GMR
The Original

ASSOCIATION OF VEGFA RS699947 POLYMORPHISM WITH BREAST CANCER: A CASE-CONTROL AND IN-SILICO FUNCTIONAL STUDY

Esha Mukhtar[1], Nageen Hussain[1*], Saima Shahzad Mirza[2], Asma Chaudhary[3], Ayesha Aihetasham[4,] Isbah Imtiaz Qazi[5]

[1.] Institute of Microbiology and Molecular Genetics, University of the Punjab, Quaid-e-Azam campus, Lahore-54590, Pakistan.
[2.] Department of Zoology, University of Education, Bank Road Campus, Lahore, Pakistan.
[3.] Department of Zoology, Division of Science and Technology, University of Education, Township, Lahore, Pakistan.
[4.] Institute of Zoology, University of the Punjab, Quaid-e-Azam campus, Lahore-54590, Pakistan.
[5.] Designation: Head Quality Control & Quality Assurance Laboratory, Lead Auditor FSSC-22000, Agro Tech Food Industries Pvt. Ltd. Gujranwala-Pakistan
*Correspondence: Nageen Hussain, Email: nageen.mmg@pu.edu.pk

Abstract
Breast cancer (BC) is a leading cause of morbidity and mortality among women globally, with genetic factors playing a significant role in its progression. This study aims to assess the association of VEGFA rs699947 polymorphism with BC risk in a Pakistani population through a case-control design. The objectives are to identify any genotypic and allelic associations, perform functional analysis, and validate findings using in-silico tools. DNA from 100 women (50 BC cases, 50 controls) is analyzed for VEGFA rs699947 polymorphism using TETRA-ARMS PCR, and statistical tests are employed to evaluate associations. The study finds no significant association between rs699947 and BC risk, though functional analysis indicated a minor regulatory role. These findings suggest limited genetic influence but provide insight into the SNP's functionality, emphasizing the need for further research. The study highlights potential therapeutic value of natural compounds, such as Quercetin and Apigenin, inhibiting VEGFA. The novelty of this study lies in its first examination of this polymorphism in a Pakistani cohort and its integration of in-silico analysis for validation.

KEYWORDS: Breast Cancer, VEGFA, rs699947, Angiogenesis, Genotyping, Polymorphism

1. INTRODUCTION

Breast cancer (BC) remains one of the most prevalent and deadly cancers globally, accounting for a significant portion of cancer-related morbidity and mortality among women. BC is a highly heterogeneous disease that is caused by a multitude of factors such as age, gender, lifestyle, family history, hormonal dysregulations, environmental conditions, and gene predispositions [1-4]. In women, it is a leading cause of death which makes up one in six of the death cases. According to GLOBOCAN 2022, about 2.3 million new cases of BC and 0.666 million BC related mortalities occurred worldwide. These figures accounted for 23.8% of the cancer cases and 15.4% of all cancer related deaths in women [5-9]. The progression and manifestation of cancer are influenced by the genotypes of tumor cells, which correspond to six distinct phenotypes associated with physiological changes [10]. Among these, one important hallmark is the induction of angiogenesis that introduces visible differences in the tumor and normal cells blood vasculatures [11]. It plays important roles in the progression and dissemination of BC in which neovascularization by pro- and anti-angiogenic factors such as fibroblast growth factor, vascular endothelial growth factor (VEGF), interleukin, transforming growth factor-β, platelet-derived growth factor and several others have been implicated [12-14]. In normal cells, the blood vessel plexus represent cohesive structures that evenly distribute nutrients and oxygen to all the cells. However, following angiogenesis, new blood vessels sprout from the existing ones by solid tumors that facilitate the growth, and metastatic spread of the cancer cells by transporting the required nutrients, oxygen, and immune cells [14-16].

The complexity of breast cancer's development arises from a multitude of factors, including genetic predispositions, hormonal influences, environmental exposures, and lifestyle choices. Despite the substantial amount of research conducted on breast cancer, the identification of specific genetic markers linked to its susceptibility continues to be an area of active investigation [17-19]. One such genetic factor is the vascular endothelial growth factor A (VEGFA), which plays a crucial role in angiogenesis, the process of new blood vessel formation. Angiogenesis is a critical mechanism that supports tumor growth and metastasis by providing the tumor with essential nutrients and oxygen [20-23]. Vascular endothelial growth factor A (VEGFA) has been found to be highly expressed in tumor cells that

might stimulate the process of angiogenesis specifically in the vascular endothelial cells. VEGFA is a highly polymorphic gene that is located on chromosome 6p21.3 and consists of seven introns and eight exons [24-26]. Important role of more than 1000 VEGFA SNPs within the gene inclusive of the upstream and downstream sequences have been found in various diseases. Among these, rs3025039 (+936C>T), rs2010963 (+405C>G), rs833061 (-460T>C), rs699947 (-2578C>A) have been identified as potential BC susceptibility gene polymorphisms [27-31]. -2578 A > C (rs699947) is an intergenic polymorphism that is present 5' upstream of the transcription start site. This SNP region has been identified as an important transcription regulator with significant involvement in promoter and enhancer activity for VEGFA expression [32-36].

Among the various polymorphisms within the VEGFA gene, the rs699947 (-2578 A>C) variant has attracted attention due to its potential involvement in regulating VEGFA expression and, consequently, its influence on cancer progression. However, studies investigating the association between this polymorphism and breast cancer have yielded conflicting results across different populations [37-44]. These discrepancies highlight the need for region-specific research to better understand the genetic factors that contribute to breast cancer risk. In particular, while many studies have focused on Western or Asian populations, there is a lack of comprehensive genetic research on Pakistani women. The present study focused on identification of VEGF genotypic and allelic associations with breast cancer and validate findings by using In-silico tools

2. METHODOLOGY

This section contains the following sub-sections:

2.1 Study population

This case–control study was conducted to study the association of VEGFA gene polymorphism rs699947 in 100 Pakistani women. A total of 50 patients with a mean age of 52.54 years with histologically confirmed BC were recruited in the medical oncology and gynecology departments of the Services Institute of Medical Sciences, Lahore, Pakistan. A structured questionnaire was designed and used to recruit patients by in-person interviews to elicit information on demographic features, gyneco-obstetric history and clinicopathological features of the patients. The control group comprising of age and gender-matched healthy individuals (n=50, mean age 52.24 years) had no history of previous or concurrent malignant diseases. The study was approved by the Ethics and Biosafety committee of the Institute of Microbiology and Molecular Genetics, University of the Punjab, Lahore, Pakistan (D/647/MMG). Data and samples were collected after obtaining written informed consent from the study participants. The research was conducted in accordance with the Declaration of Helsinki.

2.2 Sample collection and DNA extraction

After examining the clinicopathological findings of the recruited patients and age and gender matched controls, 3ml peripheral blood samples were collected in EDTA tubes by venipuncture by an experienced phlebotomist. Genomic DNA was extracted following the phenol-chloroform DNA extraction protocol by Sambrook and Russell [45]. The extracted DNA was dissolved in nuclease-free water and stored at -20°C until further analysis.

2.3 Instrument Detail

DNA was extracted from peripheral blood samples using the phenol-chloroform method as described by Sambrook and Russell (2006). The extraction was carried out using Thermo Fisher Scientific's Nalgene 50 ml tubes and processed using an Eppendorf Centrifuge 5804R. PCR amplification was performed on an Applied Biosystems 7500 Real-Time PCR System.

2.3 Genotyping of VEGFA polymorphism rs699947 (-2578A>C)

VEGFA (-2578 A > C) genotyping was performed using TETRA Amplification refractory mutation system (TETRA-ARMS) PCR. For the analysis of the polymorphism, sequence-specific outer forward and reverse (OF and OR) and internal forward and reverse (IF and IR) primers were used in table 1. These allowed the amplification of targeted DNA sequence bearing the polymorphism. The amplification products were visualized by gel electrophoresis.

Table 1: TETRA-ARMS PCR primers for detection of VEGFA polymorphism rs699947 (-2578A>C)

Sr. No.	Primers	Sequence (5' to 3')	Temperature (°C)
1	Outer Forward	CGAGTCACGAATGATGGAAAGGGAG	66
2	Outer Reverse	AAGGCCCCATCCATTCTTGCATATAGG	63
3	Inner Forward	GCCAGCTGTAGGCCAGACCCTGGTAA	67
4	Inner Reverse	CCAGTCAGTCTGATTATCCACCCAGAC	68

2.4 Primer Designing and Software Analysis:
Primers for VEGFA rs699947 were designed using Primer3 software. The genotyping was performed using TETRA-ARMS PCR, and data analysis was carried out using IBM SPSS Statistics version 21. Bioinformatics tools such as CADD and RegulomeDB were used to assess the pathogenicity of the variant.

2.5 Statistical Analysis
The data obtained from the cases and controls was subjected to detailed statistical analysis using IBM SPSS version 21. The subject characteristics including means and frequencies of clinicopathological parameters were determined by descriptive analysis. The genotypic and allelic frequencies of both the groups were calculated. The differences in the means of the genotype and allele frequencies of both the case and control groups were determined by independent sample T test. Association analysis and determination of Hardy-Weinberg Equilibrium (HWE) by the allelic frequencies was done by Chi-square analysis. Odds ratio (OR) at 95% confidence interval (CI) for both the groups and contrasting genotype models was calculated to determine the association between the genotypes and alleles with the risk of the development of BC. Correlation analysis was done to determine the relationship between genotypes and all the clinical parameters. The level of significance was kept at $p \leq 0.05$.

2.6 Bioinformatics Analysis
The functional and pathogenicity analysis of the polymorphism was done by using in-silico tools including Combined Annotation Dependent Depletion (CADD), RegulomeDB, and HaploReg v4.2. Scores were provided for Genomic Evolutionary Rate Profiling (GERP), ENCODE, Chromatin State Segmentation (cHmm), RegulomeDB, pathogenicity, conservation matrix and evolutionary constraints. Molecular docking and protein-protein network analysis was done by using Auto dock Vina (v1.5.7), MGL-Tools (v1.5.7), PyMol (v2.5.2), Discovery Studio (v2025), and Cytoscape (v3.10.3). 3D structure of VEGFA was retrived from NCBI-PDB (PDB ID 5DN2 https://www.rcsb.org/structure/5DN2) and compounds from PubChem (https://pubchem.ncbi.nlm.nih.gov/) in SDF format which were later converted into 3D format online. Binding affinities of natural compounds were recorded as kcal/mol.

3 RESULTS
The result section contains the descriptive analysis, genotyping and association analysis, genetic contrast models, Association of Clinical Parameters with rs699947 Genotypes, Functional Analysis, Pathogenicity Analysis, STRING network analysis of VEGFA, Molecular docking VEGFA inhibition analysis and Constraint Analysis of VEGFA rs699947 in the following sub-sections.

3.1 Descriptive Analysis
Association of VEGF gene polymorphism rs699947 with the occurrence of breast cancer (BC) among Pakistani women was studied. The study population consisted of 100 individuals divided in two groups i.e., a control and a case group. BC patients (n=50) made up the case group with a mean age of 52.54 ± 12.14 years and healthy individuals (n=50) made up the control group with mean age 52.24 ± 10.21 years. BC related diagnostic information was collected from the patients which included the histological type of the carcinoma, tumor grade, nodal involvement, and site of biopsy as shown in figure 1. Among the histological type, recruited patients presented with invasive ductal, invasive mammary, mixed tubular, and undefined types of carcinomas. Tumors were graded as well-differentiated cells (grade 1), moderately differentiated cells (grade 2), and poorly differentiated cells (grade 3). Depending on the nodal involvement grades, the patients were sub categorized as N0 (grade 0) with no involvement of lymph nodes, N1 (grades 1-3) with minor effect on lymph nodes, N2 (grade 4-9) lymph nodes affected, and N3 (grade 10 or more) indicating severely affected lymph nodes with spread to distant nodes as well.

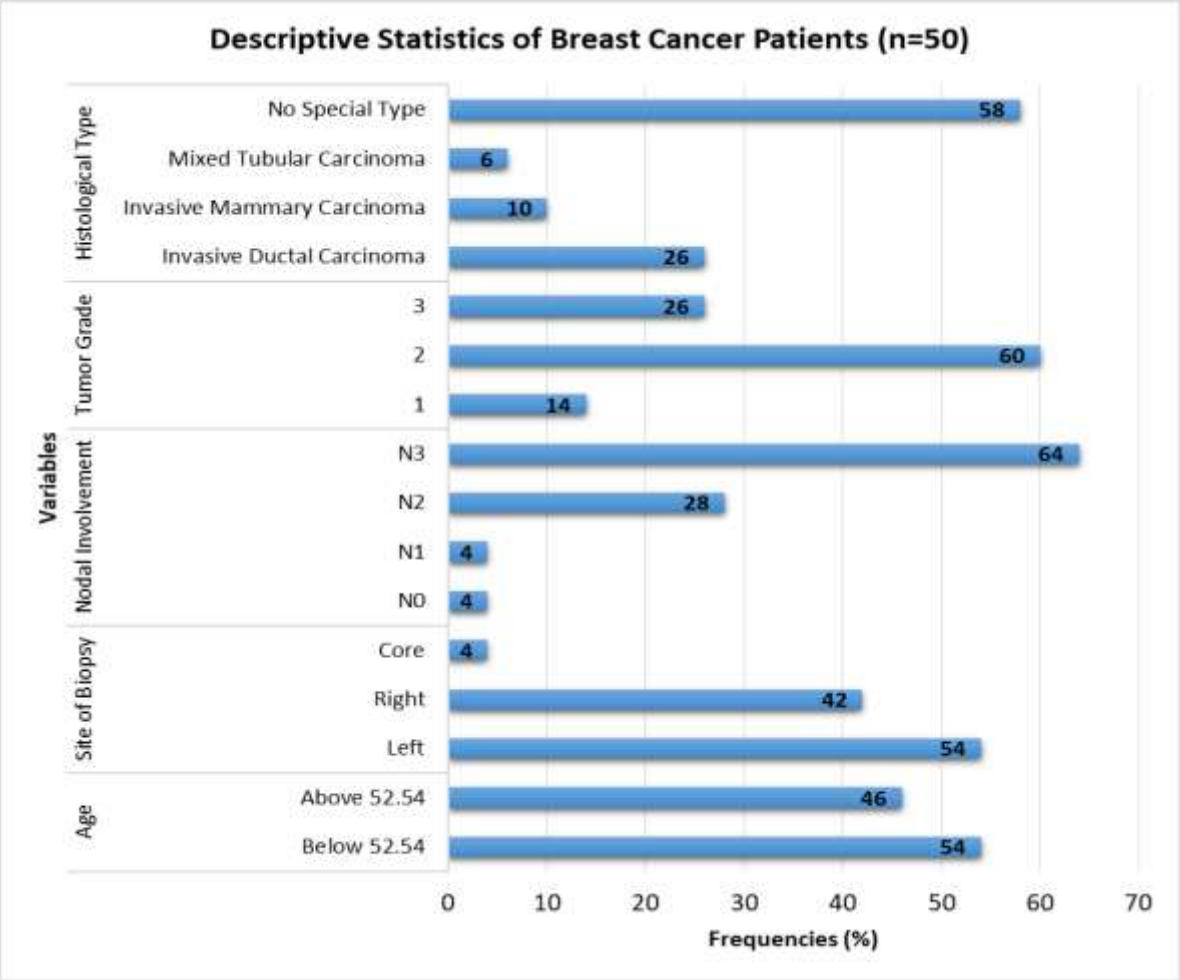


Figure 1: Frequencies of the diagnostic parameters of breast cancer patients (n=50) included in the study

3.2 Genotyping and Association Analysis

Following the blood sample collection from both the patients and controls, ARMS-PCR based genotyping was done to identify the rs699947 genotypes in the study population and estimate its relevance with development of BC. Homozygous dominant genotype AA (ancestral), homozygous recessive CC (variant), and heterozygous (AC) were found in both groups as given in the table 2. Fisher's exact test was applied to the genotypic distributions in both the cases and controls to determine the deviations from Hardy-Weinberg Equilibrium (HWE). The p-value 0.293> 0.05 suggested no significant deviation from the expected HWE proportions indicating genotypic consistency in the studied groups. This indicated that the polymorphism under investigation did not show signs of selection, genotyping errors, or population stratification. This supported the assumption that the study population was genetically stable and minimal involvement of rs699947 genotypes were found in the development of BC.

Table 2: Fisher's Exact test for Hardy-Weinberg equilibrium (n=100)

Subjects	Genotypes			Fisher's Exact Test
	A/A	C/C	A/C	
Cases (n = 50)	20	16	14	0.293
Controls (n = 50)	28	12	10	

The allelic and genotypic frequencies were then used to find out the association of BC with different genotypes. The probabilities of the occurrence of the disease were calculated by OR at 95% CI as given in the table 3. In control group, the A allele (0.66) and AA genotype (0.28) were found to be more frequent than in cases (A= 0.54; AA=0.16), suggesting a possible protective effect of the ancestral allele. The OR value 1.49 suggested a 49% increased risk of breast cancer for individuals carrying the C allele as compared to those carrying the A allele. However, CI (0.62–3.59) crossed the threshold value 1 which indicated no statistical significance. The association analysis using the chi-square

test provided with a p-value of 0.504 > 0.05 which indicated that insignificant association was present between this polymorphism and breast cancer risk among the studied population.

Table 3: Association analysis of breast cancer risk with different allelic and genotypic frequencies among the cases and the controls

Alleles/ Genotypes	A	C	A/A	C/C	A/C	Odds Ratio (95% CI)	p-value
Cases	54 (0.54)	46 (0.46)	20 (0.20)	16 (0.16)	14 (0.14)	1.49 (95% CI 0.62−3.59)	0.504
Controls	66 (0.66)	34 (0.34)	28 (0.28)	12 (0.12)	10 (0.10)		

The lack of association was validated by independent sample T-test which was performed to assess the variations in the means of both the controls and case groups with respect to their genotypic frequencies. The p-value of 0.143 > 0.05 indicated that insignificant differences were present in the means of both the groups.

3.3 Genetic Contrast Models

Different genetic contrast models; Allelic, Dominant, Recessive, Codominant, Over-dominant, were developed and compared. The model fitting information for these genetic models was analyzed which compared the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Likelihood Ratio across these models depicted in the figure 2. The Dominant Model appeared to be the best fit for the data as it had the lowest AIC, BIC, and Likelihood Ratio values. On the other hand, the codominant model was found to be the worst fit among all the models as it had the highest AIC, and BIC values. Contrarily, the Allelic, Over-dominant, and Recessive models had similar intermediate performances. Therefore, to ensure the best balance of fit and complexity, the Dominant model was selected as the best fit among all models.

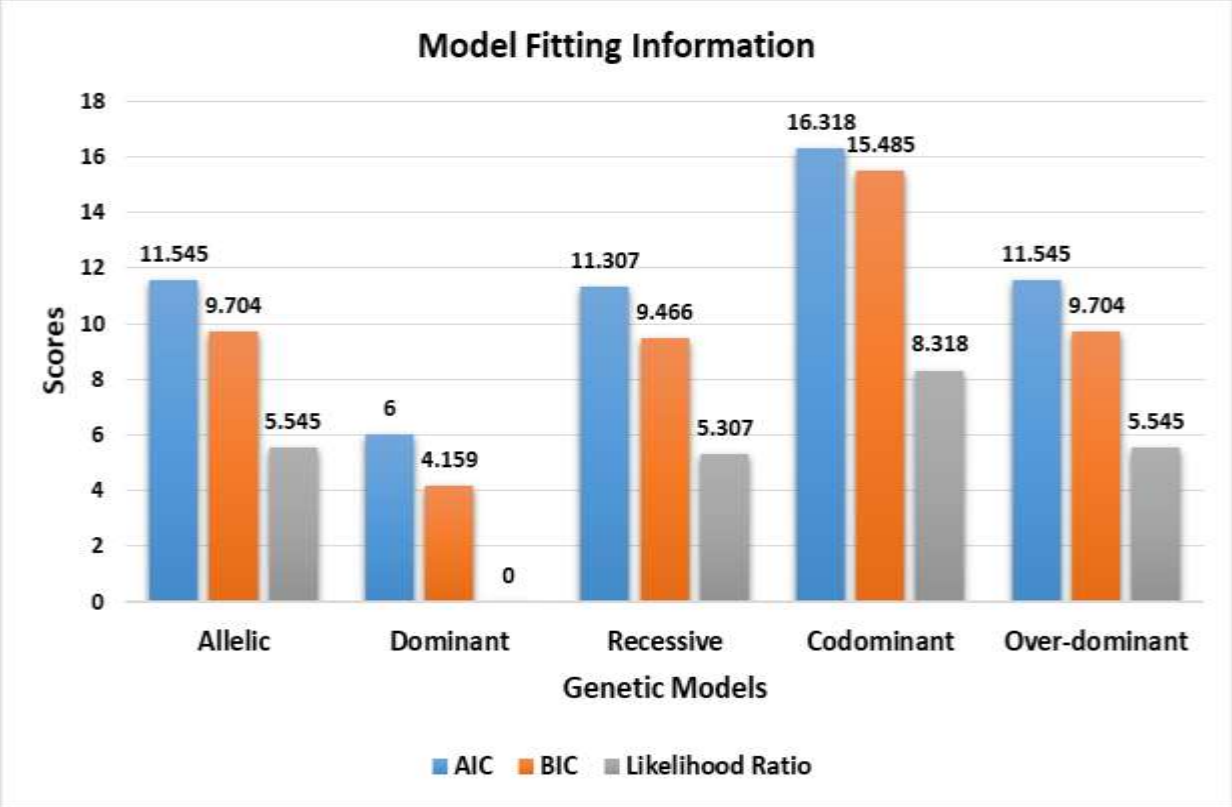


Figure 2: The model fitting information for genetic models (Allelic, Dominant, Recessive, Codominant, Over-dominant), comparing the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Likelihood Ratio.

The genetic contrast models were further used to assess the probability of the occurrence of BC in the table 4. The genotype groups for cases and controls, OR with 95% CI, and p-values for various genetic models were found. No statistically significant association was found between the VEGF rs699947 polymorphism and the occurrence of BC in any genetic model. Dominant and recessive models showed slightly increased OR values, however, their CI were wide, and p-values suggested no strong evidence of association. Over-dominant model suggested potential protective effects, however, none of the model played role in the development of the disease in association with the polymorphism under investigation. Thus, although the dominant model was found to be the best fit, however, no strong genetic association was found between rs699947 and BC risk in this population. This highlighted the need of further studies with larger sample sizes to confirm these findings and detect any potential minor effects that might not have been observed due to limited statistical power with the current dataset.

Table 4: Genetic contrast models for VEGF gene rs699947

Model	Alleles/Genotypes	Genotype Groups		OR (95% CI)	p-value
		Cases	Controls		
Allelic	A	54	66	0.605 (0.342 - 1.070)	0.112
	C	46	34		
Dominant	A/A + A/C	12	38	0.671 (0.278 - 1.618)	0.504
	C/C	16	34		
Recessive	A/A	22	28	0.524 (0.237 - 1.160)	0.161
	A/C + C/C	30	20		
Codominant	A/A	20	28	-	0.276
	C/C	16	12		
	A/C	14	10		
Over-dominant	A/C	14	10	1.556 (0.615 - 3.935)	0.482
	A/A + C/C	36	40		

3.4 Association of Clinical Parameters with rs699947 Genotypes

Correlation Analysis of the clinical parameters was done to quantify the relationship of different variables used in the study. Significant correlation of nodal involvement was found with tumor grade ($0.032 < 0.05$) and genotypes ($0.002 < 0.05$) of the BC patients as shown in the figure 3. However, no other significant relationships were found between other parameters.

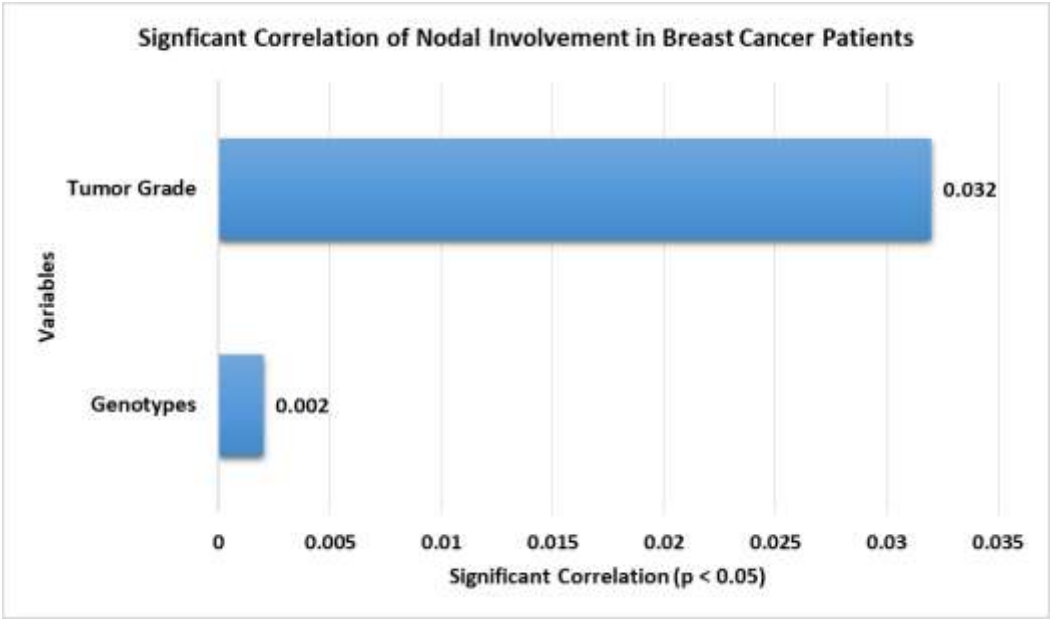


Figure 3: Correlation analysis of clinical parameters with significant correlation of Nodal involvement with tumor grade and genotypes across the BC patients

For further analysis, the association of the genotypes (AA, CC, and AC) was determined with different clinical BC parameters including nodal involvement, site of biopsy, tumor grade, and histological type shown in the figure 4. Overall, the nodal involvement and tumor grade showed some genotype-dependent variation across different groups, especially for AC genotype. Site of biopsy and histological type appeared to be unaffected by genotype. Contrarily, outliers in some groups suggested individual variations that could be clinically relevant.

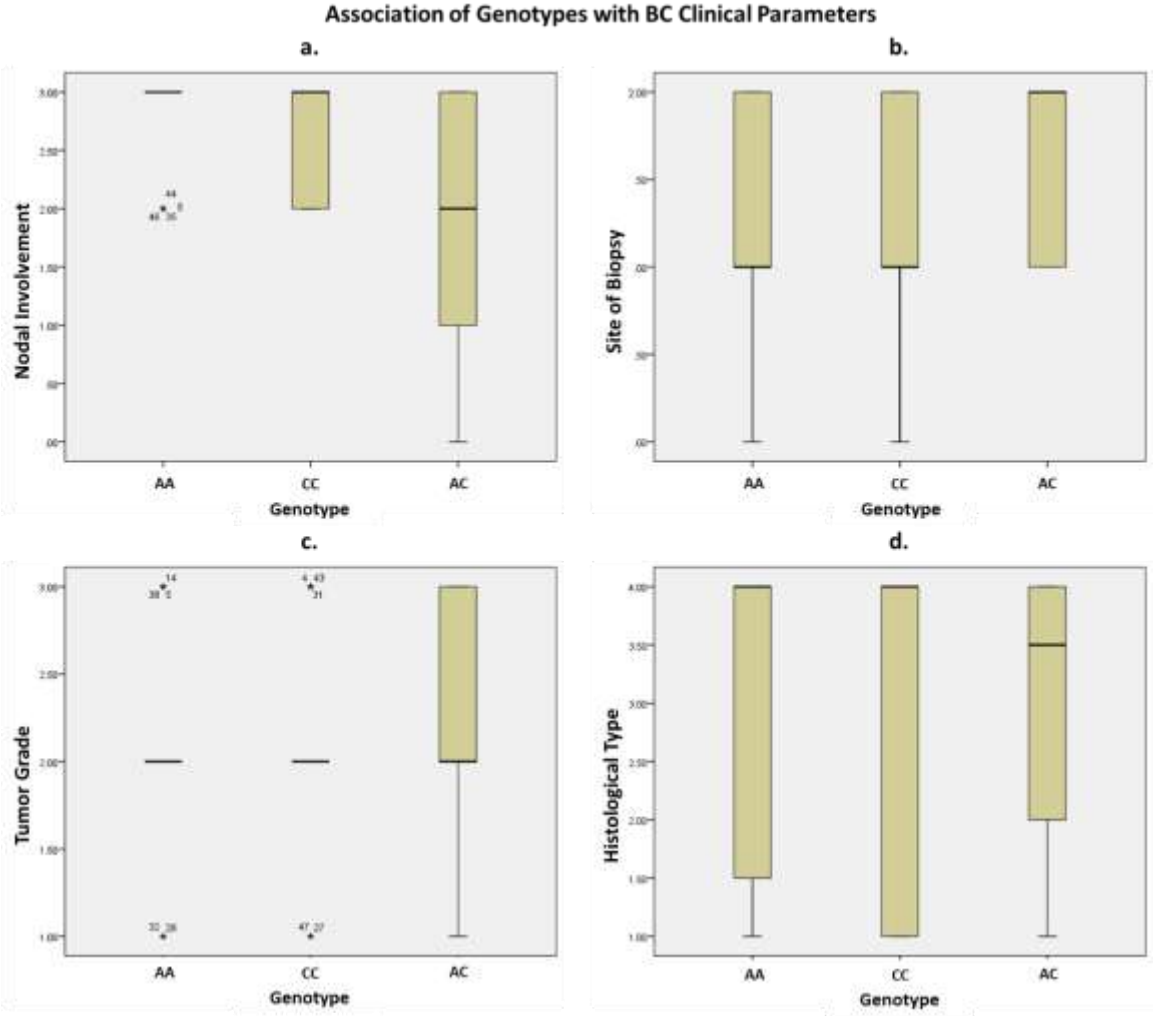


Figure 4: Association of genotypes (AA, CC, and AC) with different clinical BC parameters including (a) Nodal Involvement; (b) Site of Biopsy; (c) Tumor Grade, and (d) Histological Type.

AA and CC genotypes showed higher nodal involvement, with minimal variability. On the other hand, AC genotype exhibited more variability, with some individuals showing no nodal involvement while others indicated higher values. Outliers were present in AA and AC groups, suggesting some cases having extreme values shown in fig 4a. With respect to the biopsy sites, the distribution appeared to be similar across all three genotypes. No apparent differences appeared in median values, suggesting that the genotype does not significantly affect the biopsy site as shown in the fig 4b. AA and CC genotypes showed a uniform tumor grade, with minimal variation. AC genotype exhibited a wide range of tumor grades, suggesting potential heterogeneity. Contrarily, outliers were present in all groups that indicated the presence of some extreme cases as depicted in fig 4c. In the association of histological type with genotypes, AC genotype showed a slightly higher median histological type. However, the distribution was similar across all genotypes that suggested no strong genotypic effect on the histological type as shown in the fig 4d.

3.5 Functional Analysis

The functional role of rs699947 was further predicted by bioinformatics analysis to predict its relevance with the breast cancer risk in the studied population. The tools including CADD, RegulomeDB, and HaploReg v4.2 were used. Functional scores for GERP, ENCODE, Chromatin cHmm, and RegulomeDB parameters were found in table 5. The high GERP scores and histone modification marks suggested that this polymorphism could have influenced the

regulatory elements while potentially affecting gene expression. The cHmm values represented chromatin segmentation states, where different numerical scores corresponded to specific regulatory elements like enhancers, promoters, and heterochromatin. These values indicated that presence of this polymorphism in a regulatory region, potentially affecting transcription factor binding or enhancer activity. RegulomeDB score also represented the likelihood of this variant of having the regulatory function, however, it was not found to be among the strongest predicted functional variants involved in the development of BC.

Table 5: Different functional parameter scores found by tools including Combined Annotation Dependent Depletion (CADD), and RegulomeDB.

Parameter	Value	Description
GerpN (Neutral Score)	6.16	Strong evolutionary constraint suggesting the region is functionally important.
GerpS (Selection Score)	2.66	Presence of evolutionary constraint and possible functional importance.
EncodeH3K4me1-3	Various	Indication of active enhancers and promoters in regulatory regions.
EncodeH3K9ac	30.05	Strong mark associated with active promoters and transcription.
EncodeH3K27ac	35.96	High enhancer activity, suggesting that this region is transcriptionally active.
cHmm (Chromatin State Segmentation)	E4 = 3, E12 = 3, E13 = 7, E14 = 5, E15 = 18	Histone modification and chromatin accessibility states across different tissues, suggesting possible regulatory activity in different chromatin states (e.g., enhancers or promoters).
RegulomeDB Score	0.55436	0.55436 is a moderate to low regulatory potential score suggesting that this polymorphism may have some transcription factor (TF) binding or regulatory evidence. Therefore, it might not be among the strongest predicted functional variants.

3.6 Pathogenicity Analysis

The pathogenicity analysis revealed that this polymorphism had a low but non-negligible functional impact as given in the table 6. The raw pathogenicity and PHRED scores suggested that it might not be highly deleterious, but it could still play a regulatory role.

Table 6: Pathogenicity analysis by Combined Annotation Dependent Depletion (CADD) indicating less deleterious nature of rs699947 polymorphism

Parameter	Value	Description
Roulette-FILTER	low	A low predicted impact on function.
Roulette-MR	0.02	A minor predicted regulatory effect.
RawScore	0.67358	The raw pathogenicity prediction score.
PHRED	7.142	PHRED-scaled score indicating less deleterious nature of the polymorphism rs699947.

3.7. STRING network analysis of VEGFA

The VEGFA protein-protein interaction investigates this specific protein located within the promoter region of the VEGFA gene. The genetic variations affect the VEGFA expression levels since different alleles display various transcriptional activities. An elevated VEGFA expression rate because of this polymorphism would boost tumor blood vessel development thus ensuring tumors receive vital oxygenated blood and necessary nutrients that allow cancer growth along with metastasis evolution. The decrease of VEGFA expression levels negatively impacts angiogenesis processes which in turn could restrict tumor growth. The STRING network demonstrates why VEGFA matters through its discovery of links between this protein and fundamental proteins involved in breast cancer processes.

The receptor delivers signals that determine most of the pro-angiogenic actions of VEGFA. Endothelial cell survival along with migration and proliferation become possible when VEGFA binds to KDR receptors. The amounts of VEGFA can be modified by genetic variants that subsequently affect how KDR gets activated which controls tumor blood vessel formation. Enzymes degrade the extracellular matrix through which cancer cells locate themselves to invade other tissue structures and spread throughout the body. The expression of MMP becomes elevated through the

effects of VEGFA which creates a link between new blood vessel formation and metastatic behavior. Increased VEGFA production resulting from rs699947 could boost MMP activity which would intensify breast cancer aggressiveness. This endogenous inhibitor of angiogenesis functions to block the signaling of VEGFA. Network data shows that this factor maintains a regulatory relationship between pro-angiogenic and anti-angiogenic factors that occur in tumors. An increase in VEGFA expression due to rs699947 genotypes would potentially overpower THBS1 inhibitory influence on tumor angiogenesis.

The proteins support ECM modification and control the release of angiogenic regulatory substances (COL18A1 breaks down into the anti-angiogenic endostatin). The connection between VEGFA and these ECM proteins supports a complex matrix-stability and vascular-proliferation relationship in breast cancer tumors. The cell-surface molecules control the signaling properties of growth factors by facilitating receptor interaction with VEGFA. Modifications in VEGFA protein levels can modify the receptor-mediated interactions which subsequently affects the expression of angiogenic effects. STRING network visualizes at a systems level that VEGFA rs699947 variant affects breast cancer progression through protein interactions in the body. Elevations in VEGFA levels resulting from the polymorphism would potentially improve angiogenic processes through KDR and MMPs as well as additional protein partners to produce worse clinical results. Patients with high-risk genotypes would receive better results from treatments aimed at VEGFA or its interacting proteins including bevacizumab and MMP inhibitors as shown in figure 5.

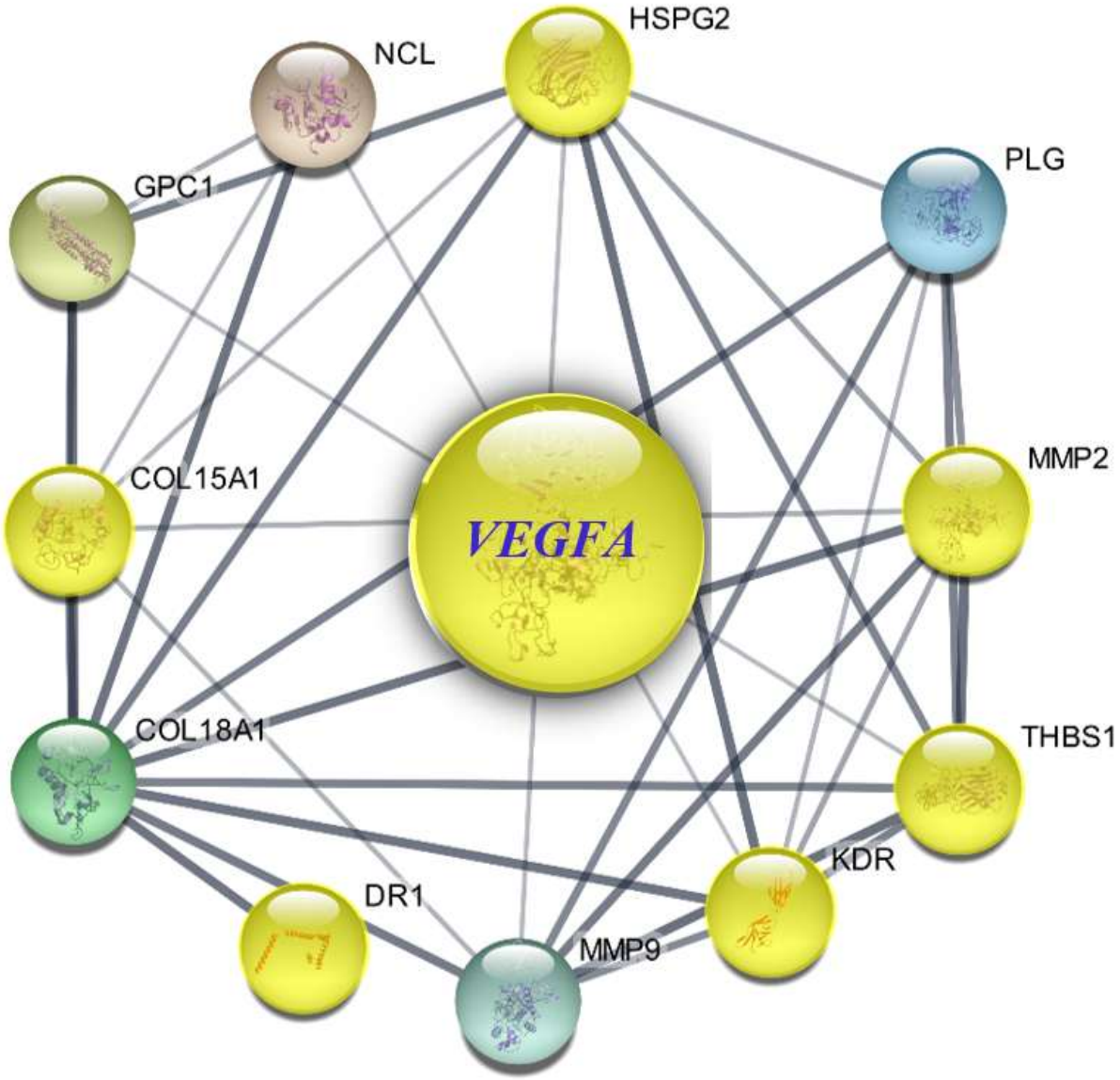
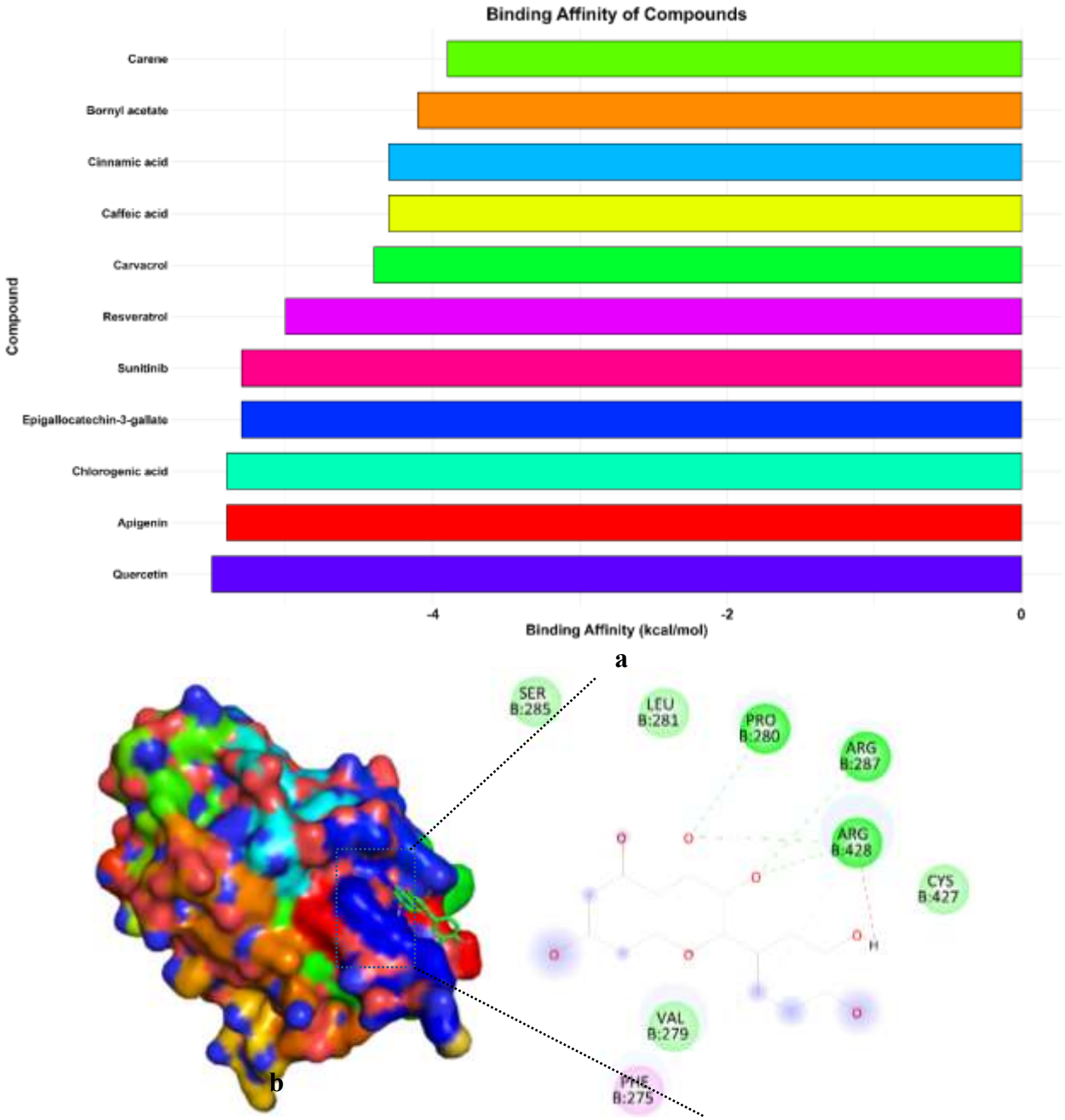


Figure 5: The STRING network displays functional protein relationships through its connections. The central element of the network functions as VEGFA which serves as the key element for controlling angiogenesis required for breast cancer tumor progression and metastasis

3.8. Molecular docking VEGFA inhibition analysis

Analyses of VEGFA binding showed natural compounds possessed different affinity levels when docked. Natural compound Quercetin achieved the most powerful binding affinity of -5.5 kcal/mol among the studied natural agents while Apigenin and Chlorogenic acid followed with -5.4 kcal/mol. The interaction strength between Epigallocatechin-3-gallate and Sunitinib as VEGFA inhibitors was comparable with an affinity of -5.3 kcal/mol. The chemical binding strength of Resveratrol appeared at -5.0 kcal/mol but Bornyl acetate and Carene exhibited the lowest binding forces at -4.1 kcal/mol and -3.9 kcal/mol respectively. Natural compounds study findings demonstrate that Quercetin together with Apigenin and Chlorogenic acid exhibit VEGFA inhibitory capabilities which match or surpass those of Sunitinib. Additional experimental studies must verify that these compounds are suitable candidates for use in cancer treatment and anti-angiogenesis therapy as shown in figure 6.



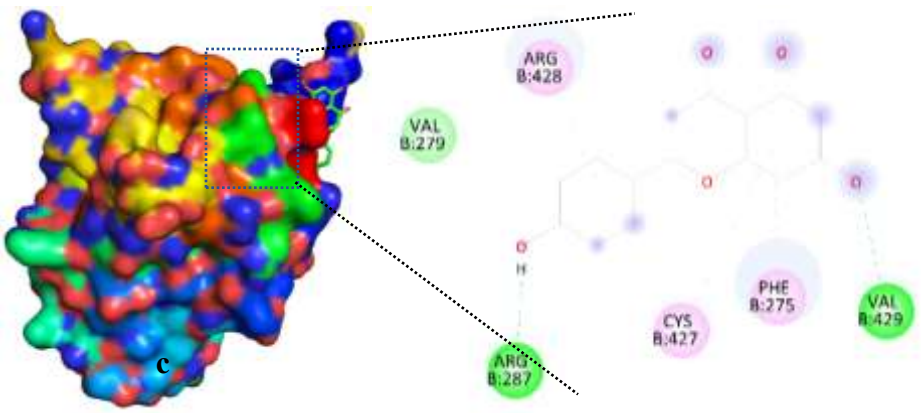


Figure 6: (a) Binding affinity of natural phytochemicals against VEGFA and 3D structure of VEGFA bound with **(b)** quercetin (-5.5 kcal/mol) and **(c)** apigenin (-5.4 kcal/mol), PyMol view (right), and Discovery Studio Mol2 view (left)

3.9 Constraint Analysis of VEGFA rs699947

The ancestral allele in VEGFA rs699947 polymorphism is an adenine (A) however, in different population, it may be altered into a cytosine (C) or a thymine (T) [46]. The population frequency and mutation proximity evaluated by CADD analysis revealed that the mutation may be present nearby multiple SNPs, suggesting that this region could be a hotspot for genetic variation and may contribute to disease risk. However, RegulomeDB revealed the presence of 208 peaks suggesting the overlap of this SNP overlaps with 208 experimentally identified regulatory elements further endorsing the presence of high regulatory activity in the region of SNP. Evolutionary constraint analysis and conservation scores were also determined for the polymorphism is enlisted in the table 7. The low PhCons scores suggested that the region is not highly conserved across species, but the PhyloP values indicated that some functional constraint, especially in vertebrates was present.

Table 7: Evolutionary constraint analysis and conservation scores of VEGFA rs699947 polymorphism

Parameter	Value	Description
priPhCons	0.028	Primate-specific conservation score revealing weak evolutionary conservation.
mamPhCons	0.001	Mammalian conservation score, suggesting little evolutionary constraint.
verPhCons	0.001	Low vertebrate conservation score indicating that the region has not been highly conserved through evolution.
priPhyloP	0.475	A moderate value suggests mild evolutionary conservation.
mamPhyloP	0.7	Slightly higher conservation.
verPhyloP	0.889	Indicating some level of evolutionary importance.

Based on the RegulomeDB ranking system and functional evidence, this variant was classified as having rank 1f. This confirmed the presence of rs699947 in a transcription factor binding site (TFBS). It also revealed its possible role in disrupting or modifying a known DNA sequence motif. Analysis also revealed that it might overlap with an experimentally validated DNase peak, indicating open chromatin and potential regulatory function. However, it was not found to be associated with eQTLs, indicating that it has not been strongly linked to gene expression changes in different population studies. Therefore, although this variant may alter transcription factor binding and chromatin accessibility, potentially modulating VEGFA expression however, it might not be classified as a highly impactful variant that it could contribute to gene regulation and lead to the development of BC as confirmed by the genotyping analysis in this population under study.

4. DISCUSSION

The present study investigated the association of the rs699947 polymorphism present upstream of the VEGFA gene with the BC risk. Comparative genotyping analysis was done and the experimental results were validated by in-silico functional and pathogenicity analysis. VEGFA rs699947 is an intergenic SNP that is located upstream of the gene. VEGFA gene, controlled by various regulatory elements, is an important gene that plays role in angiogenesis. However, disruption in the regulation may lead to dysregulated expression [47]. CADD analysis for rs699947 revealed its role in the promotor activity as it is present 1533 bp upstream of the Transcription Start Site (TSS) and 1778 bp upstream of the Transcription End Site (TSE). RemapOverlapTF and RemapOverlapCL suggested its overlap with

four transcription binding sites and five chromatin landmarks. These predictions indicated its presence at transcription factor binding sites, suggesting its regulatory functions. This increased the investigative relevance of rs699947 in this population. Genotyping analysis revealed the presence of ancestral "AA," variant "CC," and heterozygous "AC" genotypes in both the case and control groups. Higher frequencies of the wildtype "A" allele and "AA" genotype were present in the control as compared to the cases. However, the genotyping and association analysis indicated insignificant role of the homozygous dominant and recessive, or the heterozygous genotypes in the increased risk of BC in the studied population. Independent sample T test validated the results by indicating no significant differences present in the means of the genotypes of both the groups. The results were consistent with findings reported by Wang et. al. [48] in their meta-analysis involving 16703 individuals (8175 BC cases and 8528 controls) and Kumar et. al. [49] investigating 2369 BC cases and 2584 controls to determine BC risk with five VEGFA polymorphisms including rs699947. However no associations were found in their studies.

Madrid-Paredes et. al. [50] also reported no risk of BC in association with this polymorphism by examining the genotypes of 84 BC cases and 119 controls of Spanish origin. The genotypic frequencies in our study were in HW equilibrium, which was found similar to study conducted BC related rs699947 investigation conducted by Furriol et. al. [51]. In the genetic contrast models, although no model was found to be in significant association with the increased BC risk, however the dominant model (AA + CA) was found to be the best fit. Similar findings given by Vieira-Monteiro et. al. [52] on Brazilian BC patients, have revealed AA + CA model to be in significant association with high-grade (G2 + G3) tumors and with shorter disease free survival among the BC patients. Contrastingly, Rahoui et. al. [53] reported that women carrying rs699947 (AC + AA) genotypes had a reduced risk to develop BC and Al-Mohaya et. al. [54] depicted the recessive model (CC + CA) to be in strong association with the development of BC.

The STRING analysis demonstrates that VEGFA activates angiogenic partners which also remodel extracellular matrix (ECM) and activate growth factor signaling through KDR (VEGFR2) and MMP2/9 collaboration with THBS1 and PLG as well as COL18A1/COL15A1. This investigation establishes VEGFA operates as a basic requirement for cancerous tissue vascularization while it stimulates metastasis progression. Molecular docking experiments demonstrated that Quercetin together with Apigenin and Chlorogenic acid should be used since they showed effective VEGFA binding properties. The inhibitor activity of natural compounds was comparable to or greater than that of Sunitinib (-5.3 kcal/mol) indicating their potential use as therapeutic agents.

The combination of STRING network analysis and docking results improves the credibility of natural VEGFA inhibitors. STRING results support KDR as the principal VEGFA receptor because it functions as an essential transmission channel for angiogenesis processes. The molecular docking process confirmed that Quercetin (-5.5 kcal/mol) followed by Apigenin (-5.4 kcal/mol) exhibited better free energy interactions than Sunitinib (-5.3 kcal/mol) thus demonstrating their capability to block the VEGFA-KDR binding sites. The antitumoral activity of Epigallocatechin-3-gallate (EGCG; -5.3 kcal/mol) enables the compound to simultaneously bind to MMP2/9 for anti-ECM breakdown activity and suppress VEGFA-mediated metastasis activators. As an endogenous angiogenesis inhibitor THBS1 receives strengthening from Chlorogenic acid (-5.4 kcal/mol) through its direct VEGFA interaction. The minimal strength of Carene (-3.9 kcal/mol) corresponds to its limited therapeutic application because it does not occur in VEGFA signaling pathways. The recent discoveries about breast cancer have proved beneficial to its medical treatments. The natural compounds Quercetin and Apigenin bind better to VEGFA than Sunitinib yet maintain their potential as effective anti-angiogenic therapeutic agents. The natural compounds demonstrate potential for effective use as treatment agents that generate fewer adverse side effects than synthetic inhibitors. Natural compound EGCG had multiple interaction targets because it blocked VEGFA while targeting its interacting factors including MMPs to provide better tumor control. New knowledge about rs699947 polymorphisms in VEGFA genetic variations proves that healthcare strategies should be tailored specifically to individual patients. Special therapeutic methods require immediate development for European populations who carry high-risk alleles since their reaction differs from standard natural inhibitor responses.Experimental testing needs to be carried out in future studies to confirm present research findings. Laboratory investigations implementing both cellular assessments and animal testing must confirm that the best performing natural elements effectively obstruct VEGFA-KDR signaling routes. Breast cancer patient treatment responses toward compounds can be advanced by clinical research that links rs699947 genotypes with their corresponding breast cancer patient responses. The overall results of THIS study suggested no conclusive role of VEGFA rs699947 polymorphism in breast cancer predisposition. While our study does not support a strong genetic influence of rs699947 on breast cancer, however, the findings highlight the need for larger cohort studies and functional analyses to better understand its biological relevance in tumor progression and angiogenesis.

CONCLUSION

This study investigated the association between the VEGFA rs699947 polymorphism and breast cancer (BC) risk in a Pakistani population. Genotyping and statistical analyses revealed no significant association between rs699947 and BC risk. Functional analysis suggested that although rs699947 may play a role in VEGFA gene regulation, its impact on BC progression is minimal, as indicated by the moderate to low regulatory potential scores. Despite this, the study

emphasizes the need for larger cohort studies to validate these findings and explore any potential minor effects that may not have been detected due to the limited sample size. Additionally, natural compounds such as Quercetin and Apigenin exhibited promising VEGFA-blocking properties, suggesting potential therapeutic value for future cancer treatments. Further experimental validation and clinical trials are required to confirm these findings and explore their therapeutic applications in BC treatment.

Acknowledgment
We would like to acknowledge all the Breast cancer patients and healthy individuals used as controls for contributing in this study.

Compliance with Ethical Standards
All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and in accordance with the declaration of Helsinki declaration.

Informed Consent
Informed consent was obtained from all individual participants included in the study.

Conflict of Interest
The authors declare no conflict of interests.

Ethical Approval
The Departmental Research ethics and biosafety committee approved the current research study (D/647/MMG).

REFERENCES
1. Xiong, Xin, et al. "Breast cancer: pathogenesis and treatments." Signal transduction and targeted therapy 10.1 (2025): 49.
2. Kavgacı, Gözde, and Sercan Aksoy. "Breast Cancer: General Overview." Managing Side Effects of Breast Cancer Treatment (2025): 3-22.
3. Jie, Huan, Wenhui Ma, and Cong Huang. "Diagnosis, Prognosis, and Treatment of Triple-Negative Breast Cancer: A Review." Breast Cancer: Targets and Therapy (2025): 265-274.
4. Hoinoiu, Teodora, et al. "Risk factors for breast cancer recurrence in postmenopausal women: a bibliometric study." Frontiers in Oncology 15 (2025): 1522713.
5. Santos, Vanessa Emanuelle Pereira, et al. "An overview about biomarkers in breast cancer: Insights into the diagnostic and prognostic significance." Clinica Chimica Acta 567 (2025): 120030.
6. Amaro-da-Cruz, Alba, Teresa Rubio-Tomás, and Ana I. Álvarez-Mercado. "Specific microbiome patterns and their association with breast cancer: The intestinal microbiota as a potential biomarker and therapeutic strategy." Clinical and Translational Oncology 27.1 (2025): 15-41.
7. Zhang, Yunmeng, et al. "Global burden of female breast cancer: new estimates in 2022, temporal trend and future projections up to 2050 based on the latest release from GLOBOCAN." Journal of the National Cancer Center 5.3 (2025): 287.
8. Bray, Freddie, et al. "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." CA: a cancer journal for clinicians 74.3 (2024): 229-263.
9. Virani, Sehar Salim, et al. "Cancer registries in Pakistan: a scoping review." The Lancet Regional Health-Southeast Asia 38 (2025).
10. Badar, Farhana, et al. "Cancer in Balochistan, Pakistan, 2020–2022: A Descriptive Study." Journal of Cancer & Allied Specialties 11.1 (2025): 12.
11. Alsowayeh, Noorah, Azfar Jamal, and Faiz Abdulaziz Alfaiz. "Molecular Modelling and Designing of Aminopyrimidine Derivatives as a Small Molecular Inhibitor of T790m/C797s Double Mutations in the Kinase Domain of Epidermal Growth Factor Receptor to Arrest Tumour Angiogenesis." Advances in Human Biology (2025): 10-4103.
12. Shi, Jinsheng, et al. "Anti-Tumor Strategies Targeting Nutritional Deprivation: Challenges and Opportunities." Advanced Materials 37.10 (2025): 2415550.
13. Nishida, Atsushi, and Akira Andoh. "The role of inflammation in cancer: mechanisms of tumor initiation, progression, and metastasis." Cells 14.7 (2025): 488.
14. Artero, Mikel Rezola, et al. "Complement and the hallmarks of cancer." Seminars in Immunology. Vol. 78. Academic Press, 2025.
15. Nia, Hadi T., Lance L. Munn, and Rakesh K. Jain. "Probing the physical hallmarks of cancer." Nature Methods (2025): 1-19.

16. Obeagu, Emmanuel Ifeanyi, and Getrude Uzoma Obeagu. "Breast cancer: A review of risk factors and diagnosis." Medicine 103.3 (2024): e36905.
17. Roy, Himadri, Shalini Bhardwaj, and Seppo Ylä-Herttuala. "Biology of vascular endothelial growth factors." FEBS letters 580.12 (2006): 2879-2887.
18. Duffy, Angela M., David J. Bouchier-Hayes, and Judith H. Harmey. "Vascular endothelial growth factor (VEGF) and its role in non-endothelial cells: autocrine signalling by VEGF." Madame Curie Bioscience Database [Internet]. Landes Bioscience, 2013.
19. Kim, D. H., et al. "Clinical relevance of vascular endothelial growth factor (VEGFA) and VEGF receptor (VEGFR2) gene polymorphism on the treatment outcome following imatinib therapy." Annals of oncology 21.6 (2010): 1179-1188.
20. Darmadi, Darmadi, Riska Habriel Ruslie, and Cennikon Pakpahan. "Vascular Endothelial Growth Factor (VEGF) in Liver Disease." Tumor Angiogenesis and Modulators. IntechOpen, 2022.
21. Furriol Palmer, Jessica, et al. "VEGFA gene variants are associated with breast cancer progression." (2024).
22. Alagizy, Hagar A., et al. "Prognosis of vascular endothelial growth factor gene polymorphism in Egyptian patients with acute myeloid leukemia." Menoufia Medical Journal 34.2 (2021): 576-581.
23. Sa-Nguanraksa, Doonyapat, and Pornchai O-charoenrat. "The role of vascular endothelial growth factor a polymorphisms in breast cancer." International journal of molecular sciences 13.11 (2012): 14845-14864.
24. Vailati, Fabiana B., et al. "The C allele of− 634G/C polymorphism in the VEGFA gene is associated with increased VEGFA gene expression in human retinal tissue." Investigative ophthalmology & visual science 53.10 (2012): 6411-6415.
25. El Founini, Younes, et al. "Single nucleotide polymorphisms: rs833061, rs699947, and rs35569394, and the expression of the vascular endothelial growth factor gene in Moroccan patients with lung cancer." Exploration of Medicine 6 (2025): 1001293.
26. Ferrara, Napoleone, and Anthony P. Adamis. "Ten years of anti-vascular endothelial growth factor therapy." Nature reviews Drug discovery 15.6 (2016): 385-403.
27. Arcondéguy, Tania, et al. "VEGF-A mRNA processing, stability and translation: a paradigm for intricate regulation of gene expression at the post-transcriptional level." Nucleic acids research 41.17 (2013): 7997-8010.
28. Liu, Changjiang, et al. "Correlation of gene polymorphisms of vascular endothelial growth factor with grade and prognosis of lung cancer." BMC Medical Genetics 21.1 (2020): 86.
29. Lin, Ling, et al. "Four common vascular endothelial growth factor polymorphisms (− 2578C> a,− 460C> T,+ 936C> T, and+ 405G> C) in susceptibility to lung cancer: a meta-analysis." PLoS One 8.10 (2013): e75123.
30. Ribatti D, Annese T, Tamma R. Controversial role of mast cells in breast cancer tumor progression and angiogenesis. Clinical Breast Cancer. 2021;21(6):486-91. https://doi.org/10.1016/j.clbc.2021.08.010
31. Xia C, Liu Y, Yong W, Qing X. Global evolution of breast cancer incidence in childbearing-age women aged 15–49 years: a 30-year analysis. Journal of Cancer Research and Clinical Oncology. 2025;151(2):1-12. https://doi.org/10.1007/s00432-025-06113-0
32. Xu T, Yu S, Zhang J, Wu S. Dysregulated tumor-associated macrophages in carcinogenesis, progression and targeted therapy of gynecological and breast cancers. Journal of hematology & oncology. 2021;14:1-20. https://doi.org/10.1186/s13045-021-01198-9
33. Zhang Y, Ji Y, Liu S, Li J, Wu J, Jin Q, et al. Global burden of female breast cancer: new estimates in 2022, temporal trend and future projections up to 2050 based on the latest release from GLOBOCAN. Journal of the National Cancer Center. 2025. https://doi.org/10.1016/j.jncc.2025.02.002
34. Farahi A, Abedini MR, Javdani H, Arzi L, Chamani E, Farhoudi R, et al. Crocin and Metformin suppress metastatic breast cancer progression via VEGF and MMP9 downregulations: in vitro and in vivo studies. Molecular and Cellular Biochemistry. 2021;476(9):3341-51. https://doi.org/10.1007/s11010-020-04043-8
35. Ghalehbandi S, Yuzugulen J, Pranjol MZI, Pourgholami MH. The role of VEGF in cancer-induced angiogenesis and research progress of drugs targeting VEGF. European journal of pharmacology. 2023;949:175586. https://doi.org/10.1016/j.ejphar.2023.175586
36. Mou J, Li C, Zheng Q, Meng X, Tang H. Research progress in tumor angiogenesis and drug resistance in breast cancer. Cancer Biology & Medicine. 2024;21(7):571-85. https://doi.org/10.20892/j.issn.2095-3941.2023.0515
37. Badodekar N, Sharma A, Patil V, Telang G, Sharma R, Patil S, et al. Angiogenesis induction in breast cancer: A paracrine paradigm. Cell biochemistry and function. 2021;39(7):860-73. https://doi.org/10.1002/cbf.3663
38. Zhang Y, Brekken RA. Direct and indirect regulation of the tumor immune microenvironment by VEGF. Journal of Leukocyte Biology. 2022;111(6):1269-86. https://doi.org/10.1002/JLB.5RU0222-082R
39. Al Kawas H, Saaid I, Jank P, Westhoff CC, Denkert C, Pross T, et al. How VEGF-A and its splice variants affect breast cancer development–clinical implications. Cellular Oncology. 2022;45(2):227-39. https://doi.org/10.1007/s13402-022-00665-w

40. Rezaei M, Hashemi M, Sanaei S, Mashhadi MA, Taheri M. Association between vascular endothelial growth factor gene polymorphisms with breast cancer risk in an Iranian population. Breast cancer: basic and clinical research. 2016;10:BCBCR. S39649. https://doi.org/10.4137/BCBCR.S39649
41. ABDULHUSSEIN HA, ALWASITI EA, KHIRO NK. The potential impact of vascular endothelial growth factor rs699947 polymorphisms on breast tumors susceptibility in a sample of Iraqi females. ACTA Pharmaceutica Sciencia. 2024;62(2). https://doi.org/10.23893/1307-2080.APS6217
42. Maghsoodi MS, Khosroshahi NS, Beilankouhi EAV, Valilo M, Feizi MAH. VEGF-634G> C (rs2010963) gene polymorphism and high risk of breast cancer in the Northwest of Iran. Indian Journal of Gynecologic Oncology. 2023;21(1):6. https://doi.org/10.1007/s40944-022-00648-7
43. Al Balawi IA, Mir R, Abu-Duhier F. Potential impact of vascular endothelial growth factor gene variation (-2578C> A) on breast cancer susceptibility in Saudi Arabia: a Case-Control Study. Asian Pacific journal of cancer prevention: APJCP. 2018;19(4):1135. https://doi.org/10.22034/APJCP.2018.19.4.1135
44. Li Z, Wang Y, Liu C, Wang Z, Wang D, Liang X, et al. Association between VEGF single nucleotide polymorphism and breast cancer in the Northern China Han population. Breast Cancer Research and Treatment. 2021;186:149-56. https://doi.org/10.1007/s10549-020-06024-3
45. Sambrook J, Russell DW. Purification of nucleic acids by extraction with phenol: chloroform. Cold Spring Harbor Protocols. 2006;2006(1):pdb. prot4455. https://doi.org/10.1101/pdb.prot4455
46. Ben Salem A, Megdich F, Kacem O, Souayeh M, Hachani Ben Ali F, Hizem S, et al. Vascular endothelial growth factor (VEGFA) gene variation in polycystic ovary syndrome in a Tunisian women population. BMC genomics. 2016;17:71-7. https://doi.org/10.1186/s12864-016-3092-5
47. Sharma P, Chida K, Wu R, Tung K, Hakamada K, Ishikawa T, et al. VEGFA Gene Expression in Breast Cancer Is Associated With Worse Prognosis, but Better Response to Chemotherapy and Immunotherapy. World Journal of Oncology. 2025;16(1):120. https://doi.org/10.14740/wjon1993
48. Wang K, Liu L, Zhu Z-M, Shao J-H, Xin L. Five polymorphisms of vascular endothelial growth factor (VEGF) and risk of breast cancer: a meta-analysis involving 16,703 individuals. Cytokine. 2011;56(2):167-73. https://doi.org/10.1016/j.cyto.2011.06.018
49. Kumar YS, Varghese S, Kulanthaivel L, Subbaraj GK. Association of VEGF polymorphisms and breast cancer susceptibility: systemic review and meta-analysis. Meta Gene. 2021;30:100946. https://doi.org/10.1016/j.mgene.2021.100946
50. Madrid-Paredes A, Casado-Combreras MÁ, Pérez-Ramírez C, Segura-Pérez AM, Chamorro-Santos C, Vergara-Alcalde E, et al. Association of ABCB1 and VEGFA gene polymorphisms with breast cancer susceptibility and prognosis. Pathology-Research and Practice. 2020;216(4):152860. https://doi.org/10.1016/j.prp.2020.152860
51. Furriol J, Wik E, Aziz S, Askeland C, Knutsvik G, Akslen LA. VEGFA gene variants are associated with breast cancer progression. The Journal of Pathology: Clinical Research. 2024;10(5):e12393. https://doi.org/10.1002/2056-4538.12393
52. Vieira-Monteiro HdA, Freitas-Alves DR, Sobral-Leite M, Delou JMdA, Goulart-Citrangulo SMT, do Nascimento CT, et al. Prognostic evaluation of VEGFA genotypes and haplotypes in a cohort of Brazilian women with non metastatic breast cancer. Cancer biology & therapy. 2016;17(6):674-83. https://doi.org/10.1080/15384047.2016.1190486
53. Rahoui J, Laraqui A, Sbitti Y, Touil N, Ibrahimi A, Ghrab B, et al. Investigating the association of vascular endothelial growth factor polymorphisms with breast cancer: a Moroccan case–control study. Medical Oncology. 2014;31:1-9. https://doi.org/10.1007/s12032-014-0193-3
54. Al-Mohaya MA, Alfadhel AK, Mustafa M, Alquwayz TS, Al-Anazi MA. Vascular endothelial growth factor (VEGF-2578 C> A) gene polymorphism as a genetic biomarker for breast cancer: a case control study. Gene Reports. 2021;22:101007. https://doi.org/10.1016/j.genrep.2020.101007