Arch Biol Sci. 2023;75(1):57-67 https://doi.org/10.2298/ABS230115005M

Association of rs4646903 and rs1048943 CYP1A1 estrogen-metabolizing gene polymorphisms with estrogen receptor-positive breast cancer in Kenyan women

Mary Murithi[1,2,*], Steven Nyanjom[1], Victor Mobegi[3], Sayed Shahin[4] and Francis Makokha[5]

[1]*Biochemistry Department, Jomo Kenyatta University of Agriculture and Technology*, 62000-00200, Nairobi, Kenya
[2]*Pre-Clinical Department, Kabarak University*, Private Bag-20157, Nakuru, Kenya
[3]*Department of Biochemistry, University of Nairobi, 30197-00100*, Nairobi, Kenya
[4]*Department of Pathology, Aga Khan University Hospital*, 30270-00100, Nairobi, Kenya
[5]*Directorate of Research and Innovation, Mount Kenya University*, 342-01000, Thika, Kenya

Corresponding author: murithimary11@gmail.com

Received: January 15, 2023; **Revised**: January 31, 2023; **Accepted**: February 1, 2023; **Published online**: February 10, 2023

Abstract: Breast cancer is the most prevalent neoplasm and the second leading cause of death among females in Kenya. Estrogen and its metabolites are known risk factors for breast cancer. Polymorphisms in these genes and breast cancer susceptibility are unique among different populations. This study aimed to determine the probable associations between estrogen-metabolizing gene variations and other risk factors for breast cancer risk in Kenyan women. Buffy coat samples were obtained from patients diagnosed with estrogen receptor-positive breast cancer, benign breast disease, and healthy volunteers. Genotyping of target polymorphisms was conducted using polymerase chain reaction (PCR)-restriction fragment length polymorphism (RFLP) analysis. The rs4646903 variant genotype CC was associated with breast cancer in the case-control model (P=0.001); the heterozygous genotype TC (P=0.01) and the luminal B molecular subtype (P=0.02) showed increased odds of late-stage breast cancer. The rs1048943 variant genotype GG was associated with breast cancer in the case-benign model (P=0.04), whereas CG was associated with breast cancer in the case-control model (P=0.02). These findings imply that the rs4646903 and rs1048943 variant genotypes are involved in breast cancer risk in Kenyan women. Hence, they may be explored further as potential markers for the disease.

Keywords: breast cancer; gene polymorphisms; estrogen metabolizing gene; polymerase chain reaction-restriction fragment length polymorphism; genotype

INTRODUCTION

Breast cancer has consistently been recorded as the primary malignancy in females worldwide. According to the GLOBOCAN 2020 report, the disease accounts for about 2.3 million incident cases and 0.7 million deaths. Of these, 186,598 new cases and 85,787 mortalities are estimated to occur in Africa. Africa has maintained a young-age profile for breast cancer patients. Although the continent has a low age-standardized incidence rate (ASR) of 40.7/100,000, it has the highest age-standardized mortality rate of 19.4/100,000 relative to other continents. In Kenya, breast cancer was reported to have an incidence of 6799 (ASR 41/100 000) and mortalities of 3107 (ASR 19.4/100 000) [1]. A recent study reported that 54% of breast cancer cases were diagnosed in patients aged >50 years [2].

Breast cancer is managed and controlled by surgery, radiotherapy and systemic therapies (chemotherapy, targeted therapy and hormonal therapy) [3]. However, access to safe and timely therapy is scarce in sub-Saharan Africa (Kenya included) as challenges of limited specialized oncologists and infrastructure are common [4]. Late disease manifestation, limited knowledge of the disease and deficient healthcare infrastructure have contributed to the low 5-year survival rates of patients with breast cancer in the region [5,6]. Although the risk factors driving susceptibility to breast cancer in low-income countries mirror those in high-income countries, variations in the incidence

How to cite this article: Murithi M, Nyanjom S, Mobegi V, Shahin S, Makokha F. Association of rs4646903 and rs1048943 CYP1A1 estrogen-metabolizing gene polymorphisms with estrogen receptor-positive breast cancer in Kenyan women. Arch Biol Sci. 2023;75(1):57-67.

of risk factors and susceptibility to breast cancer may vary based on genetic, geographical and lifestyle factors [1]. The primary risk factors associated with urbanization and economic development, such as modern birth control methods, obesity, delayed parity, null parity and short breastfeeding periods, have contributed significantly to the occurrence of the disease [5].

Estrogen is implicated in the pathogenesis and prognosis of breast cancer [7]. Estrogen metabolism, which is mediated by heme-containing cytochrome P450 (CYP450) enzymes, generates genotoxic metabolites that play significant roles in estrogen-induced breast carcinogenesis [8]. Estrogen is oxidized by phase I enzymes (*CYP1A1*) to yield 2-hydroxy (catechol) estrogens, which are conjugated by phase II enzymes for elimination. If not conjugated, catechol metabolites are converted into quinones and semiquinones. Collectively, hydroxy estrogens, semiquinones and quinones undergo redox cycling to yield reactive oxygen species (ROS) that are associated with oxidative damage [9]. Genetic variations (single-nucleotide polymorphisms) in the CYPP450 genes can modulate metabolic activities during the transformation of substances. This may either increase or reduce exposure to estrogenic metabolites [10,11]. Allele distribution in CYP450 genes differs between populations, and these genes are of clinical importance as they can be studied to understand disease development and progression [12].

Breast cancer encompasses a vast array of conditions that result from multiple biological processes and risk factors [13]. Variations in the estrogen-metabolizing genes have been associated with breast cancer [14]. Knowledge regarding the role of genetic polymorphisms and other factors in the risk of breast cancer in Kenya is scarce. Consequently, insights into interventions that could be adopted for the prevention, monitoring and early diagnosis of the disease to improve survival are unclear [15]. This study sought to explore the possible associations of various risk factors, namely CYP1A (rs4646903) and CYP1A (rs1048943) genotypes, environmental exposure, reproductive and clinical factors with susceptibility to estrogen receptor-positive (ER+) breast cancer and its characteristics. Knowledge of variant frequencies and their associations with breast cancer will clarify the clinical relevance of these polymorphisms in the etiology and progression of the disease in Kenyan patients. This may lead to the development of interventions that can be used as potential molecular biomarkers for breast cancer diagnosis.

MATERIALS AND METHODS

Ethics statement

The study was performed in compliance with the Declaration of Helsinki; the Aga Khan University Institutional Ethical Research Committee approved the study protocol (Ref: 2020/IERC-26 (v2)). All participants gave informed consent and agreed to participate in the study.

Participants

The study participants were women with breast masses who visited the breast clinic and radiology department at Aga Khan University Hospital Nairobi (AKUHN) and AIC Hospital Kijabe (KAIC). Control individuals were healthy volunteers working at AKUHN. Recruitment took place over three years in the period between 2019 and 2021. A total of 170 subjects, including 68 pathologically confirmed ER+ breast cancer patients, 82 benign breast disease (BBD) patients and 20 healthy volunteers (controls) working at the hospital, were included in the study. The distribution of ER+ breast cancer cases across the study period of 2019, 2020 and 2021 was 14, 51 and 3, respectively, while that of BBD was 14, 33, and 35 in 2019, 2020, and 2021, respectively. Controls were recruited in 2021.

The study participants were interviewed by a trained nurse, and blood samples were collected after obtaining informed consent to participate in the study. Buffy coat aliquots were prepared from the blood samples and stored at -80°C until analysis. Individuals with other cancers were excluded from the study. In addition to the genotype data, other data considered for all participants included sociodemographic (age, level of education), anthropometric (body mass index (BMI)) and reproductive factors (age at menarche, age at menopause, parity), family medical history, comorbidities (diabetes and hypertension), lifestyle (smoking tobacco and alcohol consumption) and exogenous hormonal use (use of contraceptives). For the ER+ breast cancer cases, disease characteristics including

the stage, grade, histological and molecular subtypes were included.

DNA extraction and genotyping

Genomic DNA was extracted from the buffy coat samples using an ISOLATE II Genomic DNA Kit (Bioline, Meridian Life Science Inc., USA) according to the manufacturer's protocol. The quality and quantity of DNA were determined using a NanoDrop 2000 spectrophotometer (BioSpec-mini, Shimadzu Corporation, Tokyo, Japan). A cut-off of 50 ng/μL was used for the inclusion of samples in the PCR. Analysis of *CYP1A1* (rs4646903) and *CYP1A1* (rs1048943) by restriction fragment length polymorphism (PCR-RFLP) was performed using an Applied Biosystems™ ProFlex™ 3× 32-well PCR System thermocycler. The 340 base pairs (bp) of rs4646903 and 204 bp of rs1048943 regions were amplified using specific primers (F: 5'-CAGTGAAGAGGTGTAGCCGCT-3' and R: 5'-TAGGAGTCTTGTCTCATGCCT-3' for rs4646903; F: 5'-CTGTCTCCCTCTGGTTACAGGAAGC-3' and R: 5'-TTCCACCCGTTGCAGCAGGATAGCC-3' for rs1048943). The PCR mixtures consisted of 50 ng/μL genomic DNA, 0.3 μM of each primer, 1.5 U of MyTaq DNA polymerase, 6 μL 5× MyTaq reaction buffer and nuclease-free water to a volume of 30 μL. Amplification consisted of an initial denaturation for 3 min at 95°C, followed by 30 cycles of denaturation for 30 s at 95°C, annealing for 45 s at 55°C, extension for 45 s at 72°C and final extension for 3 min at 72°C. PCR products were visualized on a 2% agarose-Tris borate-EDTA gel stained with GelRed® nucleic acid gel stain (Biotium, Inc. Fremont, CA, USA). Amplified DNA fragments containing the regions of interest were subjected to restriction enzyme digestion. The restriction enzymes *MspI* (New England Biolabs; UK, catalog #: R0106) and *BsrDI* (New England Biolabs; UK, catalog #: R0574) were used as restriction digests. To the 10 μL PCR product, nuclease-free water, reaction buffer and four units of a predetermined enzyme were added to a total volume of 20 μL. The restriction reaction of *Msp1* was incubated at 37°C for 1 h and heat-inactivated at 80°C for 20 min; the *BsrD1* reaction was incubated at 65°C for 1 h and heat-inactivated at 80°C for 20 mins. The restricted products were visualized as band patterns on a 3% agarose-tris borate-EDTA gel stained with gel red under a UV transilluminator.

Genotyping of rs4646903

The wild-genotype TT was detected as one fragment (340 bp), while the variant genotype CC, characterized by the gain of restriction site by *MspI*, was detected as two fragments (200 and 140 bp). The heterozygous genotype TC was detected by the presence of three fragments (340, 200 and 140 bp).

Genotyping of rs1048943

The wild-genotype AA was detected as two fragments (150 and 54 bp), while the variant genotype GG, characterized by loss of cleavage by *BsrDI*, was detected as one fragment (204 bp). The heterozygous genotype AG was detected by the presence of three fragments (204, 150 and 54 bp).

Statistical analysis

Descriptive analyses of the predictors (sociodemographic, anthropometric, reproductive, lifestyle, family history of breast cancer and genotype) for all participants and disease characteristics were performed using univariate analysis. Pearson's chi-square ($\chi 2$) test was used to determine differences in the distribution of predictors among the study participants. Fischer's exact test was used for scenarios with fewer than five or zero observations. Statistical analyses were performed using SPSS (IBM Corp. Released 2019, IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp., USA). The database was resampled to handle imbalanced sample sizes using the R package version 4.4.4 [http://www.r-project.org/index.html]. Multivariable and exact logistic regression (for small sample sizes) were used to calculate the odds ratio (OR), and 95% confidence intervals (CI) served to evaluate the association between predictors and breast cancer risk, as well as the association between the genotypes and tumor characteristics. Statistical analyses were performed using the SPSS and STATA (StataCorp. 2015. Stata Statistical Software: Release 14. College Station, TX).

RESULTS

Descriptive characteristics of study participants

A total of 170 subjects comprising 68 estrogen receptor-positive (ER+) breast cancer patients (cases), 82 benign breast disease (BBDs) and 20 healthy volunteers (controls) were included in the study. The age range of the cases (28 to 78) and mean age (51 years) were higher compared to that of the BBDs (23 to 74) with a mean of 43 years and the control subjects (25 to 52) with an average of 38 years. The observed frequencies in the different age categories between cases and BBDs differed significantly ($\chi2=20.51$, df=5, P<0.001), which was analogous to the differences between cases and controls ($\chi2=20.79$, df=5, P<0.001). There was significant variation in the level of education among the study participants. The likelihood of the participants having a tertiary level of education was significantly high in both models; between cases and BBDs and cases and controls it was ($\chi2=16.963$, df=3, P<0.001) and ($\chi2=10.78$, df=3, P=0.013) respectively. The case participants reported a significantly higher proportion of family members with breast cancer ($\chi2=4.496$, df=1, P=0.034) and other types of cancers ($\chi2=8.346$, df=1, P=0.004) than the control participants (Table 1).

Table 1. Descriptive characteristics of study participants. The table shows comparative analysis of sociodemographic, anthropometric measures, family medical history, comorbidities, lifestyle, reproduction and genotype variables in estrogen receptor-positive (ER+) breast cancer (BC) cases, breast benign disease (BBD) and control participants.

	Cases (N =68)			**BBDs (N = 82)**			**Cases vs BBDs**		**Controls (N= 20)**				**Case vs Control**
	N	**n**	**%**	**N**	**n**	**%**	**χ2/Exact**	**P**	**N**	**n**	**%**	**χ2/Exact**	**P**
Sociodemographics													
Age at diagnosis (years)													
20-29	65	1	1.5	80	3	3.8	20.51	**<0.001***	20	4	20	20.79	**<0.001***
30-39		9	13.8		25	31.2				5	25		
40-49		17	26.2		34	42.5				10	50		
50-59		26	40.0		12	15.0				1	5		
60-69		8	12.3		5	6.2				0	0		
70+		4	6.2		1	1.2				0	0		
Range, mean, sd (years)	28-78, 51, 10.86			23-74, 43, 9.2					25-52, 38, 8.1				
Level of education													
None	66	2	3.0	78	0	0	16.780	**<0.001***	19	0	0	10.513	**0.006***
Primary		10	15.2		3	3.8				0	0		
Secondary		14	21.2		5	6.4				0	0		
Tertiary		40	60.6		70	89.7				19	100		
Anthropometric measures (Body mass index)													
<18.5	49	2	4.1	78	0	0	3.1444	0.354*	19	0	0	4.343	0.214*
18.6-24.9		7	18.4		16	20.5				7	36.8		
25-29.9		14	24.5		29	37.2				6	31.6		
>30		26	53.1		33	42.3				6	31.6		
Range, mean, sd	10-42, 29.7, 6.7			19-43, 28.53, 4.5					19-34, 27.3, 3.9				
Family medical history													
History of breast cancer													
No	65	52	80.0	77	60	77.9	0.091	0.762	19	19	100	4.496	**0.034***
Yes		13	20.0		17	22.1				0	0		
History of other types of cancers													
No	64	43	67.2	77	58	75.3	1.139	0.286	19	19	100	8.346	**0.004***
Yes		21	32.8		19	24.7				0	0		
Comorbidities													
Diabetes													
No	66	58	87.9	79	72	91.1	0.412	0.521	19	19	100	2.542	0.111*
Yes		8	12.1		7	8.9				0	0		

Table 1. continued

Hypertension													
No	64	42	65.6	78	60	76.9	2.218	0.136	19	14	73.7	0.434	0.510
Yes		22	34.4		18	23.1				5	26.3		
Lifestyle													
Ever smoked tobacco													
No	66	64	97.0	78	76	97.4	0.029	1.000*	19	19	100	0.590	1.000*
Yes		2	3.0		2	2.6				0	0		
Alcohol consumption in the past year													
No	66	42	63.6	79	28	35.4	11.446	**<0.001**	19	5	26.3	8.312	**0.004***
Yes		24	36.4		51	64.6				14	73.7		
Reproduction													
Age at menarche (years)													
<12	64	2	3.1	78	2	2.6	4.300	0.116*	19	3	15.8	4.421	0.119*
13-14		39	60.9		62	79.5				13	68.4		
>15		23	35.9		14	17.9				3	15.8		
Range, mean, sd	8 – 22, 15, 1.9 years			11-17, 14, 1.4 years					10-17, 13.9, 1.3 years				
Menopausal status													
No	66	30	45.5	77	60	77.9	16.060	**<0.001**	19	18	94.7	14.58	**<0.001***
Yes		36	54.5		17	22.1				1	5.3		
Gravidity													
No	66	5	7.6	79	15	19.0	3.938	**0.047***	19	6	31.6	7.54	**0.013**
Yes		61	92.4		64	81.0				13	68.4		
Ever used modern family planning methods													
No	65	22	33.8	78	16	20.5	3.230	0.072	19	3	15.8	6.369	**0.033**
Yes		43	66.2		62	79.5				16	84.2		
Genotypes													
rs4646903													
TT	68	44	65	82	62	76	2.806	0.256*	20	9	45	3.935	0.122*
TC		21	31		19	23				8	40		
CC		3	4		1	1				3	15		
rs1048943													
AA	68	61	89.7	82	72	88	0.298	0.862*	20	20	100	2.237	0.327*
AG		5	7.4		8	10				0	0		
GG		2	2.9		2	2				0	0		
Radiation therapy													
No	66	65	98.5	77	74	96.1	0.741	0.624*	19	19	100	0.291	1.000*
Yes		1	1.5		3	3.9				0	0		

N – total number of respondents; n – frequencies; $\chi 2$ – Pearson Chi-square test of independence; Exact – Fisher's exact test; * – calculated using Fisher's exact test; the P- value of statistical significance is in bold.

The difference in the proportions of alcohol consumption in the past year was significantly higher in both BBDs ($\chi 2$=11.446, df=1, P<0.001) and control participants ($\chi 2$=8.312, df=1, P=0.004) than in the cases. Most cases had attained menopausal status relative to the BBDs ($\chi 2$=16.060: df=1: P<0.001) as well as relative to the control subjects ($\chi 2$=14.58, df=1, P<0.001). Cases reported significantly higher frequencies of gravidity compared to BBDs ($\chi 2$=3.938, df=1, P=0.047) and control participants ($\chi 2$=7.54, df=1, P=0.013). A significantly higher percentage of control subjects reported using modern family planning methods compared to the cases ($\chi 2$=6.369, df=1, P=0.033) (Table 1). No significant differences were observed between the study participants in the proportions of other predictors such as comorbidities (diabetes and hypertension), smoking tobacco, radiation therapy, and *CYP1A1* genotypes (rs4646903 and rs1048943) (Table 1).

Descriptive characteristics of tumor parameters in cases

The frequency of breast cancer tumor characteristics in the cases was also analyzed. Most cases (76.5%, n=52) presented with invasive ductal carcinoma (IDC). Most tumors were moderately differentiated (grade II) (62.1%, n=41) and were diagnosed at stage II (47%, n=25). Luminal A constituted most of the cases (82.8%, n=53) (Table 2).

Association of predictors with ER+ breast cancer risk

Resampling of the dataset to handle the imbalanced sample sizes projected a high accuracy (77%) on the random over sampling model, hence its adoption in association analysis. Controls had 39% odds of developing breast cancer at the age of 40 years and above (OR=0.39, 95% CI=0.18-0.82, P=0.01). The odds of a participant in the control group having a close relative with breast cancer and developing breast cancer was significantly low at 3% (OR=0.03, 95% CI=0 0-0.19, P<0.001). The odds of BBDs and control participants having close relatives with other types of cancers and developing breast cancer were significantly low at 29% (OR=0.29, 95% CI=0.10-0.88, P=0.008) and 2% (OR=0.02, 95% CI=0-0.15, P<0.001), respectively. The odds of individuals in the BBDs and control groups being at risk of developing breast cancer owing to alcohol consumption in the past year were high. Individuals in the BBD group were three-fold more at risk of developing breast cancer (OR=3.4, 95% CI=1.38-8.71, P=0.008) while the controls were at five-fold risk (OR=5.70, 95% CI=2.77-12.08, P<0.001). Control individuals who had attained menarche at an older age (>15 years) had 25% odds of developing breast cancer (OR=0.25, 95% CI=0.10-0.58, P<0.001). The odds of postmenopausal women with benign breast disease being at risk of developing breast cancer was 26% (OR=0.26, 95% CI=0.10-0.71, P=0.008) while that of control individual was 5% (OR=0.05, 95% CI=0.01-0.19, P<0.001). A positive gravidity status showed 32% odds of breast cancer risk in control individuals (OR=0.32, 95% CI=0.08-1.00, P=0.05). Women in the control group who had never used modern family planning methods had a four-fold higher risk of developing breast cancer (OR=4.38, 95% CI=1.77-12.06, P<0.001). The variant genotype CC of *CYP1A1* (rs4646903) harbored a five-fold risk of control individuals developing breast cancer (OR=5.79, 95% CI=1.42-34.22, P<0.001). The variant genotype GG of *CYP1A1* (rs1048943) had 12% odds of developing breast cancer in BBDs (OR=0.12, 95% CI=0.01-0.98, P=0.04), whereas the heterozygous genotype AG had 11% odds of developing breast cancer in control individuals (OR=0.11, 95% CI=0-0.78, P=0.02) (Table 3).

Table 2. Descriptive characteristics of tumor parameters in cases. The table shows distribution of the tumor characteristics in ER+ BC cases.

Tumor characteristic	Levels	N	n	%
Histological type	IDC (Invasive ductal carcinoma)	68	52	76.5
	Others		16	23.5
Grade	Grade 1 (Well differentiated)	66	3	4.5
	Grade 2 (Moderately differentiated)		41	62.1
	Grade 3 (Poorly differentiated)		22	33.3
Lymph node involvement	N0	53	21	39.6
	N1		17	32.1
	N2		11	20.8
	N3		4	7.5
Tumor size	T0	53	2	3.77
	Tis		2	3.77
	T1 (<2 cm)		13	24.5
	T2 (2-5 cm)		30	56.6
	T3 (>5 cm)		5	9.43
	T4		1	1.89
Clinical stage	Stage I	53	13	25
	Stage II		25	47
	Stage III		15	28
Molecular subtypes	Luminal A	64	53	82.8
	Luminal B		11	17.2

Association analysis of rs1048943 and rs4646903 genotypes with ER+ breast tumor characteristics

The associations of genotypes in rs4646903 and rs1048943 with ER+ breast cancer tumor characteristics such as tumor stage, grade, histological subtypes and molecular subtypes were determined. Stages I and II were taken as the early stage, whereas stage III was taken as the late stage of ER+ breast cancer. The wild-type genotype TT of rs4646903 and AA of rs1048943 were used as reference points. The heterozygous genotype TC of

Table 3. Association of predictors with ER+ BC risk. The table shows association of sociodemographics, anthropometric measures, family medical history, comorbidities, lifestyle, reproduction and genotype variables with estrogen receptor positive breast cancer. Analyses were performed in case-BBDs and case-control models.

	Cases vs BBDs		Cases vs Controls	
	OR (95% CI)	**P**	**OR (95% CI)**	**P**
Sociodemographics				
Age at diagnosis (years)				
<40	Reference		Reference	
>40	0.47 (0.15-1.47)	0.19	0.39 (0.18-0.82)	**0.01***
Family history of breast cancer				
No	Reference		Reference	
Yes	1.18 (0.43-3.24)	0.74	0.03 (0 - 0.19)	**<0.001***
Family history of other types of cancers				
No	Reference			
Yes	0.29 (0.10-0.88)	**0.02**	0.02 (0 - 0.15)	**<0.001***
Comorbidity (diabetes)				
No	Reference		Reference	
Yes	0.95 (0.18 – 5.00)	0.95	0.12 (0 – 0.15)	<0.001*
Comorbidity (hypertension)				
No	Reference		Reference	
Yes	0.70 (0.23-2.11)	0.53	1.20 (0.44 – 3.54)	0.88*
Radiation therapy				
No	Reference		Reference	
Yes	2.31 (0.16 – 31.5)	0.52	0.42 (0 – 16.52)	0.59*
Ever smoked tobacco				
No	Reference		Reference	
Yes	0.28 (0.01-4.38)	0.36	0.11 (0.00-1.06)	0.05*
Alcohol consumption in the past twelve months				
No	Reference		Reference	
Yes	3.4 (1.38-8.71)	**0.008**	5.70 (2.77-12.08)	**<0.001***
Body mass index				
18 - 24.9	Reference		Reference	
<18	0.35 (0 – 3.45)	0.39	0.26 (0 – 2.5)	0.26*
25-29.9	2.20 (0.92 – 5.42)	0.07	1.75 (0.75 – 4.13)	0.21*
>30	1.31 (0.60 – 2.90)	0.57	0.61 (0.27 – 1.38)	0.27*
Reproduction				
Age at menarche(years)				
13-14	Reference		Reference	
<12	0.745 (0.19-2.80)	0.66	2.12 (0.85 – 5.62)	0.11*
>15	0.56 (0.23 – 1.36)	0.20	0.25 (0.10 -0.58)	**<0.001***
Menopausal status				
No	Reference		Reference	
Yes	0.26 (0.10 – 0.71)	**0.008**	0.05 (0.01 – 0.19)	**<0.001***
Ever been pregnant				
No	Reference		Reference	
Yes	0.99 (0.25 – 3.94)	0.99	0.32 (0.08 – 1.00)	**0.05***
Ever used any modern family planning methods				
No	Reference		Reference	
Yes	1.42 (0.46 – 4.37)	0.53	4.38 (1.77-12.06)	**<0.001***
Genotypes				
rs4646903				
TT	Reference		Reference	
TC	0.79 (0.30 – 2.06)	0.63	1.91 (0.94 – 3.94)	0.07*
CC	0.44 (0.25 – 7.62)	0.57	5.79 (1.42 – 34.22)	**<0.001**
rs1048943				
AA	Reference		Reference	
AG	1.53 (0.35 – 6.60)	0.56	0.11 (0 -0.78)	**0.02***
GG	0.12 (0.01 – 0.98)	**0.04**	0.23 (0 – 2.22)	0.21

* – Calculated using exact logistic regression; OR – odds ratio; CI – confidence interval; P values of statistical significance are shown in bold

rs4646903 conferred about 8-fold statistically significant odds of occurring in the late stage (OR=8.06, 95% CI=1.68-44.60, P=0.01) and slightly above 6-fold odds of being a luminal B molecular subtype (OR=6.56, 95% CI=0.08-148.79, P=0.02). No significant associations were found between the genotypes and grade as well as histological subtypes of tumor characteristics (Supplementary Table S1).

DISCUSSION

This study reports the prevalence of rs4646903 and rs1048943 single nucleotide polymorphisms (SNPs) in the *CYP1A1* gene in women diagnosed with ER+ breast cancer at Aga Khan University Hospital Nairobi (AKUHN) and AIC Hospital Kijabe (KAIC). Associations between SNPs and breast cancer risk as well as breast tumor characteristics were determined. Differences in the proportions of probable predictors of breast cancer among the study participants were analyzed. The rs4646903 genotype variants TC and CC were significantly associated with breast cancer risk, whereas the TC genotype was positively associated with late-stage and luminal B molecular subtype.

The first phase of estrogen metabolism increases estrogen polarity through catabolism by *CYP1A1* genes. The resulting catechol estrogens may be further oxidized to form quinones, which may cause genetic mutations and result in breast cancer. Both catechol estrogens and quinones are detoxified by conjugation with phase II enzymes. Polymorphisms within estrogen-metabolizing genes and their association with breast cancer susceptibility have generated inconsistent findings in different populations, sparking considerable research interest [16].

Table 4. Association analysis of rs4646903 and rs1048943 genotypes with ER+ breast tumor characteristics. The table shows analysis for the associations between rs4646903 and rs1048943 with breast tumor characteristics such as tumor stage and molecular subtypes.

Association analysis of rs4646903 and rs1048943 with tumor stage					
GENE (SNP RS ID)	**Genotype (N)**	**Early stage: Stage I & II (n, %)**	**Late stage: Stage III (n, %)**	**OR (95% CI)**	**P**
rs4646903	TT (38)	32 (60)	6 (11)	Reference	
	TC (13)	5 (9)	8 (15)	8.06 (1.68 – 44.60)	**0.01**
	CC (2)	1 (2)	1 (2)	5.01 (0.05 – 431.27)	0.64
rs1048943	AA (48)	34 (66)	14 (26)	Reference	
	AG (3)	2 (3)	1 (2)	1.20 (0.01 – 25.03)	1.00
	GG (2)	2 (3)	0 (0)	1.05 (0.00 – 13.89)	1.00
Association analysis of rs4646903 and rs1048943 with molecular subtype					
GENE (SNP RS ID)	**Genotype (N)**	**Luminal A (n, %)**	**Luminal B (n, %)**	**OR (95% CI)**	**P**
rs4646903	TT (41)	38 (59)	3 (4)	Reference	
	TC (20)	13 (20)	7 (10)	6.56 (1.27-45.22)	**0.02**
	CC (3)	2 (3)	1 (1)	5.86 (0.08-148.79)	0.50
rs1048943	AA (57)	47 (73)	10 (16)	Reference	
	AG (5)	4 (6)	1 (2)	1.17 (0.021-13.66)	1.00
	GG (2)	2 (3)	0 (2)	2.02 (0.00-26.94)	1.00

N – total number of respondents; n – frequencies; OR – odds ratio; CI – confidence intervals; values of statistical significance are shown in bold

Reports on breast cancer in African women have referred to an earlier age of diagnosis (below 65 years), with the majority occurring during or around menopause, compared to that in developed countries. However, early detection is limited by a lack of awareness, delayed health-seeking behavior and deficient diagnostic processes [17]. In this study, the mean age at the time of breast cancer diagnosis was 51 years. An earlier study of Kenyan women with BC showed that the mean age at diagnosis of breast cancer was 49.2 years [2]. Another study conducted in Nigeria reported an average diagnostic age of 49.5 years [18]. Amadori et al. established that the average age at diagnosis in Tanzanian and Italian women was 51 years, whereas in African Americans it was 57 years [19].

BC diagnosis is crucial for survival [20]. In this study, most cases (47%) were diagnosed as stage II of the disease. Our findings are consistent with two previous studies conducted in AKUH in which 50% and 49.1% of the cases were diagnosed at stage II in 2014 and 2021, respectively [2,21]. Contrary to our findings, it was reported that most breast cancers (77%) in the sub-Saharan African region were diagnosed at late stages [20]. This could be attributed to patient and system delays, as was established in a study conducted in Rwanda [22]. In Nigeria, >70% of the patients were diagnosed at a late stage of the disease. The observation that most breast cancers were in the early stages in this study could probably be attributed to the nature of the patients seeking health care in the two tertiary hospitals (AKUHN and KAIC). Most participants had obtained tertiary education; therefore, they were more likely to have a high socioeconomic status and better health-seeking behavior, leading to an early diagnosis [23].

Moreover, invasive ductal carcinoma (IDC) was the most common histological subtype of BC (84.2%). This observation is like that of previous studies that reported high frequencies of IDC: 76.5% in Kenyan patients [21], 82.3% in Nigerians (18), 92.8% in Tanzanians and 90.6% in Italians [19]. This finding is consistent with an earlier finding where it was reported that IDC was the most common histological subtype of breast cancer. IDC is attributed to DNA damage in breast tissue cells resulting from various triggers such as, age and hormonal exposure [24].

Most of the cases were progesterone-receptor (PR) positive, followed by human epidermal growth factor-2 (HER2) negative hormone receptors (85.3%, and 76.6%, respectively). Therefore, the luminal A (ER+/PR+/HER2-) molecular subtype was the most prevalent (82.8%). These findings contradict those of Sayed et al. who reported the luminal B (ER+/PR+/HER2+) molecular subtype to be the most prevalent 35.8% [2]. Among Nigerian women, the prevalence of PR was the highest (54.7%) [18].

CYP1A1 is located on chromosome 15 and comprises seven exons and six introns that span 5810 base pairs[25]. The contribution of *CYP1A1* to estrogen metabolism and polycyclic aromatic hydrocarbon (PAH) carcinogen metabolism points to its involvement in breast cancer [26]. *CYP1A1* is a critical member of the phase I estrogen-metabolizing enzyme family because of its role in catalyzing the formation of 2,3-hydroxyestrone, which are subsequently converted to estradiol 2,3 quinones [27]. Among the *CYP1A1* polymorphisms, *CYP1A1* (rs4646903) occurs in the 3'-UTR and results from a T/C transition. The three genotypes were TT (wild type), TC (variant heterozygous) and CC (variant homozygous) [26]. The *MspI* restriction enzyme identifies the *CYP1A1* (rs4646903) polymorphism in the non-coding region of the gene [28].

Studies exploring the association of rs4646903 genotypes, TT, TC and CC with breast cancer in variant populations have yielded inconsistent results. This is not surprising given the dual involvement of *CYP1A1* in the activation of carcinogens and estrogen metabolism [29]. Thus, no significant association between rs4646903 polymorphisms and breast cancer risk was observed in Chinese women [16]. In contrast, the homozygous variant genotype CC was significantly associated with the occurrence and stage of breast cancer [26]. The distribution of the dominant model (CT+CC versus TT) in early- versus late-stage breast cancer among Egyptian women was not significantly different; however, the odds of association of the variants TC and CC in the late stage were 3-fold more than those in the early stage [30]. An increase in the proportion of the variant genotype CC of rs4646903 with breast cancer stage was reported in contrast to no significant association between genetic polymorphisms of rs4646903 and breast cancer in East Asian, Caucasian, or African populations [31]. A study showed no association between the variant genotypes TC and CC and breast cancer, stage, grade, or molecular sub-type in Iraqi women diagnosed with breast cancer [28]. A study on a southern Indian population indicated a positive association of the variant genotypes TT and TC with breast cancer [13] where a 5-fold association of rs4646903 variant genotype CC supports the role of *CYP1A1* in breast cancer susceptibility.

The second *CYP1A1* SNP of interest, based on its association with breast cancer, was rs1048943. It is located in exon 7 of the gene and results in a A/G transition, which causes the substitution of isoleucine for valine at codon 462 [32]. A/G single nucleotide polymorphism was detected using the *BsrDI (BseMI)* restriction enzyme. Like the trend observed for rs4646903, the wild-type genotype of rs1048943 (AA) was the most prevalent in the study subjects. A Japanese study reported a significantly lower risk of breast cancer in rs1048943, AG and GG variant carriers [33]. A protective factor for the rs1048943 genotype was also observed in French and Austrian populations where the variant genotypes of rs1048943, AG and GG were significantly associated with susceptibility to breast cancer in case-control and case-BBD models, respectively [34][?].

CONCLUSIONS

The findings from this study revealed statistically significant associations between breast cancer risk and variant genotypes, CC of rs4646903, and AG and GG of rs1048943. The strength of this study is the inclusion of benign breast disease (BBD) participants alongside controls. This allowed for a comparative analysis of the associations between predictors and the risk of BC in cases and controls compared to cases and BBDs. Although BBD is an established risk factor for BC, it is not known whether the association will vary based on exposure to other predictors. The incorporation of both the genotypic as well as the non-genetic predictors enabled analysis of the interplay between the various predictors and their role in breast cancer risk. Further investigations employing larger sample sizes and other genes could provide more information on the role of SNPs in the susceptibility to breast cancer and its progression.

Funding: This research was funded by the Africa-ai-Japan Project Innovation Research Fund 2020/2021.

Author contributions: Conceptualization, MM, and FM; methodology, MM, SN, VM, SS and FM; software, MM; validation, MM, SN, VM, SS and FM; formal analysis, MM; investigation, MM; resources, MM and SN; data curation, MM; writing – original draft preparation, MM; writing, review and editing, MM, SN, VM, SS and FM; visualization, MM; supervision, SN, VM, SS and FM; project administration, SG, VM, SS and FM; funding acquisition, MM and SN. All authors have read and agreed to the published version of the manuscript.

Conflict of interest disclosure: The authors declare no potential conflicts of interest.

Data availability statements: All data underlying the reported findings have been provided as part of the submitted article and are available at:
https://www.serbiosoc.org.rs/NewUploads/Uploads/Murithi%20et%20al_8342_Data%20Report.pdf

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-49. https://doi.org/10.3322/caac.21660
2. Sayed S, Fan S, Moloo Z, Wasike R, Bird P, Saleh M, Shaikh AJ, Figueroa JD, Naidoo R, Makokha FW, Gardner K, Oigara R, Njoroge FW, Magangane P, Mutebi M, Chauhan R, Mwanzi S, Govender D, Yang XR. Breast cancer risk factors in relation to molecular subtypes in breast cancer patients from Kenya. Breast Cancer Res. 2021;23(1):1-17. https://doi.org/10.1186/s13058-021-01446-3
3. McDonald ES, Clark AS, Tchou J, Zhang P, Freedman GM. Clinical diagnosis and management of breast cancer. J Nucl Med. 2016;57:9S-16S. https://doi.org/10.2967/jnumed.115.157834
4. Pace LE, Shulman LN. Breast Cancer in Sub-Saharan Africa: Challenges and Opportunities to Reduce Mortality. Oncologist. 2016;21(6):739-44. https://doi.org/10.1634/theoncologist.2015-0429
5. Vanderpuye V, Grover S, Hammad N, Prabhakar P, Simonds H, Olopade F, Stefan DC. An update on the management of breast cancer in Africa. Infect Agent Cancer. 2017;12(1):1-12. https://doi.org/10.1186/s13027-017-0124-y
6. Gakunga R, Ali Z, Korir A, Kinyanjui AW, Ochieng E, Gikaara N, Maluni F, Subramanian S. Social determinants and individual health-seeking behaviour among women in Kenya: Protocol for a breast cancer cohort feasibility study. BMJ Open. 2019;9(1):e023171. https://doi.org/10.1136/bmjopen-2018-023171
7. Jameera BA, Jubie S, Nanjan MJ. Estrogen receptor agonists/antagonists in breast cancer therapy: A critical review. Bioorg Chem. 2017;71:257-74. https://doi.org/10.1016/j.bioorg.2017.02.011
8. Santen RJ, Yue W, Wang JP. Estrogen metabolites and breast cancer. Steroids. 2015;99:61-6. https://doi.org/10.1016/j.steroids.2014.08.003
9. Sampson JN, Falk RT, Schairer C, Moore SC, Fuhrman BJ, Dallal CM, Bauer DC, Dorgan JF, Shu XO, Zheng W, Brinton LA, Gail MH, Ziegler RG, Xu X, Hoover RN, Gierach GL. Association of estrogen metabolism with breast cancer risk in different cohorts of postmenopausal women. Cancer Res. 2017;77(4):918-25. https://doi.org/10.1158/0008-5472.CAN-16-1717
10. Blackburn HL, Ellsworth DL, Shriver CD, Ellsworth RE. Role of cytochrome P450 genes in breast cancer etiology and treatment: effects on estrogen biosynthesis, metabolism, and response to endocrine therapy. Cancer Causes Control. 2015;26(3):319-32. https://doi.org/10.1007/s10552-014-0519-7
11. Francis AM, Ramya R, Ganesan N, Kumarasamy P, Paul SFD, Munirajan AK, Divya M. Breast cancer susceptibility genes in estrogen metabolizing pathway in a southern Indian population. Meta Gene. 2019;19:225-34. https://doi.org/10.1016/j.mgene.2018.12.009
12. Zhou Y, Ingelman-Sundberg M, Lauschke V. Worldwide Distribution of Cytochrome P450 Alleles: A Meta-analysis of Population-scale Sequencing Projects. Clin Pharmacol Ther. 2017;102(4):688-700. https://doi.org/10.1002/cpt.690
13. Newman LA. Breast Cancer Disparities: Socioeconomic Factors versus Biology. Ann Surg Oncol. 2017;24(10):2869-75. https://doi.org/10.1245/s10434-017-5977-1
14. Golmohammadzadeh G, Mohammadpour A, Ahangar N, Shokrzadeh M. Polymorphisms in phase I (Cyp450) genes CYP1A1 (rs4646421), CYP1B1 (rs1056836), CYP19A1 (rs749292) and CYP2C8 (rs1058930) and their relation to risk of breast cancer: A case-control study in Mazandaran province in north of Iran. Open Access Maced J Med Sci. 2019;7(15):2488-96. https://doi.org/10.3889/oamjms.2019.667
15. Anyigba CA, Awandare GA, Paemka L. Breast cancer in sub-Saharan Africa: The current state and uncertain future. Exp Biol Med. 2021;246(12):1377-87. https://doi.org/10.1177/15353702211006047
16. Boyapati SM, Xiao OS, Gao YT, Cai Q, Jin F, Zheng W. Polymorphisms in CYP1A1 and breast carcinoma risk in a population-based case-control study of Chinese women. Cancer. 2005;103(11):2228-35. https://doi.org/10.1002/cncr.21056
17. Thomas AS, Kidwell KM, Oppong JK, Adjei EK, Osei-Bonsu E, Boahene A, Jiagge E, Gyan K, Merajver SD. Breast cancer in Ghana: Demonstrating the need for population-based cancer registries in low- And middle-income countries. J Glob Oncol. 2017;3(6):765-72. https://doi.org/10.1200/JGO.2016.006098
18. Adebamowo CA, Famooto A, Ogundiran TO, Aniagwu T, Nkwodimmah C, Akang EE. Immunohistochemical and molecular subtypes of breast cancer in Nigeria. Breast Cancer Res Treat. 2008;110(1):183-8. https://doi.org/10.1007/s10549-007-9694-5
19. Amadori D, Serra P, Bravaccini S, Farolfi A, Puccetti M, Carretta E, Medri L, Nanni O, Tumedei MM, Kahima J, Masalu N. Differences in biological features of breast cancer between Caucasian (Italian) and African (Tanzanian) populations. Breast Cancer Res Treat. 2014;145(1):177-83. https://doi.org/10.1007/s10549-014-2903-0
20. Jedy-Agba E, McCormack V, Adebamowo C, Dos-Santos-Silva I. Stage at diagnosis of breast cancer in sub-Saharan Africa: a systematic review and meta-analysis. Lancet Glob Health. 2016;4(12):e923-35. https://doi.org/10.1016/S2214-109X(16)30259-5
21. Sayed S, Moloo Z, Wasike R, Bird P, Oigara R, Govender D, Kibera J, Carrara H, Saleh M. Is breast cancer from Sub Saharan Africa truly receptor poor? Prevalence of ER/PR/HER2 in breast cancer from Kenya. Breast. 2014;23(5):591-6. https://doi.org/10.1016/j.breast.2014.06.006

22. Pace LE, Mpunga T, Hategekimana V, Dusengimana J-MV, Habineza H, Bigirimana JB, Mutumbira C, Mpanumusingo E, Ngiruwera JP, Tapela N, Amoroso C, Shulman LN, Keating NL. Delays in Breast Cancer Presentation and Diagnosis at Two Rural Cancer Referral Centers in Rwanda. Oncologist. 2015;20(7):780-8. https://doi.10.1634/theoncologist.2014-0493.
23. Cutler DM, Lleras-Muney A. Understanding differences in health behaviors by education. J Health Econ. 2010;29(1):1-28. https://doi.org/10.1016/j.jhealeco.2009.10.003
24. Mohammed AA. The clinical behavior of different molecular subtypes of breast cancer. Cancer Treat. Res. Commun. 2021;29:100469. https://doi.org/10.1016/j.ctarc.2021.100469
25. Balmukhanov TS, Khanseitova AK, Nigmatova VG, Varchenko SP, Talaeva SZ. Associations of Polymorphisms of the CYP 1 A 1 and CYP 1 B 1 Cytochrome P450 Genes with Breast Cancer in Kazakhstan. Adv Breast Cancer Res. 2013;2(3)51-5.
26. Wang H, Wang WJ. Relationship between CYP1A1 polymorphisms and invasion and metastasis of breast cancer. Asian Pac J Trop Med. 2013;6(10):835-8. https://doi.org/10.1016/S1995-7645(13)60148-0
27. Mescher M, Haarmann-Stemmann T. Modulation of CYP1A1 metabolism: From adverse health effects to chemoprevention and therapeutic options. Pharmacol Ther. 2018;187:71-87. https://doi.org/10.1016/j.pharmthera.2018.02.012
28. Ibrahem SQ, Ahmed HQ, Amin KM. Genetic Variations in Cytochrome P450 1A1 and 1B1 Genes in a Cohort of Patients from Iraq Diagnosed with Breast Cancer. Breast Cancer. 2021;15:11782234211050727. https://doi.org/10.1177/11782234211050727
29. Pescatori S, Berardinelli F, Albanesi J, Ascenzi P, Marino M, Antoccia A, di Masi A, Acconcia F. A tale of ice and fire: The dual role for 17β-estradiol in balancing DNA damage and genome integrity. Cancers (Basel). 2021;13(7):1583. https://doi.org/10.3390/cancers13071583
30. Tolba W, Rizk El Baz, Othman AI, Roshdy S, Settin A. Polymorphisms of Estrogen Biosynthesis and Metabolizing Genes in Egyptian Women With Breast Cancer. Int J Adv Res (Indore). 2017;5(5):453-64. https://doi.org/10.21474/IJAR01/4131
31. Chen C, Huang Y, Li Y, Mao Y, Xie Y. Cytochrome P450 1A1 (CYP1A1) T3801C and A2455G polymorphisms in breast cancer risk: A meta-analysis. J Hum Genet. 2007;52(5):423-35. https://doi.org/10.1007/s10038-007-0131-8
32. Moreno-Galván M, Herrera-González NE, Robles-Pérez V, Velasco-Rodríguez JC, Tapia-Conyer R, Sarti E. Impact of CYP1A1 and COMT genotypes on breast cancer risk in Mexican women: A pilot study. International J Biomark. 2010;25(3):157-63. https://doi.org/10.1177/172460081002500306
33. Miyoshi Y, Takahashi Y, Egawa C, Noguchi S. Breast cancer risk associated with CYP1A1 genetic polymorphisms in Japanese women. Breast J. 2002;8(4):209-15. https://doi.org/10.1046/j.1524-4741.2002.08404.x
34. Hussain T, Alrokayan S, Upasna U, Pavithrakumari M, Jayapriya J, Kutala VK, Naushad SM. Meta-analysis of genetic polymorphisms in xenobiotic metabolizing enzymes and their association with breast cancer risk. J Genet. 2018;97(2):523-37. https://doi.org/10.1007/s12041-018-0946-4

SUPPLEMENTARY MATERIAL

Supplementary Table S1. Association analysis of rs4646903 and rs1048943 genotypes with ER+ breast tumor characteristics. The table shows the analysis for the associations between rs4646903 and rs1048943 with breast tumor characteristics such as tumor grade and histological subtype.

Association analysis of rs4646903 and rs1048943 with tumor grade					
GENE (SNP RS ID)	**Genotype (N)**	**Well & moderate differentiation: Grade I & II (n, %)**	**Poor differentiation: Grade III (n, %)**	**OR (95% CI)**	**P**
CYP1A1 (rs4646903)	TT (42)	28 (42)	14 (21)	Reference	
	TC (21)	13 (20)	8 (12)	1.22 (0.35 – 4.13)	0.91
	CC (3)	3 (5)	0 (0)	0.55 (0.00 – 5.40)	0.63
CYP1A1 (rs1048943)	AA (60)	40 (59)	20 (29)	Reference	
	AG (4)	2 (3)	2 (3)	1.97 (0.13 – 29.14)	0.85
	GG (2)	2 (3)	0 (0)	0.85 (0.00 – 11.26)	0.91
Association analysis of rs4646903 and rs1048943 with histological subtype					
GENE (SNP RS ID)	**Genotype (N)**	**IDC (n, %)**	**Others (n, %)**	**OR (95% CI)**	**P**
CYP1A1 (rs4646903)	TT (44)	35 (51)	9 (13)	Reference	
	TC (21)	14 (21)	7 (10)	1.92 (0.50 – 7.19)	0.40
	CC (3)	3 (4)	0 (0)	1.06 (0.00 – 10.72)	1.00
CYP1A1 (rs1048943)	AA (61)	46 (68)	15 (22)	Reference	
	AG (5)	4 (6)	1 (1)	0.76 (0.01 – 8.61)	1.00
	GG (2)	2 (3)	0	1.31 (0.00 – 10.72)	1.00