; SNP, single-nucleotide polymorphism.

4 | DISCUSSION

BC is the most common malignant tumor affecting women worldwide [1]. The long-term survival of BC patients is hampered by a high risk of recurrence and metastasis. Thus, accurately predicting the risk of recurrence and metastasis for BC is crucial [18]. BC arises due to a series of genetic, epigenetic, and phenotypic changes. Genetic polymorphisms associated with BC risk have been identified in genes involved in multiple biological pathways [19]. These genetic polymorphisms further contribute to variations in disease susceptibility and severity among individuals [20]. In this study, we conducted unbiased screening of candidate SNPs through GWAS analysis and validated them using a larger clinical cohort. During the validation phase, four SNP loci, namely, *PVT1*: rs10108514 (A>G), *CASC16*: rs12920540 (C>A), *TNFRSF13B*: rs4273077 (A>G), and *NAMPT*: rs4730155 (T>C), were significantly associated with the risk of BC recurrence and metastasis. The 3D SNP analysis provides a new perspective on the spatial and functional relationships among SNPs within the genome. This method enabled us to determine the potential regulatory roles of candidate SNPs in gene expression and cellular processes associated with BC progression by identifying specific DNA binding sites and transcription factor interactions for the four SNPs.

This study revealed a significant association between the rs10108514 (A>G) mutation in *PVT1* and an increased risk of BC recurrence and metastasis, confirming the role of this gene in BC progression [21]. Previous research has shown that SNP variants of *PVT1* can induce functional changes that have been linked to the development of various malignant tumors, including BC [22]. The second SNP identified, rs12920540, was located in the intron region of the *CASC16* (*LOC643714*) gene. Our findings also align with previous studies indicating the significant association between this gene and susceptibility to BC. The genomic region of TOX3/LOC643714 has been extensively studied and linked to an increased risk of BC [23]. Recent research has shown that genetic variants (including the common rs3803662 and rs4784227 genetic

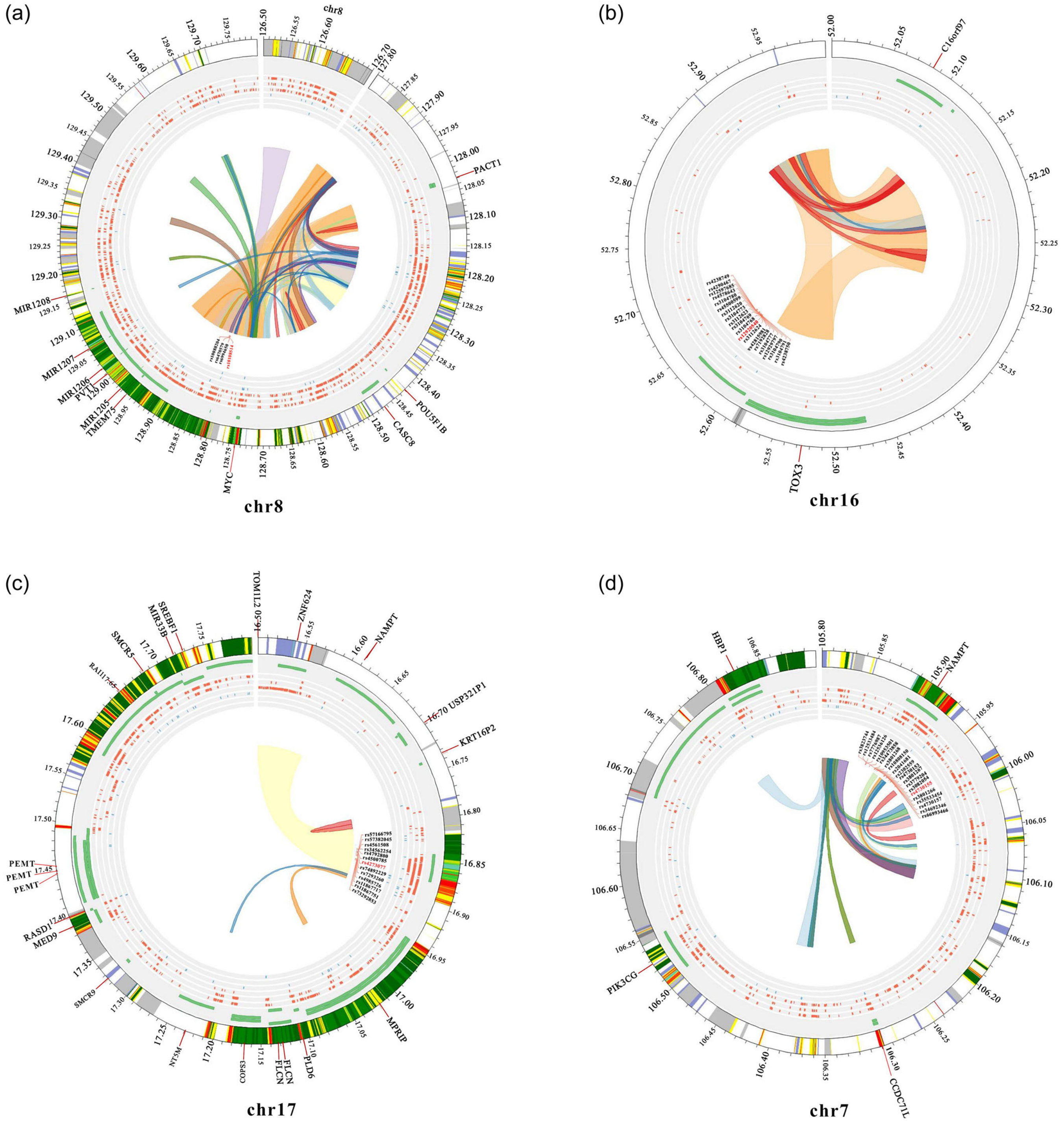

FIGURE 4 Circos diagrams visualizing chromatin interactions and SNP annotations for four SNPs. Circos diagrams for rs10108514 (a), rs12920540 (b), rs4273077 (c), and rs4730155 (d). Chromatin, annotated genes, histones (depicted in red), transcription factors (depicted in blue), current and associated SNPs, and 3D chromatin interactions (in that order) can be seen moving from the outer to the inner rings of the Circos diagram. SNP, single-nucleotide polymorphism.

variants) at multiple loci on TOX3/LOC643714 are associated with an increased risk of BC [24].

Here, we found a correlation between the rs4273077, located in the intronic region of the *TNFRSF13B* gene, and the risk of BC recurrence and metastasis. Moreover, we identified a significant association between high *TNFRSF13B* expression and a favorable BC prognosis. In a previous report, *TNFRSF13B* was proposed as a predictive marker for BC progression and a potential therapeutic target for triple-negative BC [25]. Thus, our

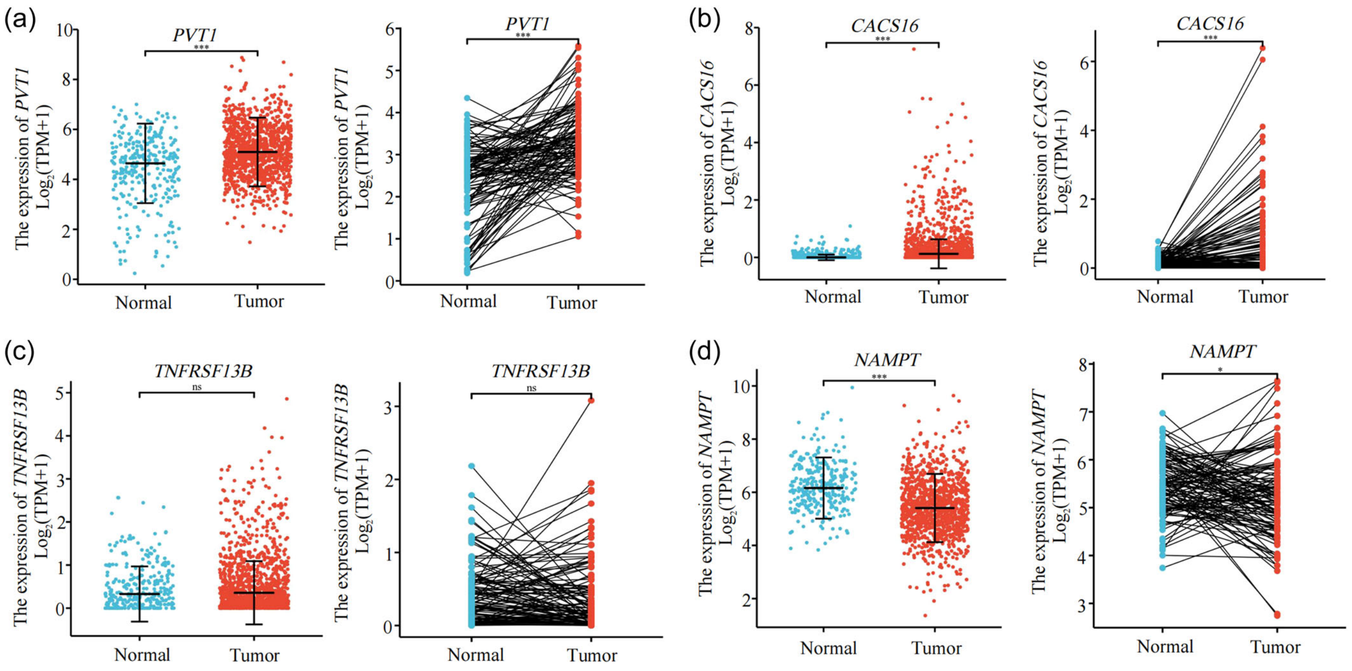

FIGURE 5 Differential expression of *PVT1* (a), *CASC16* (b), *TNFRSF13B* (c), and *NAMPT* (d) genes. ***, $p < 0.001$; *, $p < 0.05$; ns, $p \geq 0.05$.

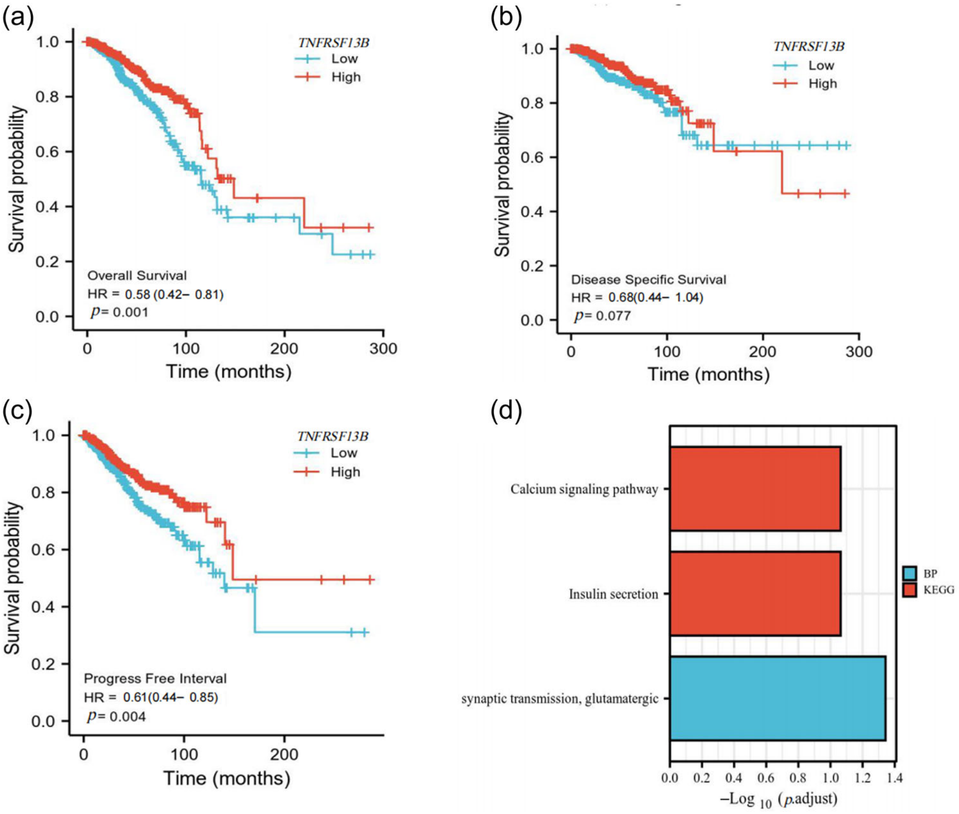

FIGURE 6 Functional enrichment analysis and survival analysis of differential expression of the *TNFRSF13B* gene. (a) OS log-rank test for the differential expression of the *TNFRSF13B* gene. (b) DSS log-rank test for the differential expression of the *TNFRSF13B* gene. (c) PFI log-rank test for the differential expression of the *TNFRSF13B* gene. (d) GO/KEGG enrichment analysis of genes corresponding to differential SNPs in GWAS analysis results. DSS, disease-specific survival; GO, gene ontology; HR, hazard ratio; KEGG, Kyoto Encyclopedia of Genes and Genomes; OS, overall survival; PFI, progress-free interval; SNP, single-nucleotide polymorphism.

findings further support the potential role of *TNFRSF13B* in BC progression. However, the underlying mechanisms and functional implications of the rs4273077 genetic variant in the development and metastasis of BC remain unclear. In addition, we discovered a significant association between the rs4730155 locus in the intronic region of *NAMPT* and susceptibility to BC recurrence and metastasis. The functional implications of *NAMPT* in tumor biology are well-established, including its involvement in DNA repair, metastasis, angiogenesis, immune regulation, and drug resistance [26]. Nevertheless, there is a lack of research examining the relationship between *NAMPT* polymorphisms and the development and progression of BC. To the best of our knowledge, the present study was the first to establish a connection between the rs4730155 locus in *NAMPT* and the susceptibility to BC recurrence and metastasis. Further research should clarify how the rs4730155 locus affects *NAMPT* function and BC progression.

Our GWAS identified genes significantly associated with BC recurrence and metastasis based on the presence of specific SNPs. The GO and KEGG enrichment analyses revealed that the genes targeted by these SNPs were involved in biological processes related to glutamatergic synaptic transmission and the calcium and insulin signaling pathways. These findings are consistent with previous research, emphasizing the importance of these pathways in BC metastasis. Glutamatergic synaptic transmission is crucial for normal brain development and function [27]. Moreover, multiple studies have demonstrated that BC brain metastasis is caused by the formation of pseudo-triple synapses between cancer cells and glutamatergic neurons [28]. The calcium signaling pathway regulates fundamental cellular processes such as cell proliferation, survival, apoptosis, and immunity; it is also associated with the development of various diseases, including cancer and autoimmune disorders [29]. Similarly, components of the insulin signaling pathway, such as PI3K and MAPK, are pivotal in the regulation of cell survival, growth, proliferation, and differentiation. Thus, the disruption of the insulin signaling pathway is also associated with cancer progression and metastasis [30, 31]. Hence, the present study provides novel evidence for the involvement of glutamatergic synaptic transmission, calcium signaling, and insulin signaling pathways in BC recurrence and metastasis. Understanding the mechanisms of these processes may have important implications for the development of targeted therapies and intervention strategies for BC.

This study identified four risk SNPs (rs10108514, rs12920540, rs4273077, and rs4730155) using GWAS, thus, providing valuable insights into the genetic variants associated with BC recurrence and metastasis. Our research lays the groundwork for further investigation and offers direction for future studies. However, there are several limitations to this study, including the small sample size, which may have produced false-negative or false-positive results. Moreover, genetic variations among different individuals may affect our predictions of SNP function. Environmental and genetic factors interact to influence an individual's phenotype. Given that our study focused solely on genetic factors while ignoring environmental influences, it may offer a simplistic view of SNP function in BC. Additionally, we acknowledge that we determined the functions of the identified SNPs through 3D SNP analysis alone, without subsequent experimental validation. In future studies, we will address these limitations by increasing the sample size, considering the role of genomic heterogeneity, and accounting for the contribution of environmental factors. We believe that this approach will help us better understand the relationship between human genetics and disease.

5 | CONCLUSION

In this study, we used a two-stage GWAS approach to demonstrate that four SNPs loci, namely, *PVT1*: rs10108514 (A>G), *CASC16*: rs12920540 (C>A), *TNFRSF13B*: rs4273077 (A>G), and *NAMPT*: rs4730155 (T>C), were associated with the risk of BC recurrence and metastasis. TCGA data were used to reveal a substantial correlation between *TNFRSF13B* expression and BC prognosis. The enrichment analysis of GWAS results revealed that the calcium signaling and insulin secretion pathways were potentially significant signaling pathways in BC recurrence and metastasis. Our study provides insights into the genetic basis of BC recurrence and metastasis, which may advance future research efforts and facilitate the development of new treatment strategies.

AUTHOR CONTRIBUTIONS

Shujuan Sun: Writing—original draft (equal). **Sha Yin**: Writing—original draft (equal). **Jie Huang**: Formal analysis (equal); writing—review and editing (equal). **Dongdong Zhou**: Formal analysis (equal). **Qiaorui Tan**: Data curation (equal). **Xiaochu Man**: Data curation (equal). **Wen Wang**: Data curation (equal). **Jiale Zhang**: Data curation (equal). **Huihui Li**: Conceptualization (equal).

ACKNOWLEDGMENTS

The authors are grateful to all our team members and the funding bodies who financially supported this study.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

There are no restrictions on data availability.

ETHICS STATEMENT

This study was approved by the Ethics Committee of Shandong Cancer Hospital and Institute (SDTHEC2022009018).

INFORMED CONSENT

All patients provided written informed consent at the time of entering this study.

ORCID

Shujuan Sun http://orcid.org/0000-0003-4160-7784

REFERENCES

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin. 2024;74(1):12–49. https://doi.org/10.3322/caac.21820
2. Malmgren JA, Mayer M, Atwood MK, Kaplan HG. Differential presentation and survival of *de novo* and recurrent metastatic breast cancer over time: 1990-2010. Breast Cancer Res Treat. 2018;167(2):579–90. https://doi.org/10.1007/s10549-017-4529-5
3. Mariotto AB, Etzioni R, Hurlbert M, Penberthy L, Mayer M. Estimation of the number of women living with metastatic breast cancer in the United States. Cancer Epidemiol Biomark Prev. 2017;26(6):809–15. https://doi.org/10.1158/1055-9965.EPI-16-0889
4. Gennari A, Conte P, Rosso R, Orlandini C, Bruzzi P. Survival of metastatic breast carcinoma patients over a 20-year period: a retrospective analysis based on individual patient data from six consecutive studies. Cancer. 2005;104(8):1742–50. https://doi.org/10.1002/cncr.21359
5. Yan J, Liu Z, Du S, Li J, Ma L, Li L. Diagnosis and treatment of breast cancer in the precision medicine era. Methods Mol Biol. 2020;2204:53–61. https://doi.org/10.1007/978-1-0716-0904-0_5
6. Si H, Esquivel M, Mendoza Mendoza E, Roarty K. The covert symphony: cellular and molecular accomplices in breast cancer metastasis. Front Cell Dev Biol. 2023;11:1221784. https://doi.org/10.3389/fcell.2023.1221784
7. Kashyap D, Pal D, Sharma R, Garg VK, Goel N, Koundal D, et al. Global increase in breast cancer incidence: risk factors and preventive measures. BioMed Res Int. 2022;2022:9605439. https://doi.org/10.1155/2022/9605439
8. Smolarz B, Nowak AZ, Romanowicz H. Breast cancer-epidemiology, classification, pathogenesis and treatment (review of literature). Cancers. 2022;14(10):2569. https://doi.org/10.3390/cancers14102569
9. Pérez-Losada J, Castellanos-Martín A, Mao JH. Cancer evolution and individual susceptibility. Integr Biol. 2011;3(4):316–28. https://doi.org/10.1039/C0IB00094A
10. Chen Y, Shi C, Guo Q. TNRC9 rs12443621 and FGFR2 rs2981582 polymorphisms and breast cancer risk. World J Surg Oncol. 2016;14(1):50. https://doi.org/10.1186/s12957-016-0795-7
11. Gudmundsdottir ET, Barkardottir RB, Arason A, Gunnarsson H, Amundadottir LT, Agnarsson BA, et al. The risk allele of SNP rs3803662 and the mRNA level of its closest genes *TOX3* and LOC643714 predict adverse outcome for breast cancer patients. BMC Cancer. 2012;12:621. https://doi.org/10.1186/1471-2407-12-621
12. Bouhniz OE, Zaied S, Naija L, Bettaieb I, Rahal K, Driss M, et al. Association between HER2 and IL-6 genes polymorphisms and clinicopathological characteristics of breast cancer: significant role of genetic variability in specific breast cancer subtype. Clin Exp Med. 2020;20(3):427–36. https://doi.org/10.1007/s10238-020-00632-5
13. Miedl H, Oswald D, Haslinger I, Gstoettner M, Wenzl R, Proestling K, et al. Association of the estrogen receptor 1 polymorphisms rs2046210 and rs9383590 with the risk, age at onset and prognosis of breast cancer. Cells. 2023;12(4):515. https://doi.org/10.3390/cells12040515
14. Cui P, Zhao Y, Chu X, He N, Zheng H, Han J, et al. SNP rs2071095 in LincRNA H19 is associated with breast cancer risk. Breast Cancer Res Treat. 2018;171(1):161–71. https://doi.org/10.1007/s10549-018-4814-y
15. Couzin J, Kaiser J. Closing the net on common disease genes. Science. 2007;316(5826):820–2. https://doi.org/10.1126/science.316.5826.820
16. Shtivelman E, Henglein B, Groitl P, Lipp M, Bishop JM. Identification of a human transcription unit affected by the variant chromosomal translocations 2;8 and 8;22 of Burkitt lymphoma. Proc Natl Acad Sci USA. 1989;86(9):3257–60. https://doi.org/10.1073/pnas.86.9.3257
17. Graham M, Adams JM. Chromosome 8 breakpoint far 3′ of the c-myc oncogene in a Burkitt's lymphoma 2;8 variant translocation is equivalent to the murine pvt-1 locus. EMBO J. 1986;5(11):2845–51. https://doi.org/10.1002/j.1460-2075.1986.tb04578.x
18. Kim MY. Breast Cancer Metastasis. Exp Med Biol. 2021;1187:183–204. https://doi.org/10.1007/978-981-32-9620-6_9
19. Coughlin SS. Epidemiology of breast cancer in women. Adv Exp Med Biol. 2019;1152:9–29. https://doi.org/10.1007/978-3-030-20301-6_2
20. Malins DC, Haimanot R. Major alterations in the nucleotide structure of DNA in cancer of the female breast. Cancer Res. 1991;51(19):5430–2.
21. Zhang Z, Zhu Z, Zhang B, Li W, Li X, Wu X, et al. Frequent mutation of rs13281615 and its association with PVT1 expression and cell proliferation in breast cancer. J Genet Genomics. 2014;41(4):187–95. https://doi.org/10.1016/j.jgg.2014.03.006
22. Lin HY, Callan CY, Fang Z, Tung HY, Park JY. Interactions of *PVT1* and *CASC11* on prostate cancer risk in African Americans. Cancer Epidemiol Biomark Prev. 2019;28(6):1067–75. https://doi.org/10.1158/1055-9965
23. Xu W, Zhong Y, Yang H, Gong Y, Dao J, Bao L. Association between the rs4784227-CASC16 polymorphism and the risk of breast cancer: a meta-analysis. Medicine. 2022;101(34):e30218. https://doi.org/10.1097/MD.0000000000030218
24. Zuo X, Wang H, Mi Y, Zhang Y, Wang X, Yang Y, et al. The association of CASC16 variants with breast cancer risk in a northwest Chinese female population. Mol Med. 2020;26(1):11. https://doi.org/10.1186/s10020-020-0137-7

25. Hinterleitner C, Zhou Y, Tandler C, Heitmann JS, Kropp KN, Hinterleitner M, et al. Platelet-expressed TNFRSF13B (TACI) predicts breast cancer progression. Front Oncol. 2021;11:642170. https://doi.org/10.3389/fonc.2021.642170
26. Gasparrini M, Audrito V. NAMPT: a critical driver and therapeutic target for cancer. Int J Biochem Cell Biol. 2022;145:106189. https://doi.org/10.1016/j.biocel.2022.106189
27. Martynyuk AE, Glushakov AV, Sumners C, Laipis PJ, Dennis DM, Seubert CN. Impaired glutamatergic synaptic transmission in the PKU brain. Mol Gen Metab. 2005;86(Suppl 1):34–42. https://doi.org/10.1016/j.ymgme.2005.06.014
28. Zeng Q, Michael IP, Zhang P, Saghafinia S, Knott G, Jiao W, et al. Synaptic proximity enables NMDAR signalling to promote brain metastasis. Nature. 2019;573(7775):526–31. https://doi.org/10.1038/s41586-019-1576-6
29. Panda S, Chatterjee O, Roy L, Chatterjee S. Targeting Ca^{2+} signaling: a new arsenal against cancer. Drug Discov Today. 2022;27(3):923–34. https://doi.org/10.1016/j.drudis.2021.11.012
30. Yee LD, Mortimer JE, Natarajan R, Dietze EC, Seewaldt VL. Metabolic health, insulin, and breast cancer: why oncologists should care about insulin. Front Endocrinol. 2020;11:58. https://doi.org/10.3389/fendo.2020.00058
31. Haeusler RA, McGraw TE, Accili D. Biochemical and cellular properties of insulin receptor signalling. Nat Rev Mol Cell Biol. 2018;19(1):31–44. https://doi.org/10.1038/nrm.2017.89

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: Sun S, Yin S, Huang J, Zhou D, Tan Q, Man X, et al. Identification of significant single-nucleotide polymorphisms associated with breast cancer recurrence and metastasis using GWAS. Cancer Innov. 2025;4:e142. https://doi.org/10.1002/cai2.142