scientific reports

Check for updates

OPEN

Genetic association and computational analysis of MTHFR gene polymorphisms rs1801131 and rs1801133 with breast cancer in the Bangladeshi population

Nazia Fairooz Alam[1,3], Rubaiat Ahmed[2,3], Zimam Mahmud[1], Sonia Tamanna[1], Md Akeruzzaman Shaon[1] & Md. Zakir Hossain Howlader[1]✉

Methylenetetrahydrofolate reductase (MTHFR) plays a crucial role in regulating one-carbon metabolism. Polymorphisms within the MTHFR gene have been found to increase the risk of breast cancer in different populations. In this study, we evaluated the association of polymorphisms of the MTHFR gene (rs1801133 and rs1801131) with the risk of breast cancer in the Bangladeshi population. This case–control study included 202 breast cancer patients and 104 healthy controls. After the organic extraction of DNA, genotyping was performed via the PCR-RFLP method. Sanger sequencing was performed to validate the RFLP data. Statistical analyses were performed to evaluate the associations of the polymorphisms. Different computational tools were used to predict the structural and functional consequences of the SNPs. Our study revealed that the MTHFR gene polymorphism rs1801131 is associated with an increased risk of developing breast cancer ($p < 0.001$, OR = 3.85, 95% CI = 2.06–7.25 for the AC genotype and $p < 0.001$, OR = 7.82, 95% CI = 2.69–22.05 for the CC genotype). An association was also observed in the dominant model (AC + CC) ($p < 0.001$, OR = 4.19, 95% CI = 2.28–7.78). For rs1801131, premenopausal status was significantly associated with breast cancer risk ($p < 0.001$). For rs1801133, no significant association was found with breast cancer risk ($p > 0.05$, OR = 1.57, 95% CI = 0.90–2.74 for the CT genotype; $p > 0.05$, OR = 1.35, 95% CI = 0.36–4.92 for the TT genotype). Computational analyses predicted rs1801131 to be tolerated and rs1801133 to be deleterious. Structural analyses demonstrated no significant changes in protein structure but revealed alterations in neighboring interactions according to both bond distances and angles. In conclusion, rs1801131 but not rs1801133 is significantly associated with breast cancer risk in the Bangladeshi population. Moreover, in silico analyses demonstrated changes in the interaction pattern of polymorphic residues with adjacent amino acids.

Keywords Methylenetetrahydrofolate reductase, Genetic polymorphisms, rs1801133, rs1801131, PCR-RFLP, Bangladeshi population

Cancer is the uncontrolled, persistent multiplication of cells, which initiates at one location within the body, and in later stages, metastasizes to other body parts[1]. Breast cancer is the most common type of cancer diagnosed in women. It is regarded as a fundamental health issue worldwide and is the leading cause of cancer-related mortality in women[2]. Most breast tumors start in duct-lining cells and are referred to as ductal cancers. Most others are lobular cancers, as they develop in the cells that cover the lobules, and a tiny minority arise in other tissues[3]. By 2020, breast cancer surpassed lung cancer as the most common type of cancer in women. Breast cancer accounts for 25% of newly diagnosed cancer cases in females worldwide. Men account for less than 1% of all diagnosed breast cancers[4]. In Bangladesh, 12,989 new cases of breast cancer were diagnosed in 2022 (https://gco.iarc.who.int/media/globocan/factsheets/populations/50-bangladesh-fact-sheet.pdf). In developed

[1]Laboratory of Nutrition and Health Research, Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka 1000, Bangladesh. [2]Molecular Biotechnology Division, National Institute of Biotechnology, Ganakbari, Ashulia, Savar, Dhaka 1349, Bangladesh. [3] These authors contributed equally: Nazia Fairooz Alam and Rubaiat Ahmed. ✉email: hhzakir@du.ac.bd

countries, there are a greater number of cases of breast cancer due to various factors, such as early menarche, older age at first birth, having no children, being overweight, drinking alcohol, lack of physical activity, and breastfeeding[5].

Breast cancer has been a global matter of concern for a considerable amount of time. The identification of breast cancer risk factors is critical to the process of screening women for this disease at an early stage, which in turn improves their chances of responding favorably to curative therapy. The polymorphic distribution of enzymes, which are responsible for the activation and/or inactivation of enzymes in humans, is one of the most important aspects to consider when determining a person's susceptibility to the development of cancer[6]. The polymorphic pattern is associated with genetic markers within the genome. Therefore, it is crucial to guide therapeutic strategies by locating genetic markers associated with breast cancer development. With the help of genetic risk profiles, personalized medication can be prescribed to patients with different needs via genetic counseling[7]. The most recent findings from epidemiological research conducted at the molecular level have demonstrated that polymorphisms in the gene encoding the 5-10-methylenetetrahydrofolate reductase (MTHFR) enzyme play a role in the development of breast cancer[2].

The MTHFR enzyme is essential for cellular homeostasis because it plays critical roles in one-carbon metabolism, including methionine and folate metabolism, as well as DNA, RNA, and protein synthesis[8]. The MTHFR gene is located on chromosome 1 at position p36.3. In humans, MTHFR is a protein consisting of 656 amino acids. The N-terminal catalytic domain of human MTHFR (residues 1-356) binds 5,10-methylenetetrahydrofolate (5,10-methylene THF), and the C-terminal regulatory domain of MTHFR (residues 363–656) contains an allosteric site that binds S-adenosylmethionine (SAM) to exert allosteric inhibition[9]. MTHFR catalyzes the irreversible conversion of 5,10-methylene THF to 5-methyl THF, which acts as a methyl donor necessary for the conversion of homocysteine to methionine. SAM, the most potent methyl donor with biological activity, is then produced from methionine[10]. In addition to being a crucial part of the synthesis of the amino acid methionine, which influences protein synthesis, MTHFR is also crucial for the de novo synthesis of purine and pyrimidine nucleosides and thus for DNA repair and maintenance[11]. Since MTHFR is involved in both methylation and DNA synthesis, the route that involves this enzyme is crucial to the growth and spread of cancer. Therefore, it is important to learn how single nucleotide polymorphisms (SNPs) influence cancer risk to determine the role of MTHFR as a potential marker for cancer.

Numerous SNPs have been identified in the MTHFR gene. It has been established that some of these factors have a direct impact on the decreased activity of the enzyme encoded by this gene, leading to SAM shortages and DNA hypomethylation. Two frequent polymorphisms, rs1801133 (C677T) and rs1801131 (A1298C), have been linked to decreased enzymatic activity and elevated plasma homocysteine levels[12]. The C677T allelic variant results from the conversion of an alanine to a valine at codon 222 (A222V) in the N-terminal catalytic domain. It is associated with increased thermolability and decreased enzyme activity, which results in increased plasma homocysteine levels[13]. It has been demonstrated that homozygous TT carriers of C677T have only 30% MTHFR enzyme activity compared to the wild genotype, whereas heterozygous CT genotype carriers have 60% enzyme activity[14]. On the other hand, the A1298C allelic variant changes a glutamic acid residue to an alanine residue at codon 429 (E429A) on the C-terminal regulatory domain of the protein. However, it is still debatable whether this results in decreased enzymatic activity. This polymorphism becomes clinically significant under conditions of severe folate deficiency[15].

The objective of this study was to compare healthy individuals with breast cancer patients to assess the significance of the rs1801133 and rs1801131 polymorphisms as risk factors. The results from different studies on these polymorphisms are inconsistent in the literature because they are heavily influenced by differences among distinct ethnic groups[16]. This study also aimed to determine the structural and functional alterations in the MTHFR protein due to these polymorphisms by using different in silico tools.

Materials and methods

Study participants

This was a population-based case-control study in which breast cancer patients were considered cases and healthy individuals without a history of breast cancer or any other chronic disease were considered controls. A total of 306 individuals were enrolled in this study. Among them, 202 were breast cancer patients (cases), and 104 were age-matched healthy controls.

The Ethical Review Committee of the Department of Biochemistry and Molecular Biology approved the study at the University of Dhaka (Ref. No. BMBDU-ERC/EC/23/014). In addition, we confirm that all methods used in this study were performed in accordance with the relevant guidelines and regulations.

Patients were enrolled from the National Institute of Cancer Research & Hospital (NICRH), Dhaka, Bangladesh, and were diagnosed with breast carcinoma by mammography, breast ultrasound, biopsy, or breast magnetic resonance imaging (MRI). Among the patients, 200 were female, and 2 were male. Control individuals were recruited from the National Institute of Ear, Nose and Throat (NIENT), Dhaka, Bangladesh. Among the controls, 100 were female and 4 were male and had no history or evidence of cancer.

Participants were informed about the nature of the study and the experimental procedures. Informed consent was taken from all the study subjects before collecting samples. Information on sociodemographic characteristics, i.e., age, height, weight, family income, residential area, education, medical history, menstrual and reproductive history, and family history of cancer, was obtained through a structured questionnaire. The detailed history of breast cancer of each patient, including pathological tumor grade, tumor size, age at diagnosis, total white blood cell (WBC) count, erythrocyte sedimentation rate (ESR), and biomarker status for progesterone receptor (PR), estrogen receptor (ER), and human epidermal growth factor receptor type 2 (HER2) were recorded from their medical files in presence of expert physicians.

Sample collection

Five (5.0) ml of venous blood was collected from all individuals using a disposable syringe by expert phlebotomists following all aseptic precautions. Then, the blood was transferred to EDTA-containing vacutainer tubes. The plasma was separated by centrifugation at 3,000 rpm for 15 min. The cellular fractions and the plasma were then stored at -20 °C until further use.

Genotyping for rs1801131 and rs1801133

DNA was extracted from the cellular fractions using an organic method described previously[17–19]. Polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) was performed to determine the genotypes of rs1801133 and rs1801131. Later, 5% of the PCR products (of both SNPs from both the case and control groups) were subjected to Sanger sequencing to validate the results obtained from RFLP. The detailed procedure is described below:

rs1801133 (C677T) PCR was carried out in a 15 µL reaction volume in a PCR tube containing 7.5 µL of GoTaq G2 Green Master Mix (Promega Corporation), 5.05 µL of nuclease-free water, 0.45 µL of DMSO, 0.5 µL of forward and reverse primers and 1 µL of extracted DNA. The primer sequences are given in Supplementary Tables 1, and the PCR conditions are shown in Supplementary Table 2. The 198 bp PCR product was digested using the HinfI restriction enzyme. Digestion was carried out in a 15 µL reaction volume for 3 h at 37 °C, and two fragments of 175 bp and 23 bp were obtained in the presence of the mutant T allele. The wild-type homozygous C/C genotype produced no cleavage, and the resulting PCR products consisted of 198 bp fragments. The heterozygous C/T genotype produced 3 bands of 198 bp, 175 bp, and 23 bp. The mutant homozygous T/T genotype produced only 2 bands of 175 bp and 23 bp (Fig. 1A, Supplementary Fig. 4A). The bands were separated and visualized on 15% polyacrylamide gels under ultraviolet light after staining with ethidium bromide.

rs1801131 (A1298C) A similar approach was taken to amplify the DNA region containing rs1801131 using the same reaction mixture as described earlier. The specific primer sequences for this region are listed in Supplementary Tables 1, and the PCR conditions are shown in Supplementary Table 22. The PCR product was 163 bp in length, and the MboII restriction enzyme was used for digestion. The wild-type homozygous A/A genotype yielded 5 bands of 56 bp, 31 bp, 30 bp, 28 bp, and 18 bp. The Heterozygous A/C genotype produced 6 bands of 84 bp, 56 bp, 31 bp, 30 bp, 28 bp, and 18 bp. The mutant homozygous C/C genotype yielded 4 bands of 84 bp, 31 bp, 30 bp, and 18 bp (Fig. 1B, Supplementary Fig. 4B). The bands were separated and visualized via the same method described in the previous section.

Sequencing of PCR products

5% of the PCR products, including both the case and control products, were randomly chosen and sequenced by the Sanger sequencing (Barcode-tagged Sequencing, BTSeq) method[20] to confirm the genotyping results obtained from the PCR-RFLP method. The results (chromatograms) were analyzed using Geneious Prime 2022.2 software as described previously[20].

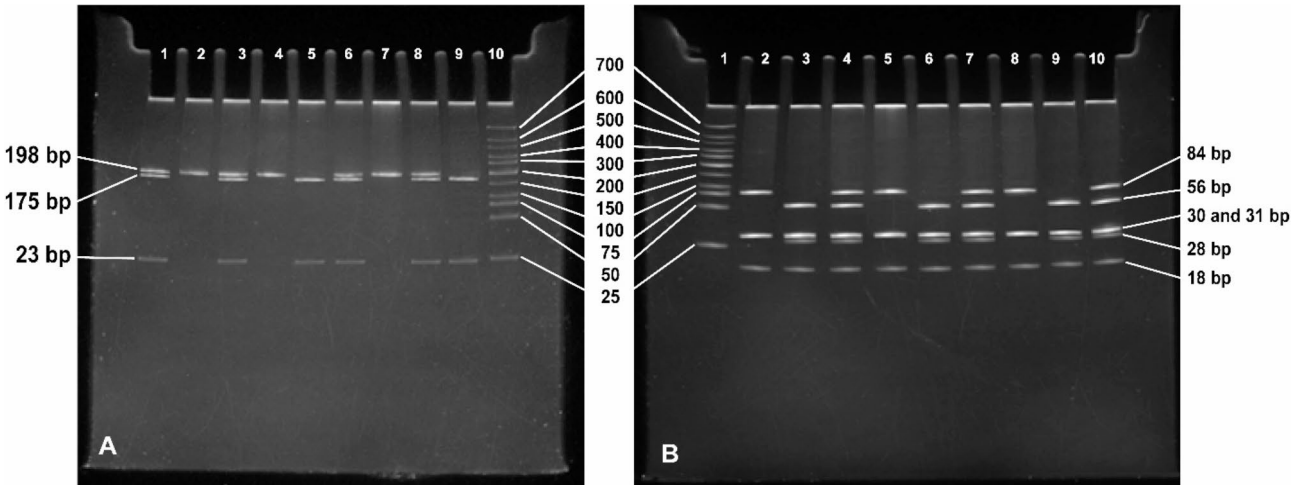

Fig. 1. Gel images of restriction digestion products of different genotypes of rs1801133 (C677T) and rs1801131 (A1298C). (**A**) Genotypes of rs1801133 (C677T). The 198 bp fragment in lanes 2, 4, and 7 indicates the wild-type C/C genotype. The presence of 198 bp, 175 bp, and 23 bp in lanes 1, 3, 6, and 8 indicates a heterozygous C/T genotype, while the presence of 175 bp and 23 bp in lanes 5 and 9 indicates the T/T genotype. Lane 10 contains a 25 bp DNA ladder. (**B**) The rs1801131 (A1298C) genotype. The 56 bp, 31 bp, 30 bp, 28 bp, and 18 bp fragments in lanes 3, 6, and 9 indicate the homozygous wild-type A/A genotype. The presence of 84 bp, 56 bp, 31 bp, 30 bp, 28 bp, and 18 bp in lanes 4, 7, and 10 indicates a heterozygous mutant A/C genotype, while the presence of 84 bp, 31 bp, 30 bp, and 18 bp in lanes 2, 5, and 8 indicates a homozygous mutant C/C genotype. Lane 1 contains a 25 bp DNA ladder.

Statistical analyses

Statistical analyses were performed using GraphPad Prism (version 8.4.2) and the R programming language (version 4.1.2). Quantitative data (age and BMI) are expressed as the mean ± standard deviation (SD). Categorical data are summarized as percentages (%). For analysis of parameters related to female participants only (menstrual status, age at menarche, number of pregnancies), male participants (both case and controls) were not included. Therefore, for those parameters, total 200 cases and 100 controls were taken into consideration. The difference in means between continuous variables was assessed by unpaired t test. The associations between different categorical variables were calculated using the chi-square test and Fisher's exact test. The odds ratio (OR) at the 95% confidence interval (CI) was calculated to estimate risk. A p value less than *0.05* was considered to indicate statistical significance in all analyses. Hardy–Weinberg equilibrium, linkage disequilibrium (LD), and haplotype analysis were performed using the "SHEsisPlus" (http://shesisplus.bio-x.cn/SHEsis.html) web tool.

In silico analysis of rs1801131 and rs1801133 polymorphisms in the MTHFR protein

Multiple in silico analysis tools were employed according to the methods described in our previously published studies[21,22]. For example, Sorting Intolerant From Tolerant (SIFT) (https://sift.bii.a-star.edu.sg/)[23] and Polymorphism Phenotyping v2 (PolyPhen-2) (http://genetics.bwh.harvard.edu/pph2/)[24] were used to predict the effects of the target SNPs on the resulting proteins. Predictors of human deleterious single nucleotide polymorphisms (PhD-SNPs) (https://snps.biofold.org/phd-snp/phd-snp.html)[25], SNPs&GO (https://snps.biofold.org/snps-and-go/snps-and-go.html)[26] and meta-SNPs (https://snps.biofold.org/meta-snp/)[27] were utilized to study disease associations with the SNPs. MutationAssessor (http://mutationassessor.org/r3/)[28] was used to predict the functional impact of the variants. MUpro (https://mupro.proteomics.ics.uci.edu/)[29] and the impact of nonsynonymous mutations on protein stability – multi dimension (INPS-MD) (https://inpsmd.biocomp.unibo.it/inpsSuite/)[30] were used to predict the impact of the SNPs on protein stability. The ConSurf server (https://consurf.tau.ac.il/consurf_index.php)[31] was used to identify the conserved functional and structural residues on the protein surface and core.

The three-dimensional structure of the MTHFR protein (PDB ID: 6FCX) was obtained from the Protein Data Bank (https://www.rcsb.org/structure/6FCX). Chain A was used to create homology models using SWISS-MODEL (https://swissmodel.expasy.org/). The SWISS-MODEL structure assessment tool (https://swissmodel.expasy.org/assess), ProSA-web (https://prosa.services.came.sbg.ac.at/prosa.php)[32], and ERRAT (https://saves.mbi.ucla.edu/)[33] were used to evaluate the quality of the predicted three-dimensional (3D) models. The structures were analyzed for their template modeling score (TM-score) and root mean square deviation (RMSD) values in TM-align (https://seq2fun.dcmb.med.umich.edu//TM-align/)[34] and PyMol 2.5.4, respectively. The structures were visualized, superimposed, and analyzed using PyMol and BIOVIA Discovery Studio Visualizer 2021 Client.

Results

Demographic characteristics of the study participants

The demographic characteristics of the study participants, including age, BMI, family history of cancer, occupation, residence area, educational status, monthly income, menstrual status, age at menarche, and number of pregnancies are presented in Table 1.

In our study population, more than 75% of participants (both case and control) are within the 30 to 50 years age group. In both groups, the majority (around 90%) of the participants are housewives. Only a handful (10.89% patients and 13.46% controls) of the participants in each group received higher secondary or above levels of education. Participants from both groups (91.09% patients and 89.42% control individuals) predominantly belong to the lower income group (< 20,000 BDT) and reside in the rural area of Bangladesh (74.25% patients and 68.27% control individuals). Breast cancer patients have more than twice (12.87%) the incidence of cancer in family history, compared to control individuals (5.77%). But no significant differences were found between breast cancer patients and healthy controls regarding age, BMI, family history of cancer, occupation, residence area, educational status, monthly income, number of pregnancies, or age at menarche ($p > 0.05$). However, a statistically significant difference in menopausal status was observed between the patients and control subjects.

Clinicopathological data of breast Cancer patients

The clinical and pathological data of the breast cancer patients enrolled in this study are shown in Table 2.

Almost all (99.5%) patients included in this study had invasive ductal carcinoma (IDC), and only 1 patient had invasive lobular carcinoma (ILC). In terms of hormone receptor status, 74.14% of patients were ER+, 65.49% were PR+, and 69.9% were HER2+. A total of 63.86% of patients had a tumor within 2 to 5 cm, and 74.75% of patients had grade 2 (G2) tumors.

MTHFR polymorphisms and the risk of developing breast cancer

Gel images of restriction digestion products of different genotypes of rs1801133 (C677T) and rs1801131 (A1298C) are shown in Fig. 1.

The frequencies and associations of different genotypes according to different genetic models with the risk of breast cancer, as measured by odds ratios (ORs) at the 95% confidence intervals (CIs), are shown in Table 3.

In the control group, the genotype frequencies of rs1801131 (A1298C) were 32.69%, 61.54%, and 5.77% for the homozygous wild type (AA), heterozygous variant (AC), and homozygous variant (CC) strains, respectively. In the patient group, the frequencies were 10.39% homozygous wild type (AA), 75.25% heterozygous variant (AC), and 14.36% homozygous variant (CC). The allele frequency of alternate allele C was 36.60% in the control population and 51.98% in breast cancer patients. On the other hand, for rs1801133 (C677T), the control group had 76.92% homozygous wild type (CC), 20.19% heterozygous variant (CT), and 2.89% homozygous variant (TT) genotypes. In addition, in breast cancer patients, the frequencies were 68.32%, 28.22%, and 3.46% for

Variables	Case ($n=202$) Mean ± SD	Control ($n=104$) Mean ± SD	*p* value
Age (year)	43.25 ± 10.10	42.07 ± 8.43	*0.304*
BMI (kg/m^2)	24.25 ± 4.02	23.55 ± 3.69	*0.139*

Variables		*n* (%)	*n* (%)	*p* value
Age group	< 30	11 (5.45)	7 (6.73)	*0.602*
	30–50	151 (74.75)	81 (77.89)	
	> 50	40 (19.80)	16 (15.38)	
BMI group	≤ 25	129 (63.86)	69 (66.35)	*0.667*
	> 25	73 (36.14)	35 (33.65)	
Family history of cancer	Yes	26 (12.87)	6 (5.77)	*0.084*
	No	176 (87.13)	98 (94.23)	
Occupation	Housewife	189 (93.56)	93 (89.42)	*0.236**
	Working Woman	11 (5.45)	7 (6.73)	
	Working Man	2 (0.99)	4 (3.85)	
Residence	Urbal	52 (25.75)	33 (31.73)	*0.331*
	Rural	150 (74.25)	71 (68.27)	
Educational status	No Formal Education	53 (26.24)	25 (24.04)	*0.858*
	Primary	61 (30.20)	27 (25.96)	
	Secondary	66 (32.67)	38 (36.54)	
	Higher Secondary	12 (5.94)	7 (6.73)	
	Graduation or above	10 (4.95)	7 (6.73)	
Monthly family income	< 20,000 BDT	184 (91.09)	93 (89.42)	*0.416**
	20,000–30,000 BDT	15 (7.43)	7 (6.73)	
	> 30,000 BDT	3 (1.49)	4 (3.85)	
Menstrual status**	Premenopausal	104 (52)	67 (67)	***0.013***
	Postmenopausal	96 (48)	33 (33)	
Age at menarche**	10–13	144 (72)	76 (76)	*0.460*
	14–17	56 (28)	24 (24)	
Number of pregnancies**	< 3	110 (55)	51 (51)	*0.512*
	≥ 3	90 (45)	49 (49)	

Table 1. Demographic characteristics of the study subjects. BDT = Bangladeshi Taka. *Fisher's exact test; **Two (2) case and four (4) control individuals were male, data for these sections are not available for them.

Variables ($n=202$)		Number (*n*)	Percentage (%)
Age at diagnosis	≤40	107	52.97
	> 40	95	47.03
Histologic type	IDC	201	99.51
	ILC	1	0.49
Receptor status	ER +	150	74.25
	ER-	52	25.75
	PR+	132	65.35
	PR-	70	34.65
	HER2+	141	69.80
	HER2-	61	30.20
Tumor size	T1 (≤ 2 cm)	53	26.24
	T2 (> 2 and ≤ 5 cm)	129	63.86
	T3 (> 5 cm)	20	9.9
Grading of tumor	G1	12	5.94
	G2	151	74.75
	G3	39	19.31

Table 2. Clinicopathological data of the patients included in this study.

SNP	Genetic model	Genotypic status	Control ($n=104$) n (%)	Case ($n=202$) n (%)	p value	OR (95% CI)
MTHFR (rs1801131)	Codominant model	AA	34 (32.69)	21 (10.39)	***< 0.001***	1
		AC	64 (61.54)	152 (75.25)		3.85 (2.06–7.25)
		CC	6 (5.77)	29 (14.36)		7.82 (2.69–22.05)
	Dominant model	AA	34 (32.69)	21 (10.40)	***< 0.001***	1
		AC + CC	70 (67.31)	181 (89.60)		4.19 (2.28–7.78)
	Recessive model	AA + AC	98 (94.23)	173 (85.64)	***0.025***	1
		CC	6 (5.77)	29 (14.36)		2.74 (1.15–6.45)
	Over Dominant model	AA + CC	40 (38.46)	50 (24.75)	***0.013***	1
		AC	64 (61.54)	152 (75.25)		1.90 (1.14–3.15)
	Allele frequency	A	132 (63.40)	194 (48.02)	***< 0.001***	1
		C	76 (36.60)	210 (51.98)		1.89 (1.33–2.66)
MTHFR (rs1801133)	Codominant model	CC	80 (76.92)	138 (68.32)	*0.283*	1
		CT	21 (20.19)	57 (28.22)		1.57 (0.90–2.74)
		TT	3 (2.89)	7 (3.46)		1.35 (0.36–4.92)
	Dominant model	CC	80 (76.92)	138 (68.32)	*0.115*	1
		CT + TT	24 (23.08)	64 (31.68)		1.55 (0.90–2.65)
	Recessive model	CC + CT	101 (97.12)	195 (96.53)	*> 0.999**	1
		TT	3 (2.88)	7 (3.47)		1.21 (0.33–4.37)
	Over dominant model	CC + TT	83 (79.81)	145 (71.78)	*0.127*	1
		CT	21 (20.19)	57 (28.22)		1.55 (0.89–2.69)
	Allele frequency	C	181 (87.02)	333 (82.43)	*0.142*	1
		T	27 (12.98)	71 (17.57)		1.43 (0.89–2.34)

Table 3. Distribution of MTHFR rs1801131 (A1298C) and rs1801133 (C677T) genotypes in study participants and assessment of the risk of breast cancer. *Fisher's exact test.

homozygous wild type (CC), heterozygous variant (CT), and homozygous variant (TT), respectively. The minor allele frequency is 12.98% in the control population and 17.57% in the patient population.

According to all the genetic models analyzed (codominant, dominant, recessive, and overdominant), rs1801131 was significantly associated with the risk of breast cancer ($p < 0.05$). According to the OR analysis, AC and CC genotype carriers had 3.85- and 7.82-fold greater risks of developing breast cancer, respectively, in the codominant model. According to the dominant model, AC + CC genotype carriers had a 4.19-fold greater risk of developing breast cancer, and according to the recessive model, carriers of the CC genotype had a 2.74-fold greater risk of developing breast cancer. Overall, compared with carriers of the reference A allele, carriers of the mutant C allele had an increased (OR = 1.89, $p < 0.05$, 95% CI = 1.33–2.66) risk of developing breast cancer.

For rs1801133, no significant association was detected in any of the genetic models. Although the OR analysis revealed a greater risk of developing cancer for both heterozygous (CT) and homozygous mutant (TT) T allele carriers, the difference was not statistically significant ($p > 0.05$).

The genotype distributions of both SNPs in the study participants are shown in Fig. 2.

Distribution of MTHFR rs1801131 (A1298C) and rs1801133 (C677T) genotypes in the study subjects according to menopausal status

The female subjects were stratified according to their menstrual status (premenopause and postmenopause), and the risk of developing breast cancer in these groups was analyzed. In premenopausal women, rs1801131 was significantly associated with breast cancer risk. AC and CC genotype carriers had 5.06 and 16.25 times greater risks of developing breast cancer, respectively, than did the reference AA genotype carriers. In postmenopausal women, no significant association was found between genotype rs1801131 and breast cancer risk. For rs1801133, no significant association was found with the risk of developing the disease in either group. The results are presented in Table 4.

Association of the target SNPs with tumor grade and size in breast cancer patients

In the patient group, associations of the polymorphisms with tumor size and grade were analyzed. The results are presented in Tables 5 and 6.

The presence of the genotype CT in rs1801133 was significantly associated with an increased risk of larger tumor size (OR: 4.46, 95% CI: 2.22–9.45), but no significant association was found with tumor grade. Conversely, rs1801131 was not associated with either tumor size or grade.

Confirmation of RFLP genotyping results by sequencing

PCR products of different genotypes were sequenced by the Sanger sequencing method (BTSeq), and all samples matched the results obtained by PCR-RFLP. A representative chromatogram is shown in Fig. 3.

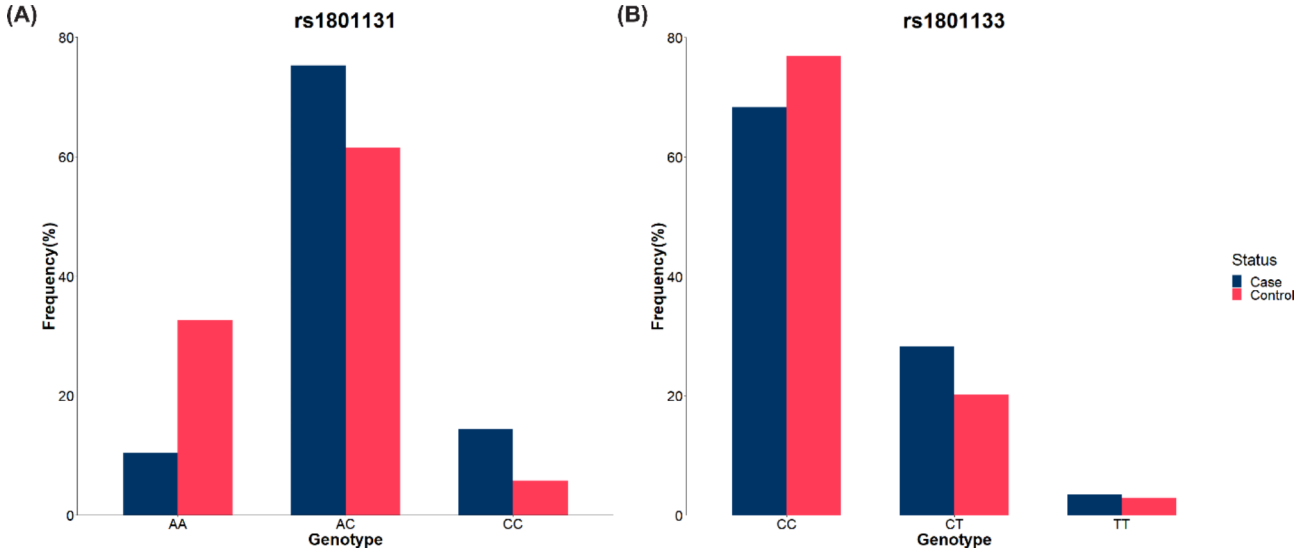

Fig. 2. Genotypic distribution of the two SNPs in the study participants. (**A**) Distribution of rs1801131 and (**B**) distribution of rs1801133.

SNP	Menopausal status	Genotype	Control *n* (%)	Case *n* (%)	*p* value	OR (95% CI)
MTHFR (rs1801131)	Premenopausal	AA	25 (37.31)	10 (9.62)	***< 0.001***	1
		AC	40 (59.70)	81 (77.89)		5.06 (2.24–11.39)
		CC	2 (2.99)	13 (12.49)		16.25 (3.33–77.57)
	Postmenopausal	AA	9 (27.27)	11 (11.46)	*0.095*	1
		AC	20 (60.61)	70 (72.92)		2.86 (1.06–7.27)
		CC	4 (12.12)	15 (15.62)		2.72 (0.61–9.55)
MTHFR (rs1801133)	Premenopausal	CC	49 (73.13)	73 (70.19)	*0.895**	1
		CT	16 (23.88)	28 (26.92)		1.18 (0.58–2.38)
		TT	2 (2.99)	3 (2.89)		1.01 (0.19–5.84)
	Postmenopausal	CC	27 (81.82)	63 (65.63)	*0.645**	1
		CT	5 (15.15)	29 (30.21)		2.48 (0.85–6.39)
		TT	1 (3.03)	4 (4.16)		1.71 (0.26–21.67)

Table 4. Distribution of rs1801131 (A1298C) and rs1801133 (C677T) genotypes in the participants stratified against menopausal status. *Fisher's exact test.

SNP	Genotype	Tumor grade		*p* value*	OR (95% CI)
		G1 n (%)	G2 + G3 n (%)		
rs1801133	CC	7 (58.34)	131 (68.94)	*0.395*	1
	CT	4 (33.33)	53 (27.90)		0.70 (0.20–2.87)
	TT	1 (8.33)	6 (3.16)		0.30 (0.04–8.44)
rs1801131	AA	1 (8.33)	20 (10.53)	*0.460*	1
	AC	8 (66.67)	144 (75.79)		1.01 (0.04–6.07)
	CC	3 (25)	26 (16.68)		0.48 (0.02–4.44)

Table 5. Distribution of rs1801131 (A1298C) and rs1801133 (C677T) genotypes in patients with different tumor grades. *Fisher's exact test.

SNP	Genotype	Tumor size		p value	OR (95% CI)
		T1 (≤2 cm) n (%)	T2 + T3 (> 2 cm) n (%)		
rs1801133	CC	37 (69.81)	56 (37.58)	***< 0.001****	1
	CT	13 (24.53)	89 (59.74)		4.46 (2.22–9.45)
	TT	3 (5.66)	4 (2.68)		0.87 (0.17–4.97)
rs1801131	AA	3 (5.66)	18 (12.08)	*0.388*	1
	AC	41 (77.36)	111 (74.50)		0.47 (0.10–1.50)
	CC	9 (16.98)	20 (13.42)		0.39 (0.07–1.57)

Table 6. Distribution of rs1801131 (A1298C) and rs1801133 (C677T) genotypes in patients with different tumor sizes. *Fisher's exact test.

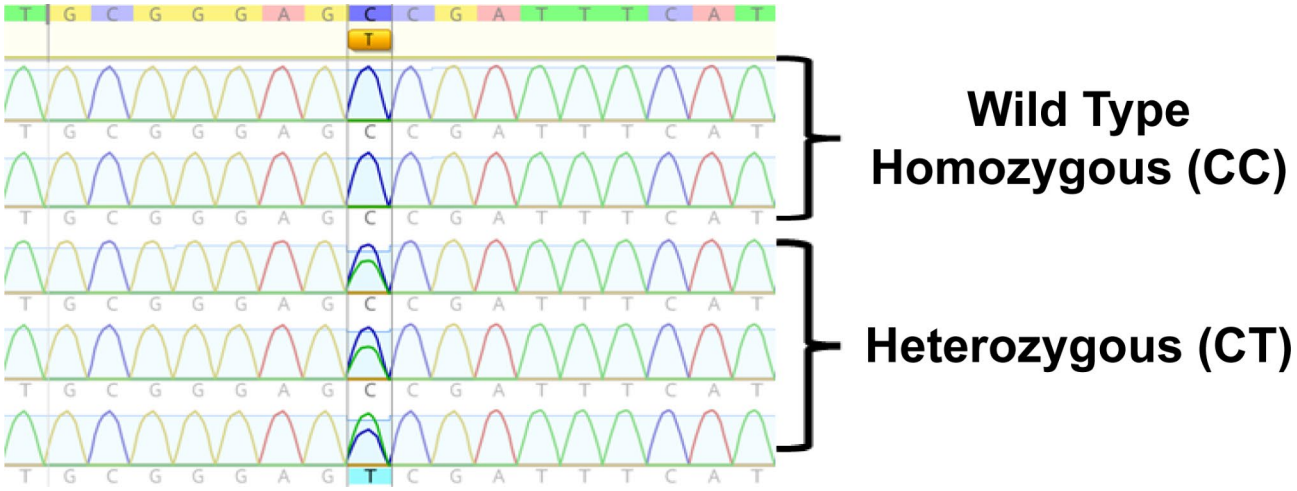

Fig. 3. Chromatogram showing the sequencing results of different genotypes of rs1801133.

SNP	Study subjects	Chi^2	Pearson's p value	Fisher's p value
rs1801131	Case	51.988	***< 0.001***	***< 0.001***
	Control	11.117	***0.003***	*0.064*
	Both	53.419	***< 0.001***	NA
rs1801133	Case	0.136	*0.933*	*0.999*
	Control	1.172	*0.556*	*0.670*
	Both	0.837	*0.657*	NA

Table 7. Hardy–Weinberg equilibrium test of 2 SNPs (rs1801131 and rs1801133) in the study subjects.

Constancy of genotype frequencies, linkage disequilibrium (LD) and haplotype analysis

The Hardy–Weinberg equilibrium was used to analyze the constancy of the genotype frequencies observed in this study. For rs1801131, deviation from the Hardy–Weinberg equilibrium was observed in both the control and patient groups, and both combined. However, for rs1801133, the study population was in equilibrium. The results are presented in Table 7.

LD analysis and R^2 results for the two SNPs are shown in Fig. 4. Although D' is 0.71, R^2 is very low (0.08), which represents weak LD between the SNPs.

Four (4) possible haplotype combinations are possible for two (2) SNPs studied. The results of the haplotype analysis are presented in Table 8.

The AC haplotype (reference alleles at both loci) is protective against breast cancer development (OR = 0.468, 95% CI = 0.332–0.659, $p < 0.05$), and the CC haplotype is associated with an increased risk of breast cancer development (OR = 1.757, 95% CI = 1.245–2.48, $p < 0.05$).

Total WBC count and ESR

Breast cancer patients had significantly greater ($p < 0.05$) total WBC counts (9,271 ± 2,608 cells/mm^3) than did control subjects (8,139 ± 1,140 cells/mm^3). The results are shown in Fig. 5.

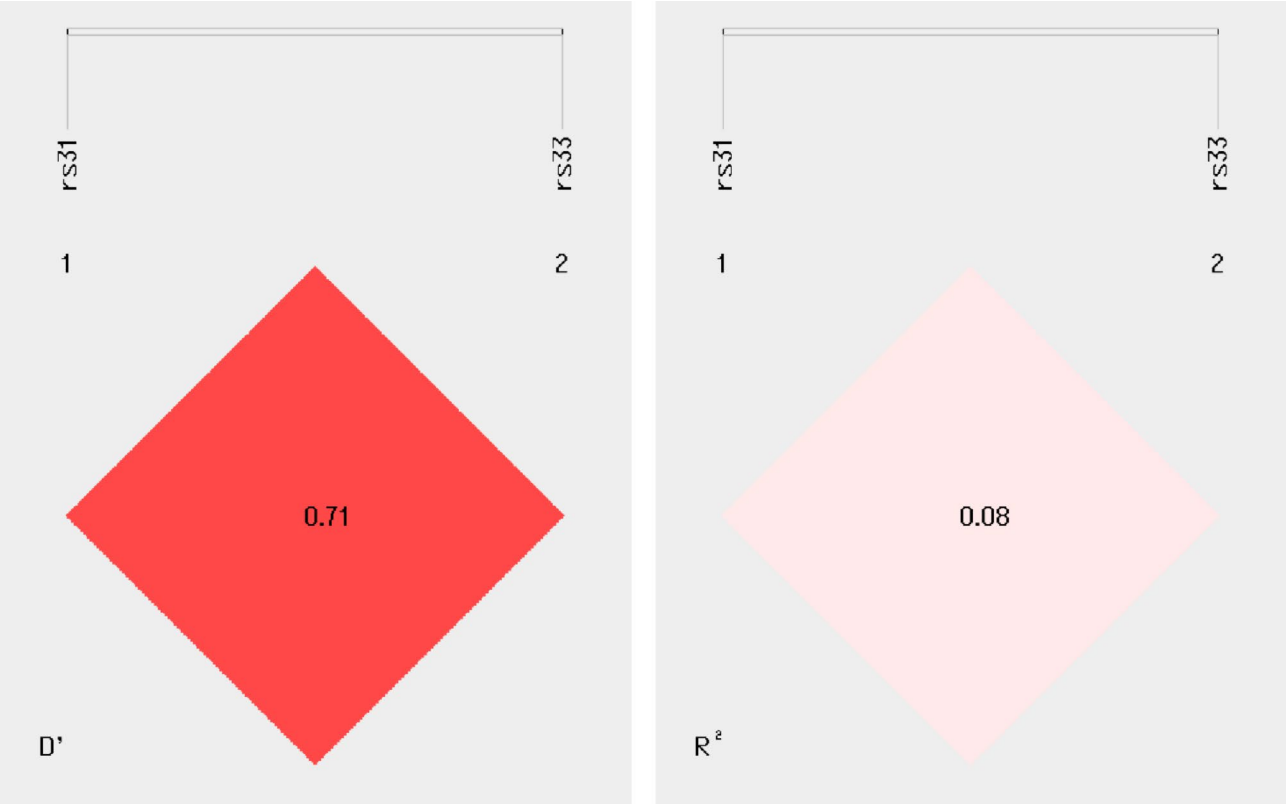

Fig. 4. Linkage disequilibrium (LD) plot for 2 SNPs of the MTHFR gene.

Haplotype	Case (frequency)	Control (frequency)	Chi²	Fisher's *p* value	Pearson's *p* value	OR (95% CI)
AC	134 (0.331)	107 (0.514)	19.207	***< 0.001***	***< 0.001***	0.468 (0.332–0.659)
AT	60 (0.148)	25 (0.12)	0.92	*0.388*	*0.337*	1.276 (0.774–2.104)
CT	11 (0.027)	2 (0.009)	2.048	*0.236*	*0.152*	2.882 (0.633–13.129)
CC	199 (0.492)	74 (0.355)	10.4	***0.001***	***0.001***	1.757 (1.245–2.48)

Table 8. Haplotype analysis of 2 SNPs (rs1801131 and rs1801133) in the participants. (The first locus represents rs1801131).

The ESR was also significantly higher in breast cancer patients (36.69 ± 178.06 mm/1st hour) than in healthy controls (23.8 ± 8.26 mm/1st hour). Comparison between patient and control ESR is shown in Fig. 6.

In silico analysis of the effects of SNPs on the MTHFR protein

Analysis of functional consequences of rs1801131 (E429A) and rs1801133 (A222V)

Different web-based tools have been used to predict the effect of polymorphisms on the MTHFR protein. All the tools, except SIFT, predicted the A222V (rs1801133) polymorphism as damaging or disease-causing. However, the E429A (rs1801131) polymorphism was predicted to be tolerated or neutral by all the tools. Both MUpro and INPS-MD predicted that both mutations decrease the stability of the protein. The results are summarized in Table 9. The scores from the different tools are shown in Supplementary Table 3.

The results from the ConSurf server revealed the functional and/or structural conservation of amino acid residues at the SNP sites. Functional residues are defined as being highly conserved and being in an exposed position on the protein surface, whereas structural residues are defined as being highly conserved and in a buried position. Alterations of amino acid residues in highly conserved regions are more detrimental than those in less stable regions. In MTHFR, A222 was found to be a conserved and buried residue, whereas E429 was an exposed but variable residue. Therefore, the A422V polymorphism might be damaging to the protein.

Homology modeling

The 3D structure of the human MTHFR protein (6FCX) obtained from the RCSB PDB had 2 mutations in its sequence. Therefore, the reference sequence from UniProtKB was obtained in FASTA format and used to construct the reference 3D model. Two (2) other models were generated by substituting A222 and E429

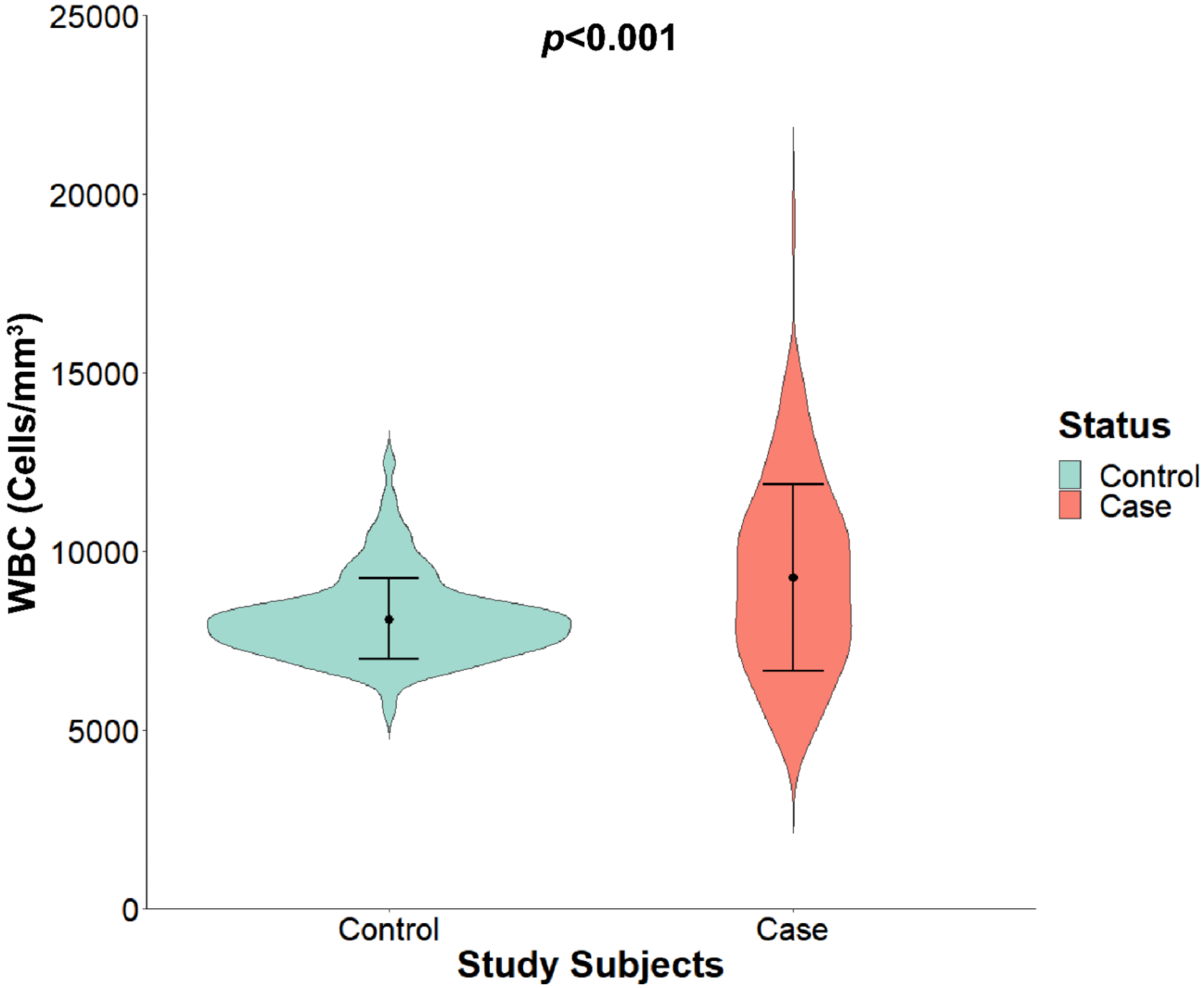

Fig. 5. Violin plot demonstrating Total WBC count in breast cancer patients and healthy controls.

with valine and alanine, respectively. Evaluation of the predicted structure by the SWISS-MODEL structure assessment tool, ProSA-Web, and ERRAT yielded satisfactory results, and the models were of good quality. The evaluation scores from the different tools are presented in Supplementary Table 4. The changes in amino acid residues in 3D structures were also visualized and are presented in Supplementary Figs. 1 and 2.

Analysis of the impact of mutated residues on protein structures

Mutated structures were superimposed onto the predicted wild-type structure using PyMol, and RMSD values were calculated. A higher RMSD indicates a greater deviation between the structures. Both structures had the same RMSD (0.062 Å) as the wild-type structure. Additionally, both structures had the same TM score of 0.99995. TM scores are calculated within a range of 0 to 1, where a score of 0–0.3 indicates random structural similarity and 0.5-0.99999 indicates a good fit between the models. A score of 1 represents a perfect match between the C alpha atoms of the protein backbone. TM-align also provided an RMSD between the protein structures, which was similar to that of PyMol (0.06 Å vs. 0.062 Å). Combining the RMSD value and TM score, it can be deduced that the protein structure, on a broad scale, was not substantially altered by these substitutions. However, interactions with neighboring residues or the cellular environment might be affected by these polymorphisms.

The superimposed structures were further analyzed for the impact of the mutated residues on neighboring interactions and bonding patterns. The mutated residues on superimposed structures and bonding patterns of wild-type and mutated residues are shown in Fig. 7.

As shown in Fig. 7(A), the introduction of valine, which substitutes for the alanine residue at position 222, caused an increase in residue size, but because both residues are hydrophobic, neither the charge nor the hydrophobicity changed. Fig. 7(C) shows that the H-bond distance between Ala222 and Val218 was 2.978 Å, whereas the hydrophobic interactions between Ala222 and Ile192 and between Ala222 and Val194 were 5.247 Å and 3.937 Å, respectively. On the other hand, for the mutated valine residue at 222, the H-bond distance with Val218 decreased to 2.961 Å, but another new hydrophobic interaction was introduced with Val179, which had a

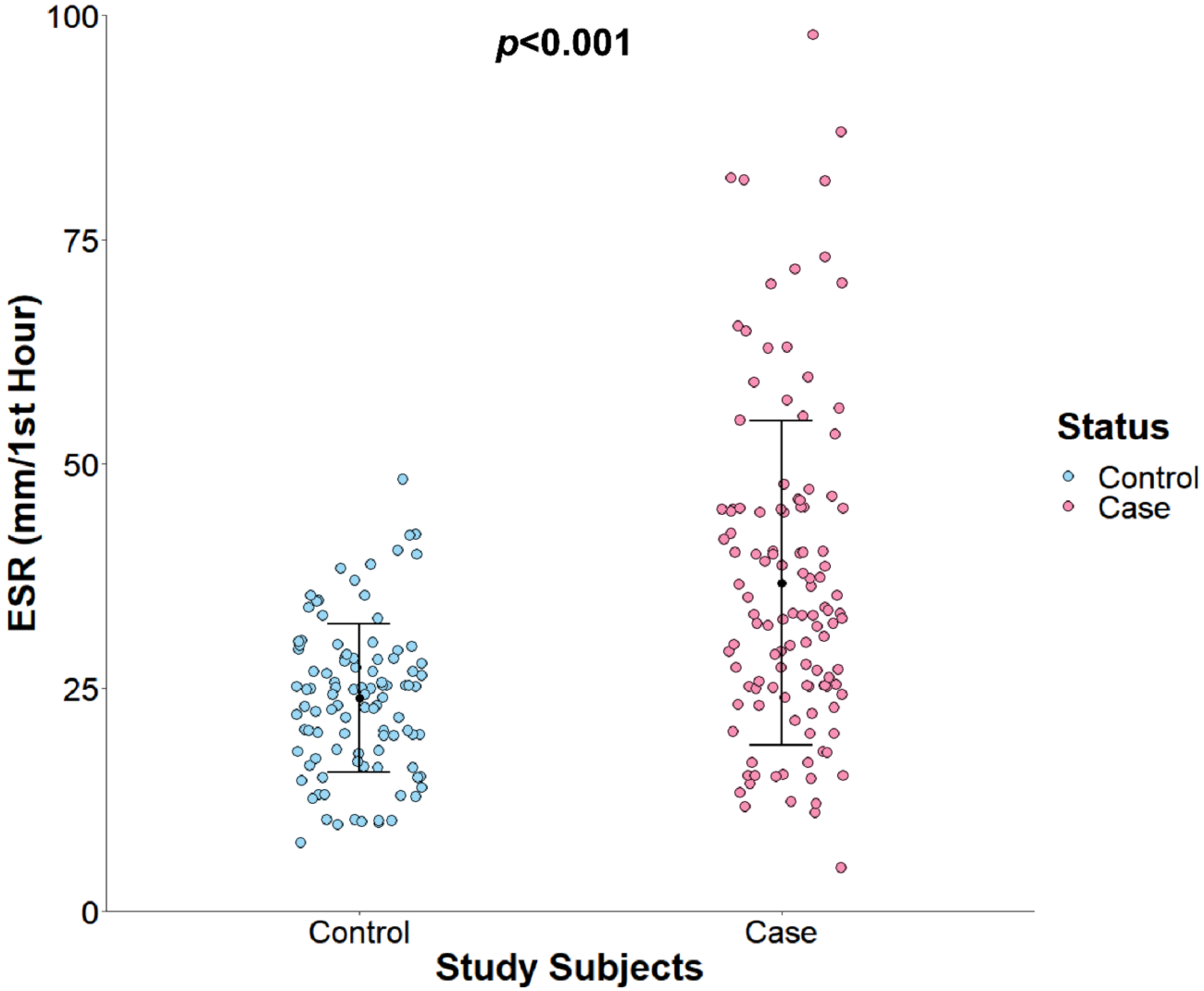

Fig. 6. Jitter plot showing ESR in breast cancer patients and healthy controls.

Tool	rs1801131 (E429A)	rs1801133 (A222V)
SIFT	Tolerated	Tolerated
Polyphen2	Benign	Most likely, damaging
PhD-SNP	Neutral	Disease
SNPs&GO	Neutral	Disease
Meta-SNP	Neutral	Disease
MutationAssessor	Low functional Impact	Medium functional Impact
MuPro	Decrease stability	Decrease stability
INPS-MD	Decrease stability	Decrease stability

Table 9. Functional effects of the two SNPs.

distance of 4.722 Å (Fig. 7(E)). The previous interactions with Ile192 and Val194 were retained but were reduced to 5.192 Å and 3.46 Å, respectively.

For the E429A polymorphism, glutamic acid was replaced by alanine (Fig. 7(B)). This caused differences in size, charge, and hydrophobicity, as glutamic acid is a polar, acidic residue, but alanine is a nonpolar, hydrophobic amino acid. However, no bonds were diminished with neighboring residues. In both cases, 3 H-bonds were observed with Ser427, Ser432, and Glu433. For Glu429, the distances were 3.276 Å, 3.089 Å, and 3.135 Å (Fig. 7(D)). When alanine was introduced at position 429, the distances were 3.270 Å, 3.154 Å, and 3.236 Å, respectively (Fig. 7(F)).

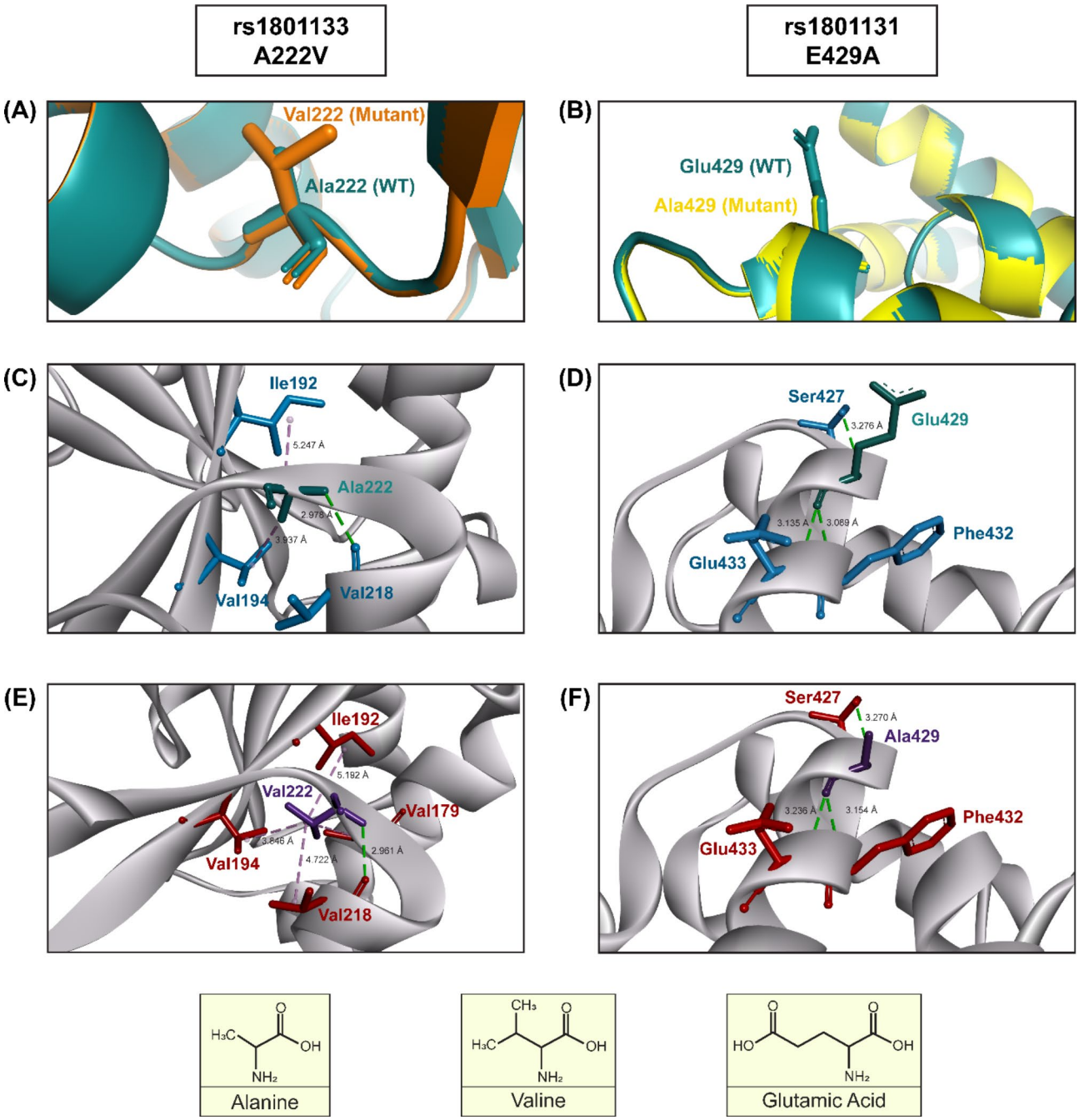

Fig. 7. Impact of the mutated residues on neighboring interactions. (**A**), (**C**), and (**E**) correspond to the A222V polymorphism (rs1801133), and (**B**), (**D**), and (**F**) correspond to the E429A polymorphism (rs1801131). (**A**) The superimposed structure of wild-type alanine and mutated valine at position 222 (the wild-type structure is shown in teal ribbons, and the mutated structures are shown in orange). (**B**) The structures of wild-type glutamate and mutated alanine at position 429 (wild-type in teal and mutated in yellow.) (**C**) and (**E**) show the differences in neighboring interactions due to the A222V polymorphism. (**D**) and (**F**) demonstrate the effects of the E429A polymorphism on neighboring interactions. Hydrogen bonds (H-bonds) are denoted in green, and hydrophobic interactions are shown in purple.

The polymorphisms affected the bond angles and bond distances between the C alpha of adjacent amino acids. For the A222V polymorphism, in the wild-type structure, the bond angle between Gly221-Ala222-Asp223 was 121.74°, and in the mutated structure, the angle between Gly221-Val222-Asp223 was 125.32°. The C alpha distance for Gly221-Ala222 was 3.818 Å for the wild-type structure and 3.824 Å for Gly221-Val222 for the mutated structure. Between the C alphas of Ala222-Asp223, a distance of 3.813 Å was observed in the wild-type structure. In the mutated structure, the distance was 3.818 Å. In the structure predicted for E429A, the bond angle between Glu428-Ala429-Ser430 was 87.85°, but it was 88.49° in the wild-type structure, between Glu428-

Glu429-Ser430. The bond distance between Glu428 and Glu429 was 3.854 Å, and that between Glu429 and Ser430 was 3.853 Å in the wild-type structure. For the structure with alanine at position 429, the bond distances were 3.836 Å and 3.845 Å for Glu428-Ala429 and Ala429-Ser430, respectively.

During the superimposition of the mutated structures with the wild-type structure, as shown in Fig. 7(A) and (B), little to no fluctuations were observed for either amino acid. For A222V, the C alpha was displaced by 0.1 Å compared to that of the wild-type structure. The C beta moved by 0.3 Å, the N by 0.4 Å, and the O atom moved 0.2 Å away from their positions in the wild-type structure. For the E429A polymorphism, the values were 0.1 Å for all four atoms mentioned.

The substrate binding site of MTHFR differs from that of the polymorphisms analyzed. The residues important for binding 5-methyl THF and NAD(P)H are Q228, Q267, K270, L271, and L323. The residues were visualized in the superimposed position, and the residues in the mutated structures did not change in position and superimposed perfectly to the reference structure. The residues in the superimposed position are shown in Supplementary Fig. 3.

Discussion

This study aimed to evaluate the association of rs1801133 and rs1801131 in the MTHFR gene with the risk of breast cancer in a Bangladeshi population. Additionally, the impacts of these polymorphisms on the MTHFR protein were assessed using different in silico tools.

These two polymorphisms are well studied for their association with different types of cancers, e.g., colorectal cancer, gastric cancer, bladder cancer, lung cancer, ovarian cancer, and liver cancer, in different populations[35–42]. However, the association with breast cancer has been most commonly reported in different populations, including China, Thailand, Iran, Morocco, Ecuador, Egypt, Turkey, and Jordan[2,43–49]. Apart from cancer, significant associations of these polymorphisms with other diseases, such as Parkinson's disease, ischemic stroke, coronary artery disease, gestational diabetes mellitus, rheumatoid arthritis, spontaneous abortion, etc., have been reported by different groups[13,50–55].

In the South Asian region, the associations of rs1801133 and rs1801131 with breast cancer have been studied in India and Pakistan[56,57]. However, no reports analyzing their association with the Bangladeshi population were found. In Bangladesh, however, several other studies have indicated the relationships of different polymorphisms and genes, such as rs13181 of ERCC2, rs2276466 of ERCC4, rs80357713, and rs80357906 of BRCA1, rs11571653 of BRCA2, rs1136201 of HER2, rs1042522 of TP53, rs16260 of CDH1, rs25487 of XRCC1, rs861539 of XRCC3, rs1219648, rs2420946, and rs2981582 of the FGFR2 gene, with breast cancer[58–62]. This study is therefore the first to evaluate the associations of rs1801133 and rs1801131 of the MTHFR gene with breast cancer in the Bangladeshi population.

Studies that sought to determine the association of these polymorphisms with breast cancer have reported different results. Some studies found significant associations, and others did not. Most studies have shown significant associations between rs1801133 and breast cancer risk, but some reports have suggested otherwise. Reports have suggested that the presence of rs1801133 increases breast cancer risk in Egyptian, Ecuadorian, Moroccan, Chinese, Iranian, and Jordanian populations[2,43–49]. However, in the Thai population, Sangrajrang et al. reported no association[45]. On the other hand, research from Kaya et al. reported no significant association in the Turkish population, but their meta-analysis showed a significant association in a larger Turkish cohort[48]. We did not find any associations in our population either. Although the ORs in all the genetic models indicated a greater risk of developing breast cancer, the differences were not statistically significant. We did not find a substantial difference in altered allele frequency between the case (18%) and control (13%) groups. Therefore, our results align with those observed in the Turkish and Thai populations, which report no association for this SNP.

On the other hand, the association of rs1801131 with breast cancer has been less well documented. However, the results are ambiguous due to differences in the findings of different groups. Research conducted on Ecuadorian, Thai, Chinese, and Jordanian populations did not find any association between increased breast cancer risk and rs1801131 genotype. However, Omran et al. reported a strong association of both AC and CC genotypes with increased risk in the Egyptian population[2]. Our results are in concordance with their report. We observed a significantly increased risk of breast cancer due to rs1801131 in all the genetic models. An OR greater than 1 indicated an increased risk, with $p < 0.05$ indicating a statistically significant difference.

When stratified against menopausal status, the role of these polymorphisms in increasing breast cancer risk is inconclusive in the previous literature. In postmenopausal women, some reports suggest no association of either SNPs[63] or only rs1801133[64], whereas other research shows increased risk due to rs1801133[65]. However, in premenopausal women, previous studies have suggested that these polymorphisms are associated with increased risk[64,66]. We also observed a greater risk in premenopausal women due to rs1801131. In postmenopausal women, the AC and CC genotypes increased breast cancer risk 3-fold, but none of them were a significant risk factor. However, for rs1801133, we did not find any significant association in either pre- or postmenopausal women.

We observed an association between the CT genotype of rs1801133 and larger tumor size (T2 or higher). Our results are in concordance with those of Omran et al., who reported similar findings[2]. However, they also reported an association with rs1801131, which we did not find. We did not observe an association between tumor grade and either SNP.

LD analysis revealed a weak linkage between the two SNPs. This finding is consistent with a previous report from Hardi et al., who also reported weak LD between these two SNPs[44].

Our study population followed the Hardy–Weinberg equilibrium for rs1801133 but deviated from the equilibrium for rs1801131. Notably, for rs1801131, we found a significant association with an increased risk of breast cancer. There was a greater occurrence of heterozygous AC (61.54% vs. 75.25%) and homozygous mutant

CC (5.77% vs. 14.36%) in patients than in controls, possibly because, as in the diseased population, the genotype frequencies are not in equilibrium. A similar deviation was also reported in a previous study[45].

In the in silico analyses conducted in this study, different tools were used to predict the effects of the polymorphisms on the MTHFR protein structure and function. For rs1801133, which results in the A222V substitution, the tools predicted the substitution to be deleterious. This residue is buried in the protein core, and alterations might impact the folding of the protein. However, as the substitution introduced Val222 instead of Ala222, which has similar physicochemical properties, we did not observe a large change in protein structure. However, the introduction of Val222 created new hydrophobic interactions with neighboring residues. As this position is part of the catalytic domain and is buried, this substitution might hamper protein structure and catalytic function in a dynamic situation such as the cellular environment.

However, E429A, which is associated with rs1801131, was predicted to be tolerated or to have a neutral effect on the protein structure. We also found similar results after superimposing the structures. Even after the substitution, the altered residue maintained the H-bonds that were observed for the wild-type residue. This indicates that the neighboring interactions were not hampered by this change. However, as this residue is located at the surface of the protein, the substitution of the polar Glu429 residue by the hydrophobic Ala429 residue might affect the interaction of the protein with the cellular environment, which may alter the stability of the protein. Analysis by different tools also predicted decreased stability due to this polymorphism. As Glu429 is part of the regulatory domain of the protein, substitution at this position might not affect the catalytic region directly but may interfere with the regulation of protein activity. The bond angles and C alpha bond distances, which in turn affect the dihedral angles, were also altered by both polymorphisms. These findings may also have further implications for protein function.

This study has some limitations. We analyzed a limited number of samples, which may influence the results of such case-control studies. The patients were enrolled from only one hospital and mostly belonged to the lower-income group, which may not properly represent the whole population. Furthermore, we have not assessed the effects of different factors like diet, physical activity, lifestyle, and other environmental factors on our results. We also didn't measure serum homocysteine levels, which could be an important biochemical parameter in assessing the effects of the polymorphisms. Apart from these, in the in silico investigation, performing molecular dynamics simulations could better reflect the effect of the polymorphisms on protein structure and possibly stability and enzymatic activity.

In conclusion, this study revealed the association of the MTHFR polymorphism rs1801131 with an increased risk of breast cancer in the Bangladeshi population. In addition, this study investigated the effects of rs1801131 (E429A) and rs1801133 (A222V) on the MTHFR protein via computational approaches. The effects predicted by in silico approaches may assist in further elucidation of the role of these polymorphisms in altering protein stability and enzymatic activity by more advanced techniques. By understanding the structure and function of specific proteins involved in breast cancer, we can design targeted therapies that are more effective and have fewer side effects compared to traditional treatments. This can lead to the development of personalized medicine approaches, tailored to the unique molecular profile of each patient's cancer. The findings of the in vitro experiments also need to be further validated in a larger cohort.

Data availability

All data generated or analyzed during this study are included in this manuscript and its supplementary material files.

Received: 28 February 2024; Accepted: 7 October 2024
Published online: 16 October 2024

References

1. Sarkar, S. et al. Cancer development, progression, and therapy: an epigenetic overview. *Int J Mol Sci*https://doi.org/10.3390/ijms141021087 (2013).
2. Omran, M. H. et al. Strong Correlation of MTHFR Gene Polymorphisms with Breast Cancer and its Prognostic Clinical Factors among Egyptian Females. *Asian Pac J Cancer Prev* **22**(2), 617–626. https://doi.org/10.31557/APJCP.2021.22.2.617 (2021).
3. Lei, S. et al. Global patterns of breast cancer incidence and mortality: A population-based cancer registry data analysis from 2000 to 2020. *Cancer Commun (Lond)*https://doi.org/10.1002/cac2.12207 (2021).
4. Houghton, S. C. & Hankinson, S. E. Cancer progress and priorities: breast cancer. *Cancer Epidemiol. Biomarkers Prev.* **30** (5), 822–844. https://doi.org/10.1158/1055-9965.EPI-20-1193 (2021).
5. Arthur, R. S., Wang, T., Xue, X., Kamensky, V. & Rohan, T. E. Genetic factors, adherence to healthy lifestyle behavior, and risk of invasive breast cancer among women in the UK biobank. *J. Natl. Cancer Inst.* **112** (9), 893–901. https://doi.org/10.1093/jnci/djz241 (2020).
6. Hamasaki, T. et al. N-acetyltransferase-2 gene polymorphism as a possible biomarker for prostate cancer in Japanese men. *Int. J. Urol.* **10** (3), 167–173. https://doi.org/10.1046/j.1442-2042.2003.00586.x (2003).
7. Naushad, S. M. et al. Association of aberrations in one-carbon metabolism with molecular phenotype and grade of breast cancer. *Mol Carcinog*https://doi.org/10.1002/mc.21830 (2012).
8. S. Raghubeer and T. E. Matsha, "Methylenetetrahydrofolate (MTHFR), the One-Carbon Cycle, and Cardiovascular Risks.," *Nutrients*, vol. **13**, 12, (2021). https://doi.org/10.3390/nu13124562.
9. Wan, L. et al. Methylenetetrahydrofolate reductase and psychiatric diseases. *Transl Psychiatry*https://doi.org/10.1038/s41398-018-0276-6 (2018).
10. Eldeeb, M. K. et al. Genetic polymorphisms and gene expression of one-carbon metabolizing enzymes and their relation to breast cancer. *Egypt. J. Med. Hum. Genet*https://doi.org/10.1186/s43042-022-00296-8 (2022).
11. Castiglia, P. et al. Methylenetetrahydrofolate reductase (MTHFR) C677T and A1298C polymorphisms in breast cancer: a sardinian preliminary case-control study. *16*. https://doi.org/10.7150/ijms.32162 (2019).
12. Oliveira, K. C. et al. C677T and A1298C polymorphisms of MTHFR gene and their relation to homocysteine levels in Turner syndrome. *Genet. Test. Mol. Biomarkers*https://doi.org/10.1089/gtmb.2011.0222 (2012).

13. Shivkar, R. R. et al. Association of MTHFR C677T (rs1801133) and A1298C (rs1801131) Polymorphisms with Serum Homocysteine, Folate and Vitamin B12 in Patients with Young Coronary Artery Disease. *Indian J Clin BiochemBold"*>*37*(2), 224–231. https://doi.org/10.1007/s12291-021-00982-1 (2022).
14. Frosst, P. et al. A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *United States.*https://doi.org/10.1038/ng0595-111 (1995).
15. Czechowicz, P. et al. Polymorphisms of the MTHFR gene in mothers of children with trisomy 21 Down syndrome in a Polish population. *Advances in Clinical and Experimental Medicine.* **29**(2), 251–256 (2020).
16. Samii, A. et al. MTHFR gene polymorphisms and susceptibility to myocardial infarction: Evidence from meta-analysis and trial sequential analysis. *IJC Hear. & Vasc.,* **49**, 101293. https://doi.org/10.1016/j.ijcha.2023.101293 (2023).
17. Ahmed, R. et al. Intronic variants of the angiotensin-converting enzyme 2 gene modulate plasma ACE2 levels and possibly Confer Protection against severe COVID-19. *Biomed. Res. Int.* **2023**, 5705076. https://doi.org/10.1155/2023/5705076 (2023).
18. Nila, N. N. et al. The IKAROS Transcription Factor Gene IKZF1 as a Critical Regulator in the Pathogenesis of Childhood Acute Lymphoblastic Leukemia Insights from a Bangladeshi Population. *Indian J. Clin. Biochem.,*https://doi.org/10.1007/s12291-024-01218-8 (2024).
19. Nabil, I. K. et al. Vitamin D deficiency and the vitamin D receptor (VDR) gene polymorphism rs2228570 (FokI) are associated with an increased susceptibility to hypertension among the Bangladeshi population. *PLoS One* **19**(3), e0297138 https://doi.org/10.1371/journal.pone.0297138 (2024).
20. Saba Al, A. et al. Single nucleotide variants rs7975232 and rs2228570 within vitamin D receptor gene confers protection against severity of COVID-19 infection in Bangladeshi population. *Gene Rep.* **36**, 101981 https://doi.org/10.1016/j.genrep.2024.101981 (2024).
21. Nila, N. N. et al. Investigating the structural and functional consequences of germline single nucleotide polymorphisms located in the genes of the alternative lengthening of telomere (ALT) pathway. *Heliyon*. https://doi.org/10.1016/j.heliyon.2024.e33110 (2024).
22. Ahmed, H. U., Paul, A., Mahmud, Z., Rahman, T. & Hosen, M. I. Comprehensive characterization of the single nucleotide polymorphisms located in the isocitrate dehydrogenase isoform 1 and 2 genes using in silico approach. *Gene Rep.* **24**, 101259. https://doi.org/10.1016/j.genrep.2021.101259 (2021).
23. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. *Nat. Protoc.* **11** (1), 1–9. https://doi.org/10.1038/nprot.2015.123 (2016).
24. Adzhubei, I. A. et al., A method and server for predicting damaging missense mutations. *Nature methods,* **7**(4), 248–249. https://doi.org/10.1038/nmeth0410-248 (2010).
25. Capriotti, E., Calabrese, R. & Casadio, R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics.* **22**(22), 2729–2734 https://doi.org/10.1093/bioinformatics/btl423 (2006).
26. Calabrese, R., Capriotti, E., Fariselli, P., Martelli, P. L. & Casadio, R. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum Mutat* **30**(8), 1237–1244 https://doi.org/10.1002/humu.21047 (2009).
27. Capriotti, E., Altman, R. B. & Bromberg, Y. Collective judgment predicts disease-associated single nucleotide variants. *BMC Genom* **14**(3), S2. https://doi.org/10.1186/1471-2164-14-S3-S2 (2013).
28. Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* **39**(17), e118–e118. https://doi.org/10.1093/nar/gkr407 (2011).
29. Cheng, J., Randall, A. & Baldi, P. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins* **62**(4), 1125–1132. https://doi.org/10.1002/prot.20810 (2006).
30. Savojardo, C., Fariselli, P., Martelli, P. L. & Casadio, R. INPS-MD: a web server to predict stability of protein variants from sequence and structure. *Bioinformatics* **32**(16), 2542–2544. https://doi.org/10.1093/bioinformatics/btw192 (2016).
31. Ashkenazy, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* **44**(W1), W344–50. https://doi.org/10.1093/nar/gkw408 (2016).
32. Wiederstein, M. & Sippl, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* **35**, W407–10 https://doi.org/10.1093/nar/gkm290 (2007).
33. Colovos, C. & Yeates, T. O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci* **2**, 1511–1519 https://doi.org/10.1002/pro.5560020916 (1993).
34. Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* **33**(7), 2302–2309. https://doi.org/10.1093/nar/gki524 (2005).
35. Teng, Z. et al., The 677C > T (rs1801133) polymorphism in the MTHFR Gene contributes to Colorectal Cancer risk: A Meta-analysis based on 71 Research studies. *PLoS One* **8**(2), e55332. https://doi.org/10.1371/journal.pone.0055332 (2013).
36. Han, Z., Sheng, H., Gao, Q., Fan, Y. & Xie, X. Associations of the mthfr rs1801133 polymorphism with gastric cancer risk in the Chinese Han population. *Biomed. Rep.* **14**(1), 1–7. https://doi.org/10.3892/br.2020.1390 (2021).
37. Lin, K. M. et al. The role of MTHFR genotype in colorectal cancer susceptibility in Taiwan. *Anticancer Res* **38**(4), 2001–2006 https://doi.org/10.21873/anticanres.12438 (2018).
38. Farshid, S. et al. Associations of MTHFR rs1801133 (677C > T) and rs180113 (1298A > C) polymorphisms with susceptibility to bladder Cancer: A systematic review and Meta-analysis. *Asian Pac. J. Cancer Prev.* **23**(5), 1465–1482. https://doi.org/10.31557/APJCP.2022.23.5.1465 (2022).
39. Liu, F. et al. Methylenetetrahydrofolate Reductase (MTHFR) gene rs1801133 C > T polymorphisms and Lung Cancer susceptibility: an updated Meta-analysis. **30**(1), 65–73. https://doi.org/10.1515/pteridines-2019-0008 (2019).
40. Terry, K. L. et al. MTHFR polymorphisms in relation to ovarian cancer risk. *Gynecol Oncol* **119**(2), 319–324. https://doi.org/10.1016/j.ygyno.2010.08.007 (2010).
41. Cecchin, E. et al. MTHFR-1298 A > C (rs1801131) is a predictor of survival in two cohorts of stage II/III colorectal cancer patients treated with adjuvant fluoropyrimidine chemotherapy with or without oxaliplatin. *Pharmacogenomics J.* **15** (3), 219–225. https://doi.org/10.1038/tpj.2014.64 (2015).
42. Liang, T-J. et al. Quantitative assessment of the association between MTHFR rs1801131 polymorphism and risk of liver cancer. *Tumor Biology* **35**(1), 339–343 https://doi.org/10.1007/s13277-013-1046-y (2014).
43. López-Cortés, et al. Breast cancer risk associated with gene expression and genotype polymorphisms of the folate-metabolizing MTHFR gene: a case-control study in a high altitude Ecuadorian mestizo population. *Tumor Biology.Bold"*>*36*(8), 6451–6461. https://doi.org/10.1007/s13277-015-3335-0 (2015).
44. Hardi, H. et al. Significant association between ERCC2 and MTHR polymorphisms and breast cancer susceptibility in Moroccan population: genotype and haplotype analysis in a case-control study. *BMC Cancer*. **18** (1), 292. https://doi.org/10.1186/s12885-018-4214-z (2018).
45. Sangrajrang, S. et al. Genetic polymorphisms in folate and alcohol metabolism and breast cancer risk: a case–control study in Thai women. *Breast Cancer Res. Treat.* **123** (3), 885–893. https://doi.org/10.1007/s10549-010-0804-4 (2010).
46. Lu, Q. et al. Polymorphisms in the MTHFR gene are associated with breast cancer risk and prognosis in a Chinese population. *Tumor Biology*. **36**(5), 3757–3762 https://doi.org/10.1007/s13277-014-3016-4 (2015).
47. Hesari, et al. Evaluation of the two polymorphisms rs1801133 in MTHFR and rs10811661 in CDKN2A/B in breast cancer. *J Cell Biochem* **120**(2), 2090–2097. https://doi.org/10.1002/jcb.27517 (2019).
48. Kaya, E. F., Karakus, N., Ulusoy, A. N., Özaslan, C. & Kara, N. Association of the MTHFR Gene C677T polymorphism with breast Cancer in a Turkish Population. *Oncol Res Treat* **39**(9), 534–538. https://doi.org/10.1159/000448084 (2016).

49. Awwad, A. -M., Yousef, A., Abuhaliema, I., Abdalla & Yousef, M. Relationship between genetic polymorphisms in MTHFR (C677T, A1298C and their haplotypes) and the incidence of breast Cancer among Jordanian females–case-control study. *Asian Pac. J. Cancer Prev.* **16** (12), 5007–5011. https://doi.org/10.7314/apjcp.2015.16.12.5007 (2015).
50. Li, S. et al. Associations of maternal rs1801131 genotype in MTHFR and serum folate and vitamin B12 with gestational diabetes Mellitus in Chinese pregnant women. *Nutrients*, **14**. https://doi.org/10.3390/nu14061169 (2022).
51. Boughrara, W. et al. Impact of MTHFR rs1801133, MTHFR rs1801131 and ABCB1 rs1045642 polymorphisms with increased susceptibility of rheumatoid arthritis in the West Algerian population: a case-control study. *Acta Reumatol Port*. **40** (4), 363–371 (2015).
52. Yuan, L. et al. Association of the MTHFR rs1801131 and rs1801133 variants in sporadic Parkinson's disease patients. *Neurosci. Lett.* **616**, 26–31. https://doi.org/10.1016/j.neulet.2016.01.031 (2016).
53. Hubacek, A. et al. Association of MTHFR genetic variants C677T and A1298C on predisposition to spontaneous abortion in slavonic population. *Clin. Chim. Acta*. **440**, 104–107. https://doi.org/10.1016/j.cca.2014.11.018 (2015).
54. Luo, Z. et al. Associations of the MTHFR rs1801133 polymorphism with coronary artery disease and lipid levels: a systematic review and updated meta-analysis. *Lipids Health Dis.* **17** (1), 191. https://doi.org/10.1186/s12944-018-0837-y (2018).
55. Zhao, et al. Association of methylenetetrahydrofolate reductase (MTHFR) rs1801133 (677C > T) gene polymorphism with ischemic stroke risk in different populations: an updated meta-analysis. *Front. Genet.* **13**. https://doi.org/10.3389/fgene.2022.1021423 (2023).
56. Waseem, M. et al. Association of MTHFR (C677T) Gene Polymorphism with Breast Cancer in North India. *Biomark Cancer* **8**, BIC.S40446. https://doi.org/10.4137/BIC.S40446 (2016).
57. Akram, F. A., Malik & Kayani, M. A. Mutational analysis of the MTHFR gene in breast cancer patients of Pakistani population. *Asian Pac. J. Cancer Prev.* **13** (4), 1599–1603. https://doi.org/10.7314/apjcp.2012.13.4.1599 (2012).
58. Sahaba, S. A. et al. The link of ERCC2 rs13181 and ERCC4 rs2276466 polymorphisms with breast cancer in the Bangladeshi population. *Mol. Biol. Rep.* **49** (3), 1847–1856. https://doi.org/10.1007/s11033-021-06994-7 (2022).
59. Parvin, S. et al. Association of BRCA1, BRCA2, RAD51, and HER2 gene polymorphisms with the breast cancer risk in the Bangladeshi population. *Breast Cancer*. **24** (2), 229–237. https://doi.org/10.1007/s12282-016-0692-5 (2017).
60. Shabnaz, S. et al. Breast cancer risk in relation to TP53 codon 72 and CDH1 gene polymorphisms in the Bangladeshi women. *Tumor Biology*. **37**(6), 7229–7237. https://doi.org/10.1007/s13277-015-4612-7 (2016).
61. Howlader, R. et al. Genetic polymorphisms in DNA repair genes XRCC1 and 3 are associated with increased risk of breast cancer in Bangladeshi population. *Breast Cancer Res. Treat.* **182** (3), 739–750. https://doi.org/10.1007/s10549-020-05738-8 (2020).
62. Jahan et al. Evaluation of the association between FGFR2 gene polymorphisms and breast Cancer risk in the Bangladeshi Population. *Genes* **14**(4). https://doi.org/10.3390/genes14040819 (2023).
63. Ziva Cerne, J., Stegel, V., Gersak, K. & Novakovic, S. Lack of association between methylenetetrahydrofolate reductase genetic polymorphisms and postmenopausal breast cancer risk. *Mol. Med. Rep.* **4** (1), 175–179. https://doi.org/10.3892/mmr.2010.406 (2011).
64. Semenza, J. C., Delfino, R. J., Ziogas, A. & Anton-Culver, H. Breast Cancer Risk and Methylenetetrahydrofolate Reductase Polymorphism. *Breast Cancer Res. Treat.* **77** (3), 217–223. https://doi.org/10.1023/A:1021843019755 (2003).
65. Maruti, S. S., Ulrich, C. M., Jupe, E. R. & White, E. MTHFR C677T and postmenopausal breast cancer risk by intakes of one-carbon metabolism nutrients: a nested case-control study. *Breast Cancer Res.* **11** (6), 1–9. https://doi.org/10.1186/bcr2462 (2009).
66. Ergul, E., Sazci, A., Utkan, Z. & Canturk, N. Z. Polymorphisms in the MTHFR Gene Are Associated with Breast Cancer. *Tumor Biology*https://doi.org/10.1159/000076460 (2004).

Acknowledgements
Not Applicable.

Author contributions
Conceptualization: ZHH; Study design: ZHH, ZM, ST; Sample collection: NFA, MAS; Data curation and analysis: NFA, RA, ZM, ST, MAS, ZHH; Result interpretation: NFA, RA, ZM, ST, MAS, ZHH; Supervision: ZHH, ZM, ST; Writing original draft: NFA, RA, ZM, ZHH; Writing revised draft: NFA, RA, ZM, ZHH; Review and editing: ZM, ST, ZHH.

Declarations

Competing interests
The authors declare no competing interests.

Disclosure
We declare that the work was carried out by the concerned authors & co-authors and all the authors have read the draft manuscript.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-75656-y.

Correspondence and requests for materials should be addressed to M.Z.H.H.

Reprints and permissions information is available at www.nature.com/reprints.

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