medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Independent and Combined Associations of 677C/T and 1298A/C Polymorphisms in the *MTHFR* gene with Breast Cancers in a South-Asian Population

Sadia Ajaz[*1], Sani-e-Zehra Zaidi[1], Saleema Mehboob Ali[1], Aisha Siddiqa[2], Muhammad Ali Memon[2], Aiysha Abid[3], Shagufta Khaliq[4]

*** Corresponding Author:**

sadiaajaz1@gmail.com

Affiliations[o]:

1. Dr. Panjwani Center for Molecular Medicine and Drug Research (PCMD), International Center for Chemical and Biological Sciences (ICCBS), University of Karachi, Karachi, Pakistan.

2. Atomic Energy Medical Centre (AEMC), Jinnah Postgraduate Medical Centre (JPMC), Karachi, Pakistan.

3. Center for Human Genetics and Molecular Medicine, Sindh Institute of Urology and Transplantation (SIUT), Karachi, Pakistan.

4. Department of Human Genetics and Molecular Biology, University of Health Sciences, Lahore, Pakistan.

[o]for the Duration of Reported Research

Conflict of Interest: The authors declare no conflict of interest.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Abstract:

MTHFR is a pivotal enzyme in the folic acid cycle. Two functional SNPs, rs 1801133 (677C/T) and rs 1801131 (1298A/C), which affect the function of the MTHFR, are associated with different cancers. In the present study, these SNPs were investigated in breast cancer patients from the Pakistani population. The pilot study includes 187 participants with 124 breast cancer patients and 63 medically confirmed healthy individuals as controls. PCR-RFLP methods validated by Sanger sequencing were used for the polymorphic investigations. Here, we report the significant and unique associations of these polymorphisms with breast cancers in a South-Asian population for the first time in the literature. The case-control analysis showed that in case of 1298A/C polymorphism, a significant protective effect of homozygous C genotype was observed in recessive [CC vs AA+AC; OR: 0.320 (95% CI: 0.259 – 0.397)] and co-dominant multiplicative [CC vs AA; OR 0.379 (95% CI: 0.273 – 0.527)] models. In the case of 677C/T analysis, no significant association was observed with the risk of breast cancers. However, homozygous T genotype was more frequent in the advanced age group (>35 years) patients as compared to the young age-group (<35 years) i.e. 6.7% vs 0%. Thus, carriers of this genotype are less likely to develop breast cancers at younger age. The combined genotype analysis at two loci revealed that 677CC+1298AC [OR: 2.688 (95% CI: 1.247-5.795)] and 677CT+1298AA [OR.: 20.91 (95% CI: 1.156-378.2)] increased risk of breast cancers, significantly. The latter association (677T*1298A) was also observed in a semi-parametric haplotype analysis (p-value: 0.03). Interestingly, despite the proximity, these loci were in linkage equilibrium (r^2 = 0.029 and 0.049 in cases and controls, respectively). The study implies translational potentials of these polymorphisms for breast cancer management in the studied population.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

1 INTRODUCTION

Methylene Tetrahydrofolate Reductase (MTHFR) (UniProtKB-P42898) is a 77kDa key enzyme of the folate pathway. It converts 5-10 methylene tetrahydrofolate (5,10 MTHF) to 5-methyl tetrahydrofolate (5 MTHF) (Chittiboyina, Chen, Chiorean, Kamendulis, & Hocevar, 2018). The latter is a co-substrate in the homocysteine-based re-methylation of methionine. The folate cycle is involved in DNA synthesis, repair and methylation, as well as detoxification pathways (Kawakita et al., 2017). The genomic instability and aberrant DNA methylation are two critical hallmarks of cancer(s) (Flavahan, Gaskell, & Bernstein, 2017; Macheret & Halazonetis, 2015). Thus, folate pathway provides protection against neoplastic transformation and progression (Friso, Udali, De Santis, & Choi, 2017; Stover, James, Krook, & Garza, 2018).

Decreased MTHFR activity has been associated with gastrointestinal stromal tumour, neural tube defects, folate sensitivity, MTHFR-deficiency, schizophrenia, as well as dosage and toxicity response to adriamycin and cyclophosphamide (M. J. Landrum et al., 2016; OMIM #235250). The gene for MTHFR is located on chromosome 1p36.3. It is one of the ten most investigated genes globally, due to its potential in clinical applications (Dolgin, 2017). Two functional Single Nucleotide Polymorphisms (SNPs), 677C/T (rs1801133) and 1298A/C (rs1801131) encode for thermolabile isoforms of the MTHFR, affecting its enzymatic activity. The SNP 677C/T results in the substitution of alanine with valine (Ala222Val). In 1298A/C variant, glutamic acid is replaced with alanine (Glu429Ala) (Nefic, Mackic-Djurovic, & Eminovic, 2018). These SNPs result in decreased enzymatic activity: almost 60% reduced activity as compared to the wild type (D'Angelo et al., 2011).

In cancer cell, thermolabile isoform of the MTHFR leads to the mis-incorporation of uracil instead of thymine with consequent DNA damage (Rai, 2014). Additionally, reduced enzymatic

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

activity affects DNA methylation and hence dysregulates the gene expression (Friso et al., 2002).

The SNPs are located within a few kilobases of each other and due to the short distance are expected to be in high degree of linkage disequilibrium (LD) (Balding, 2006; Bodmer & Bodmer, 1978). However, several discrepancies in other closely-related genomic regions have been reported (Schaid, Chen, & Larson, 2018). Therefore, the quantitative value of LD among proximate genetic variations need to be assessed separately for each population. This is essential to map the population-specific contribution of closely situated polymorphisms in the disease phenotype.

Data regarding the role of these functional SNPs in the breast cancers remains inconclusive in different populations (McEwen, 2016; Naushad et al., 2016). In the developing countries, breast cancer is the leading cause of cancer-related deaths among females (Torre, Siegel, Ward, & Jemal, 2015). In Pakistan, age-specific incidence and mortality rates of breast cancer(s) are one of the highest among Asian countries and globally, respectively. Given this background, the present study is a population-focused systematic attempt to investigate the role of these functional *MTHFR* SNPs in breast cancers in Pakistani population. The analysis includes individual and combined contribution of these two loci, as well as the evaluation of LD, in the risk and pathology of breast cancers in this population.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

2 METHODOLOGY

2.1 Ethics Statement

The study was conducted in accordance with the Declaration of Helsinki (WMA, 2018). The project was approved by the ethical review committees (ERCs) of the participating institutions: the independent ERC, International Center for Chemical and Biological Sciences (ICCBS), University of Karachi, Karachi, Pakistan [ICCBS/IEC-016-BS/HT-2016/Protocol/1.0], and the Atomic Energy Medical Centre (AEMC), Jinnah Postgraduate Medical Centre (JPMC), Karachi, Pakistan [Admin-3(257)/2016]. All the samples were collected after obtaining written informed consent from each participant. The ethical approval from controls has been submitted and published earlier (Ajaz S, et al. 2012).

2.2 Study Participants

The pilot study comprised case-control design and included 187 participants. The cases included 124 diagnosed and histologically confirmed primary breast cancer patients. The patients were either first-time visitors or under regular treatment at (AEMC), JPMC, Karachi, Pakistan during the period of July 2016 – July 2017. The exclusion criteria for patients were: lack of biopsy report for breast cancer. In case of controls, 63 samples of medically-confirmed, healthy individuals were included (Ajaz S et al., 2012). The controls were matched on the basis of age, gender and ethnicity. The exclusion criteria for controls were: age or gender based mismatch, previous diagnosis of any cancer or co-morbidity with any other chronic disease. The participants belonged to Southern-Pakistan.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

2.3 Breast Cancer Data

At the time of sampling, after obtaining written informed consent, participants' relevant information was obtained through a questionnaire, including age, ethnicity, place of residence, contact number, family history of cancers, age at menarche, obstetrics and gynaecology history, and if applicable, the age at menopause. Clinical characteristics including, tumor histology, size, grade, stage, axillary lymph node metastasis, ER status, PR status, and Her-2 status were collected from the patients' medical records.

2.4 DNA Extraction and Genotyping

DNA samples were extracted using standard phenol-chloroform method with slight modifications (Sambrook & Russell, 2001). Subsequently, target DNA sequences were amplified for the identification of selected genetic polymorphisms in the *MTHFR*. The restriction sites for two polymorphisms are shown in Figure 1.

2.4.1 MTHFR 677C/T

For 677C/T, DNA fragment of 198bp was amplified in 25µl total volume. PCR mix contained 1X $(NH_4)_2SO_4$ PCR buffer, 0.4mM $MgCl_2$, 0.2mM dNTPs, 0.4U Taq polymerase, 0.35µM of each primer (Forward primer: 5'-TGAAGGAGAAGGTGTCTGCGGGA-3'; Reverse primer: 5'-AGGACGGTGCGGTGAGAGTG-3') and 130ng of DNA template in a final reaction volume of 25 µl. Cycling conditions have already been published (Ajaz et al., 2012). Amplification was carried out at annealing temperature (T_a) of 64°C. 20µl of the amplified product was digested overnight with 10U of *HinfI* and 3.5µl of the recommended buffer (Thermo Scientific®, USA). The digested products were run on 10% polyacrylamide gel, stained with ethidium bromide, and observed under UV.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

2.4.2 MTHFR 1298A/C

For 1298A/C, DNA fragment of 168bp was amplified. PCR mix of 25µl contained 1X $(NH_4)_2SO_4$ based PCR buffer with 0.8mM of $MgCl_2$, 0.25mM of dNTPs, 2U of Taq polymerase, 1.25µM of each primer (Forward primer: 5'-CTTTGGGGAGCTGAAGGACTACTA-3'; 5'-CACTTTGTGACCATTCCGGTTTG-3') and 70ng of DNA template. Amplification was carried out at T_a of 60°C. Cycling conditions were the same as for 677C/T. 20µl of the amplified product was digested overnight with 2.5U of *MboII* and 2µl of the recommended buffer (Thermo Scientific®, USA). The digested products were run on 10% polyacrylamide gel, stained with ethidium bromide, and observed under UV. Results of both polymorphisms were validated by Sanger sequencing.

Representative gels for genotyping of the *MTHFR* 677C/T and 1298A/C are shown in Figures 2 and 3, respectively. The validation of each methodology by Sanger sequencing is shown in the inset.

2.5 Bioinformatics Analyses

In order to predict the impact of the missense substitutions on the protein function, bio-informatic analyses were performed. The softwares for investigation included MutPred2 (Balding, 2006), Provean (Gaunt, Rodríguez, & Day, 2007), and Polyphen2 (Adzhubei IA, 2010). The variant correlation with the disease-related phenotype was investigated using ClinVar (Landrum MJ et al., 2015) and with cancer was investigated by using CIViC database (Griffith M et al., 2017).

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

2.6 Statistical Analyses

Statistical analyses were carried out using IBM (SPSS®) v.21.0 software (Arbuckle, 2012). The genotype and allele frequencies were determined by gene counting method along with the weighted percentages. Genotype distributions among the cases and controls were analyzed for Hardy-Weinberg equilibrium using Chi-squared test. The association between qualitative variables such as clinico-pathological characteristics including tumour stages and grade were also assessed by Pearson's Chi-squared test (Cochran-Armitage trend test was used for the assessment of the association where Hardy-Weinberg disequilibrium was observed (Balding, 2006). In order to measure the allelic, genotypic and haplotype risks for breast cancers, odds ratios (OR) with 95% confidence interval (95% CI) were calculated using logistic regression. For all the analyses p-value <0.05 was considered to be significant.

2.7 Haplotype Estimation and LD Statistics

Estimation of haplotype frequencies and LD analysis were carried out using cubeX webtool (Gaunt et al., 2007). The maximum likelihood estimations were carried out for the calculations of haplotype frequencies. The programme was used for the calculation of Lewontin's standardized disequilibrium coefficient (D´), correlation co-efficient (r^2) and χ^2 test (with significant value at $p<0.05$) were used for the estimation of LD between two loci. The results were compared with the Phase 3 (version 5) 1000 Genomes Project data for different populations on LDlink, a National Cancer Institute website tool (Machiela MJ & Chanock SJ, 2015).

3 RESULTS

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

The average age of cases was 44.35 ± 0.932 years while that of the controls was 46.04 ± 0.80 years, respectively. Clinico-pathological data for the studied cohort is shown in Table 1. Briefly, majority of the patients had invasive ductal carcinoma (79%) with tumour size >5cm (48%) and stage III (55%), grade 3 (52%) tumours.

3.1. Hardy-Weinberg Equilibrium Estimation

The distributions of genotypes and allele frequencies of the *MTHFR* 677C/T and 1298A/C polymorphisms are given in Tables 2 and 3, respectively. Genotypes and allele distributions, for both markers, were determined in the groups of breast cancer patients, controls and the subject population combining both the groups. χ^2 values show that the distributions of genotypes for *MTHFR* 677C/T occurred in Hardy-Weinberg proportions in controls, and breast cancer patients, however, in the combined group these were not in Hardy-Weinberg equilibrium. The genotypes were in Hardy-Weinberg dis-equilibrium for *MTHFR* 1298A/C in all the groups and the disequilibrium is highly significant in the breast cancer patients and the combined group, implying the role of this polymorphism in breast cancer susceptibility.

3.2. Significant Differences in the Distribution of the *MTHFR* 677C/T Genotypes in Breast Cancers on the Basis of Age:

The distribution of the *MTHFR* 677C/T polymorphism and genotypes on the basis of age (≤35 years) and advanced age-group (>35years) are shown in Table 4. No overall significant associations were found between the polymorphism and the risk for breast cancers. However, Homozygous T genotype is associated with advanced age, while heterozygous CT is associated with younger age-group of the patients.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

3.3. Significant Protective Effect of the *MTHFR* 1298 C-variant with Polymorphism in Breast Cancer

In the studied cohort, a significant protective effect of CC-genotype in *MTHFR* 1298A/C polymorphism against breast cancer was observed in Pakistani population. Statistical analysis of this SNP with breast cancers is shown in Table 5.

3.4. Significant Associations of Combined Genotypes of 677CC/1298AC and 677CT/1298AA with the Risk for Breast Cancers

The nine possible genotype combinations for two *MTHFR* polymorphisms are shown in Table 6. Case-control association analysis was carried out for each combination. The analysis revealed that homozygous C at 677C/T with heterozygous AC at 1298A/C genotypic combination when compared with reference genotype (CC/AA) significantly increased the risk of breast cancers (χ^2 = 6.558, OR: 2.688, 95%CI: 1.247-5.795. Similarly, heterozygous 677CT with 1298 homozygous A (CT/AA) when compared with (CC/AA) showed significant association (p value<0.05, OR: 20.91, 95% CI: 1.156-378.2) with increased breast cancer susceptibility (Table 6).

3.5. LD Analysis at *MTHFR* 677C/T and 1298A/C Locus:

The disequilibrium spread in pairwise allelic combinations at the *MTHFR* 677C/T and 1298A/C loci was quantified by maximum likelihood calculation from the frequency of diploid genotypes.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Haplotype frequencies and the LD statistics, D´, r^2, and χ^2 values demonstrate that the variants at two loci are not associated (Table 7).

3.6. Bioinformatics analysis:

The bioinformatics analysis for the effect of SNPs on protein function is shown in Table 8. *MTHFR* 677C/T was predicted to be deleterious by three software tools. On the other hand, *MTHFR* 1298A/C was predicted to have benign effect by two programmes, while deleterious by PROVEAN analysis (Table 8).

4 DISCUSSION

Breast cancers are a heterogeneous group of disorders, which are characterized by abnormal cellular proliferation in the mammary tissue (Kumar, Abbas, Aster, & Robbins, 2013). The contributing gene effects are complex in nature, which can vary in penetrance (Haines & Pericak-Vance, 2006). The present study is a component of the molecular genetic investigation of the breast cancers in Pakistani population. The aim is to decipher the molecular and genetic architecture of the breast cancers in a population, characterized by high consanguinity, multiparity, and low exposure to putative risk factors for breast cancers, particularly low levels of alcohol and pork consumption. Thus, the present population provides a unique model for the epidemiological and molecular investigations of breast cancers. Additionally, folate deficiency is an established severe public health concern in women of reproducible ages (Soofi et al., 2017). This insufficiency, often attributed to low intake, is likely to be compounded by the genetic factors, with consequent effect on the breast cancer spectrum observed across Pakistani population.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

In the present study, we report significant associations of the *MTHFR* 677C/T (rs 1801133) and 1298A/C (rs 1801131) polymorphisms with the risk of breast cancers in Pakistani population. To the best of our knowledge, the age-specific differences in the distribution of the *MTHFR* 677C/T genotypes in breast-cancer patients are reported for the first time from any population. As shown in Table 8, the bio-informatic analysis of the *MTHFR* 677C/T polymorphism demonstrates that Ala222Val substitution has deleterious effect on the enzyme function. The breast cancer patients with *MTHFR* 677 heterozygous genotype (CT) aggregates in the young (≤35 years) group [OR: 2.81; 95% CI: 1.08 - 7.3]. On the other hand, all the homozygous T genotype breast cancer patients were found in the advanced age group (>35 years) [OR: 1.281; 95% CI: 1.162 – 1.412].

The other important observation in the present study is the lack of *MTHFR* 1298CC genotype in breast cancer patients. Additionally, the Mid-P chi-square test and the conditional maximum likelihood estimate (CMLE) odds ratio show that the protective effect is highly significant. A number of studies from different regions have reported the protective effect of C-allele against cancers including leukaemias and bladder cancer (Robien & Ulrich, 2003; Skibola et al., 1999; You et al., 2013). Mechanistically, it is proposed that the double-stranded breaks lead to chromosomal instability, translocation and aberrations with consequent increase in cancer risk and progression. It has been shown through the comet assay that C-allele decreases the frequency of double-stranded breaks, thereby conferring a protective effect against cancers (Fragkioudaki et al., 2017; Jackson & Bartek, 2009). Secondly, decreased activity of the MTHFR results in the accumulation of the substrate 5,10 THF (Figure 4). Although the effect of increased concentration of 5, 10 THF has not been investigated recently, the earlier studies showed subsequent increased purine and pyrimidine synthesis. Hence, the DNA replication becomes stable and has the least chance of mutations (Bagley & Selhub, 1998; Chen et al., 1996;

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Fintelman-Rodrigues, Correa, Santos, Pimentel, & Santos-Reboucas, 2009; Giovannucci et al., 1998; Ulrich et al., 1999).

In the genotype combination analysis (*MTHFR* 677C/T and 1298A/C), CT/AA combination genotype was not observed in 63 controls as compared to 11 cases in 124 breast cancer patients [OR: 20.91; 95% CI: 1.156 – 378.2]. A second combination i.e. CC/AC also showed significant association with breast cancers [OR: 2.68; 95% CI: 1.25 – 5.8]. The pattern in haplotype analysis also supports this result (χ^2 test p-value = 0.03). Both of these heterozygous/homozygous combination genotypes (CC/AC and CT/AA) were also associated with the increased risk of renal cell carcinoma (RCC) in Pakistani population (Ajaz et al., 2012).

Three genotype combinations are absent in the studied breast cancer patients: CT/CC, CC/CC, and TT/CC. These results underscore the significant protective effect of the *MTHFR* 1298CC genotype against breast cancers and the comparatively young age of breast cancer patients in the present cohort (average age: 44.35 $\pm$ 0.932). Interestingly, TT/CC combination was absent in controls and RCC patients as well (Ajaz et al., 2012).

The linkage equilibrium between two loci, which are in close proximity is a unique finding of the current report. In 1000 genome project (Kumar et al., 2013), among South-Asian populations, these two SNPS have been shown to be in LD in PJL (Punjabis from Lahore), GIH (Gujrati Indians in Houston), ITU (Indian Telugu from the UK) populations; whereas these are known to be not associated in BEB (Bengali from Bangladesh), and STU (SriLankan Tamil from the UK). In the present study, in contrast to PJL, where D´ has been shown to be 1, and r^2 to be 0.119, for breast cancer patients these values were -0.502, and 0.0273, respectively. In case of controls, the values were 0.41 and 0.0496, respectively. Thereby, lack of association between the variants at

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

studied loci is reported in the present study. Thus, underscoring the importance of validating data from genome wide association studies by appropriately designed case-control studies.

Conclusions: The present study reports unique associations of the *MTHFR* 677C/T and *MTHFR* 1298A/C independently and in combinations with breast cancers in Pakistani population. A distinctive lack of association over the short genomic sequence is also a significant finding of the present molecular investigative report. The results have important implications in devising folate pathway specific strategies for breast cancer management in distinctive sub-population groups.

Acknowledgements: The authors wish to thank Pakistan Health Research Council for funding the study. ICCBS for core facilities. AEMC, JPMC staff for their co-operation. The authors are especially grateful to the participants in the study.

Conflict of Interest: The authors declare no conflict of interest.

References

1. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods. 2010 Apr;7(4):248-9. doi: 10.1038/nmeth0410-248. PMID: 20354512; PMCID: PMC2855889.

2. Ajaz, S., Khaliq, S., Hashmi, A., Naqvi, S. A. A., Rizvi, S. A.-u.-H., & Mehdi, S. Q. (2012). Polymorphisms in the methylene tetrahydrofolate reductase gene and their unique combinations are associated with an increased susceptibility to the renal cancers. *Genetic testing and molecular biomarkers, 16*(5), 346-352.

3. Arbuckle, J. L. (2012). IBM SPSS Amos 21. *Chicago, IL: Amos Development Corporation*.

4. Bagley, P. J., & Selhub, J. (1998). A common mutation in the methylenetetrahydrofolate reductase gene is associated with an accumulation of formylated tetrahydrofolates in red blood cells. *Proceedings of the National Academy of Sciences, 95*(22), 13217-13220.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

5. Balding, D. J. (2006). A tutorial on statistical methods for population association studies. *Nature reviews genetics, 7*(10), 781.

6. Bodmer, W., & Bodmer, J. (1978). Evolution and function of the HLA system. *British Medical Bulletin, 34*(3), 309-316.

7. Chen, J., Giovannucci, E., Kelsey, K., Rimm, E. B., Stampfer, M. J., Colditz, G. A., . . . Hunter, D. J. (1996). A methylenetetrahydrofolate reductase polymorphism and the risk of colorectal cancer. *Cancer research, 56*(21), 4862-4864.

8. Chittiboyina, S., Chen, Z., Chiorean, E. G., Kamendulis, L. M., & Hocevar, B. A. (2018). The role of the folate pathway in pancreatic cancer risk. *PloS one, 13*(2), e0193298.

9. D'Angelo, V., Ramaglia, M., Iannotta, A., Crisci, S., Indolfi, P., Francese, M., . . . Fusco, C. (2011). Methotrexate toxicity and efficacy during the consolidation phase in paediatric acute lymphoblastic leukaemia and MTHFR polymorphisms as pharmacogenetic determinants. *Cancer chemotherapy and pharmacology, 68*(5), 1339-1346.

10. Dolgin, E. (2017). The most popular genes in the human genome. *Nature, 551*(7681), 427-431. doi:10.1038/d41586-017-07291-9

11. Fintelman-Rodrigues, N., Correa, J. C., Santos, J. M., Pimentel, M. M., & Santos-Reboucas, C. B. (2009). Investigation of CBS, MTR, RFC-1 and TC polymorphisms as maternal risk factors for Down syndrome. *Dis Markers, 26*(4), 155-161. doi:10.3233/dma-2009-0626

12. Flavahan, W. A., Gaskell, E., & Bernstein, B. E. (2017). Epigenetic plasticity and the hallmarks of cancer. *Science, 357*(6348). doi:10.1126/science.aal2380

13. Fragkioudaki, S., Nezos, A., Souliotis, V. L., Chatziandreou, I., Saetta, A. A., Drakoulis, N., . . . Koutsilieris, M. (2017). MTHFR gene variants and non-MALT lymphoma development in primary Sjogren's syndrome. *Scientific reports, 7*(1), 7354.

14. Friso, S., Choi, S.-W., Girelli, D., Mason, J. B., Dolnikowski, G. G., Bagley, P. J., . . . Corrocher, R. (2002). A common mutation in the 5, 10-methylenetetrahydrofolate

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

reductase gene affects genomic DNA methylation through an interaction with folate status. *Proceedings of the National Academy of Sciences, 99*(8), 5606-5611.

15. Friso, S., Udali, S., De Santis, D., & Choi, S.-W. (2017). One-carbon metabolism and epigenetics. *Molecular aspects of medicine, 54*, 28-36.

16. Gaunt, T. R., Rodríguez, S., & Day, I. N. (2007). Cubic exact solutions for the estimation of pairwise haplotype frequencies: implications for linkage disequilibrium analyses and a web tool'CubeX'. *BMC bioinformatics, 8*(1), 428.

17. Giovannucci, E., Stampfer, M. J., Colditz, G. A., Hunter, D. J., Fuchs, C., Rosner, B. A., . . . Willett, W. C. (1998). Multivitamin use, folate, and colon cancer in women in the Nurses' Health Study. *Annals of internal medicine, 129*(7), 517-524.

18. Griffith M, Spies NC, Krysiak K, McMichael JF, Coffman AC, Danos AM, Ainscough BJ, Ramirez CA, Rieke DT, Kujan L, Barnell EK, Wagner AH, Skidmore ZL, Wollam A, Liu CJ, Jones MR, Bilski RL, Lesurf R, Feng YY, Shah NM, Bonakdar M, Trani L, Matlock M, Ramu A, Campbell KM, Spies GC, Graubert AP, Gangavarapu K, Eldred JM, Larson DE, Walker JR, Good BM, Wu C, Su AI, Dienstmann R, Margolin AA, Tamborero D, Lopez-Bigas N, Jones SJ, Bose R, Spencer DH, Wartman LD, Wilson RK, Mardis ER, Griffith OL. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. Nat Genet. 2017 Jan 31;49(2):170-174. doi: 10.1038/ng.3774. PMID: 28138153; PMCID: PMC5367263.

19. Haines, J. L., & Pericak-Vance, M. A. (2006). *Genetic Analysis of Complex Disease*: Wiley.

20. Institute, T. M. G. CIViC. Retrieved from https://civicdb.org

21. Institute, T. M. G. CIViC. Retrieved from https://civicdb.org/home

22. Jackson, S. P., & Bartek, J. (2009). The DNA-damage response in human biology and disease. *Nature, 461*(7267), 1071-1078. doi:10.1038/nature08467

23. Kawakita, D., Lee, Y.-C. A., Gren, L. H., Buys, S. S., La Vecchia, C., & Hashibe, M. (2017). The impact of folate intake on the risk of head and neck cancer in the prostate, lung, colorectal, and ovarian cancer screening trial (PLCO) cohort. *British Journal Of Cancer, 118*, 299. doi:10.1038/bjc.2017.383

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

24. Kumar, V., Abbas, A. K., Aster, J. C., & Robbins, S. L. (2013). *Robbins Basic Pathology*: Elsevier/Saunders.

25. Landrum, M. J., Lee, J. M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., . . . Hoover, J. (2015). ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic acids research, 44*(D1), D862-D868.

26. Landrum, M. J., Lee, J. M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., . . . Maglott, D. R. (2016). ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res, 44*(D1), D862-868. doi:10.1093/nar/gkv1222

27. Macheret, M., & Halazonetis, T. D. (2015). DNA replication stress as a hallmark of cancer. *Annual Review of Pathology: Mechanisms of Disease, 10*, 425-448.

28. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics. 2015 Nov 1;31(21):3555-7. doi: 10.1093/bioinformatics/btv402. Epub 2015 Jul 2. PMID: 26139635; PMCID: PMC4626747. (https://ldlink.nci.nih.gov/.)

29. McEwen, B. J. (2016). Methylenetetrahydrofolate reductase (MTHFR): Mythology or polymorphism (ology)? *Advances in Integrative Medicine, 3*(3), 79-81.

30. Naushad, S. M., Divya, C., Janaki Ramaiah, M., Hussain, T., Alrokayan, S. A., & Kutala, V. K. (2016). Population-level diversity in the association of genetic polymorphisms of one-carbon metabolism with breast cancer risk. *J Community Genet, 7*(4), 279-290. doi:10.1007/s12687-016-0277-1

31. Nefic, H., Mackic-Djurovic, M., & Eminovic, I. (2018). The Frequency of the 677C> T and 1298A> C Polymorphisms in the Methylenetetrahydrofolate Reductase (MTHFR) Gene in the Population. *Medical Archives, 72*(3), 164-169.

32. OMIM. Retrieved from https://www.omim.org/entry/607093?search=mthfr&highlight=mthfr

33. project, g. Retrieved from http://www.internationalgenome.org/

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

34. Rai, V. (2014). Methylenetetrahydrofolate reductase A1298C polymorphism and breast cancer risk: A meta analysis of 33 studies. *Annals of medical and health sciences research*, *4*(6), 841-851.

35. Resource, I. T. I. G. S. 1000 genome Project. Retrieved from http://www.internationalgenome.org

36. Robien, K., & Ulrich, C. M. (2003). 5,10-Methylenetetrahydrofolate Reductase Polymorphisms and Leukemia Risk: A HuGE Minireview. *American Journal of Epidemiology*, *157*(7), 571-582. doi:10.1093/aje/kwg024

37. Sambrook, J., & Russell, D. W. (2001). Molecular cloning: a laboratory manual. 2001. In: Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

38. Schaid, D. J., Chen, W., & Larson, N. B. (2018). From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nature reviews. Genetics*.

39. Skibola, C. F., Smith, M. T., Kane, E., Roman, E., Rollinson, S., Cartwright, R. A., & Morgan, G. (1999). Polymorphisms in the methylenetetrahydrofolate reductase gene are associated with susceptibility to acute leukemia in adults. *Proceedings of the National Academy of Sciences*, *96*(22), 12810-12815. doi:10.1073/pnas.96.22.12810

40. Soofi, S., Khan, G. N., Sadiq, K., Ariff, S., Habib, A., Kureishy, S., . . . Rizvi, A. (2017). Prevalence and possible factors associated with anaemia, and vitamin B 12 and folate deficiencies in women of reproductive age in Pakistan: analysis of national-level secondary survey data. *BMJ open*, *7*(12), e018007.

41. Stover, P., James, W., Krook, A., & Garza, C. (2018). Emerging Concepts on the Role of Epigenetics in the Relationships between Nutrition and Health. *Journal of internal medicine*.

42. Torre, L. A., Siegel, R. L., Ward, E. M., & Jemal, A. (2015). Global cancer incidence and mortality rates and trends—an update. *Cancer Epidemiology and Prevention Biomarkers*.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

43. Ulrich, C. M., Kampman, E., Bigler, J., Schwartz, S. M., Chen, C., Bostick, R., . . . Potter, J. D. (1999). Colorectal adenomas and the C677T MTHFR polymorphism: evidence for gene-environment interaction? *Cancer Epidemiology and Prevention Biomarkers, 8*(8), 659-668.

44. WMA. (2018, 09-01-2018). WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects Retrieved from https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/

45. You, W., Li, Z., Jing, C., Qian-Wei, X., Yu-Ping, Z., Weng-Guang, L., & Hua-Lei, L. (2013). MTHFR C677T and A1298C polymorphisms were associated with bladder cancer risk and disease progression: a meta-analysis. *DNA Cell Biol, 32*(5), 260-267. doi:10.1089/dna.2012.1931

Table 1. Clinico-pathologic Characteristics of the Breast Cancer Samples

Sr. No.	***Tumour characteristic***	***Value/ numbers***
1.	**Tumour Size**	**<2cm :** 11 (9%)
		2-5cm : 52 (43%)
		>5cm : 58 (48%)
2.	**Tumour Stage**	**T_1:** 03 (03.84%)
		T_2: 26 (33.33%)
		T_3: 43 (55.12%)
		T_4: 06 (07.69%)
3.	**Tumour Grade**	**G1:** 01 (0.94%)
		G2: 48 (45.28%)

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

		G3: 55 (51.88%)
		G4: 2 (1.88%)
4.	**Histopathology**	**IDC:** 88 (79.3%)
		DCIS: 01 (0.9%)
		LCIS: 0
		ILC: 02 (1.80%)
		DCIS + IDC: 15 (13.5%)
		Others: 05 (4.50%)

IDC: Invasive Ductal Carcinoma; DCIS: Ductal Carcinoma Insitu; LCIS: Lobular Carcinoma In situ; ILC: Invasive Lobular Carcinoma

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 2. Distribution of *MTHFR* 677C/T genotypes and allele frequencies (with standard errors) in controls, breast cancer patients, and in the two combined groups.

***MTHFR* 677C/T Polymorphism**	**Controls (n = 62[¶])**	**Breast cancer patients (n = 124)**	**Combined (n = 186)**
Genotypes			
CC	47	90	137
CT	12	28	40
TT	03	06	09
Allele Frequencies			
p[C]	0.85±0.032	0.84±0.033	0.84±0.033
q[T]	0.15±0.032	0.16±0.033	0.16±0.033

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Hardy-Weinberg Equilibrium:			
χ^2	3.0	3.39	6.23*
P-value	NS	NS	Significant

*= Significant; ¶ Result was unavailable for one sample.

Table 3. Distribution of *MTHFR* 1298A/C genotypes and allele frequencies (with standard errors) in controls, breast cancer patients, and in the two combined groups.

***MTHFR* 1298A/C Polymorphism**	**Controls (n = 63)**	**Breast cancer patients (n = 124)**	**Combined (n = 187)**
Genotypes			
AA	21	37	57
AC	33	87	120
CC	4	0	3
Allele Frequencies			
p[A]	0.65±0.043	0.65±0.043	0.65±0.043
q[C]	0.35±0.043	0.35±0.043	0.35±0.043

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Hardy-Weinberg Equilibrium:			
χ^2	4.98*	36.21**	38.95**
P-value	Significant	Highly-significant	Highly-significant

**= Highly Significant

Table 4. Distribution of the *MTHFR* 677C/T polymorphism and genetic models stratified on the basis of young age group (≤35 years) and advanced age-group (>35years)

***MTHFR* 677C/T**	**Patients (n%)**		**OR**	**95% CI**	**p-value**
	≤35 years	**>35 years**			
Genetic Model					
Dominant Model					
TT+CT vs CC	09 vs 16	25 vs 70	0.635	0.249-1.618	0.332
Over dominant model					
CT vs CC+TT	09 vs 16	19 vs 95	**0.356**	**0.137-0.923**	**0.05**
Recessive model					
TT vs CC+CT	0 vs 25	06 vs 89	**1.281***	**1.162-1.412**	**<0.05**
Homozygous codominant					
TT vs CC	00 vs 16	06 vs 70	**1.229***	**1.11-1.359**	**<0.05**
Heterozygous codominant					
CT vs CC	09 vs 16	19 vs 70	2.072	0.793-5.419	0.187

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Allele contrast					
T vs C	09 vs 41	31 vs 159	1.126	0.497-2.550	0.831

***=** Significant

Table 5. Distribution of the *MTHFR* 1298A/C polymorphism and genetic models

***MTHFR* 1298A/C**	**OR (95% CI)**	***p* value**
Genetic model		
Dominant model		
CC+AC vs AA	1.33 (0.694 – 2.556)	0.404
Over dominant model		
AC vs AA+CC	1.754 (0.925 – 3.326)	0.098
Recessive model		
CC vs AA+AC	0.320 (0.259 – 0.397)	0.012*
Co-dominant multiplicative Model		
CC vs AA	0.379 (0.273 – 0.527)	0.027*
Allele contrast		
C vs A	0.988 (0.623 – 1.568)	1.00

*= Significant

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 6. Distribution of combined genotypes of *MTHFR* 677CT and 1298AC polymorphism in control and cases (breast cancer patients) and their risk assessment using OR (95%CI)

Sr. no.	*MTHFR* 677C/T polymorphism	*MTHFR* 1298A/C polymorphism	Breast cancer patients (n=124)	Control (n= 63)	OR (95%CI)	p value
1	CC	AA	22 (17.8%)	20 (31%)	REFERENCE	
2	CC	AC	68 (55.2%)	23 (36.5%)	**2.688 (1.247-5.795)**	**0.016***
3	CC	CC	0	3 (4.77%)	0.3255 (0.0116-3.298)	0.1901
4	CT	AA	11 (8.9%)	0	**20.91 (1.156-378.2)***	**0.0006***
5	CT	AC	17 (13.8%)	10 (15.88%)	1.545 (0.575-4.152)	0.41
6	CT	CC	1	0	2.733 (0.105 – 70.91	0.28
7	TT	AA	4	1	3.636 (0.374-	0.29

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

			(3.25%)	(1.59%)	35.319)		
8	TT	AC	2 (1.6%)	2 (3.18%)	0.909 7.071)	(0.117-	0.93
9	TT	CC	0	0	-		-

*= Significant

Table 7. Semiparametric haplotype analysis-based and frequencies of *MTHFR* 677C/T and 1298A/C haplotypes in breast cancer patients and control

Sr.no.	Haplotypes	No. of cases (%)	No. of control (%)	χ^2/ Fisher's exact Test (p-value)	OR (95%CI)
1	**677C*1298A**	64 (51.2%)	35 (59.3%)	REFERENCE	
2	**677C*1298C**	41 (31.8%)	16 (27.1%)	0.38	1.401 (0.69 – 2.9)
3	**677T*1298A**	17 (13.6%)	3 (5.1%)	**0.03***	3.1 (0.9034-13.95)
4	**677T*1298C**	4 (3.2%)	5 (8.5%)	0.13	0.437 (0.11-1.735)
	Linkage disequilibrium (LD) measurements				
	-0.502				
	0.41				
	Correlation coefficient (r^2)				
	0.0273				
	0.0496				
	χ^2 (*p*-value) for LD				
	3.41				
	2.93				

*Significant

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 8. Bioinformatics analysis for the effects of SNPs on protein function

Amino Acid change	**MutPred2 score**	**Prediction**	**PROVEAN score**	**Prediction (cutoff= -2.5)**	**Polyphen2 score**	**Prediction**
Ala>Val (677C/T)	0.892	deleterious	-3.76	Deleterious	0.998	probably damaging
Glu>Ala (1298A/C)	0.457	benign	-3.088	Deleterious	0.021	Benign

medRxiv preprint doi: https://doi.org/10.1101/2021.03.13.21253512; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure Legends:

Figure 1 (A). 677C/T locus and restriction site of *HinfI*; (B). 1298A/C locus and restriction site of *MboII*

Figure 2 *MTHFR* 677C/T genotyping after digestion with *HinfI* restriction enzyme using 10% polyacrylamide gel stained with ethidium bromide, CC genotype: 198-bp fragment, TT genotype: 175-bp fragment, CT genotype: 198-bp and 175-bp fragments. (Inset) Sanger sequencing validation (reverse sequence).

Figure 3 *MTHFR* 1298A/C genotyping after digestion with *MboII* restriction enzyme using 10% polyacrylamide gel stained with ethidium bromide, AA genotype: 56-bp fragment, CC genotype 87-bp fragment, AC genotype: 56-bp and 87-bp fragments. (Inset) Sanger sequencing validation (forward sequencing). No CC genotype was observed in cases.

Figure 4. Integration of the folate cycle with different pathways for protection against cancers.

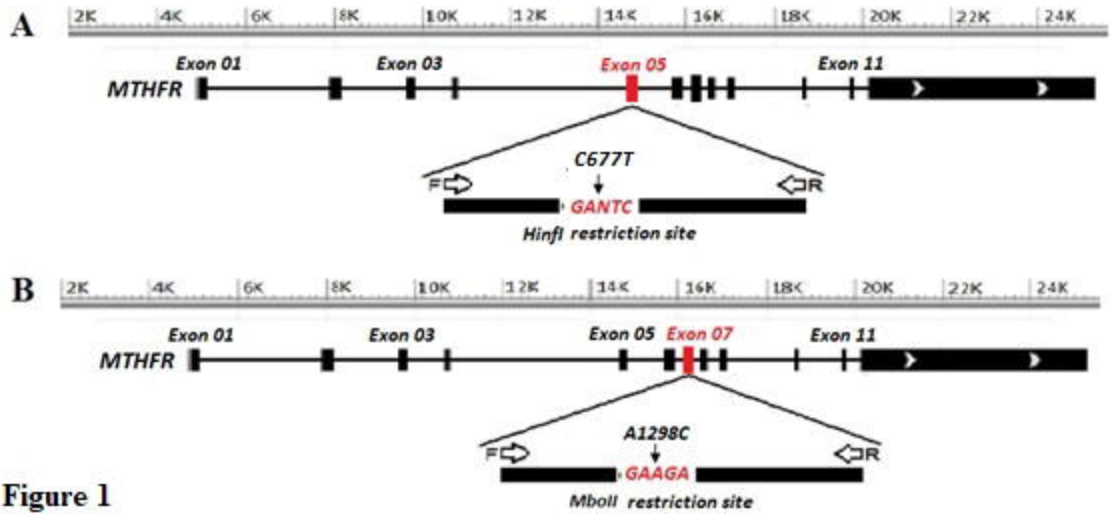
A
2K
4K
6K
8K
10K
12K
14K
16K
18K
20K
22K
24K
Exon 01
Exon 03
Exon 05
Exon 11
MTHFR
C677T
F
R
GANTC
HinfI restriction site
B
2K
4K
6K
8K
10K
12K
14K
16K
18K
20K
22K
24K
Exon 01
Exon 03
Exon 05
Exon 07
Exon 11
MTHFR
A1298C
F
R
GAAGA
MboII restriction site

Figure 1

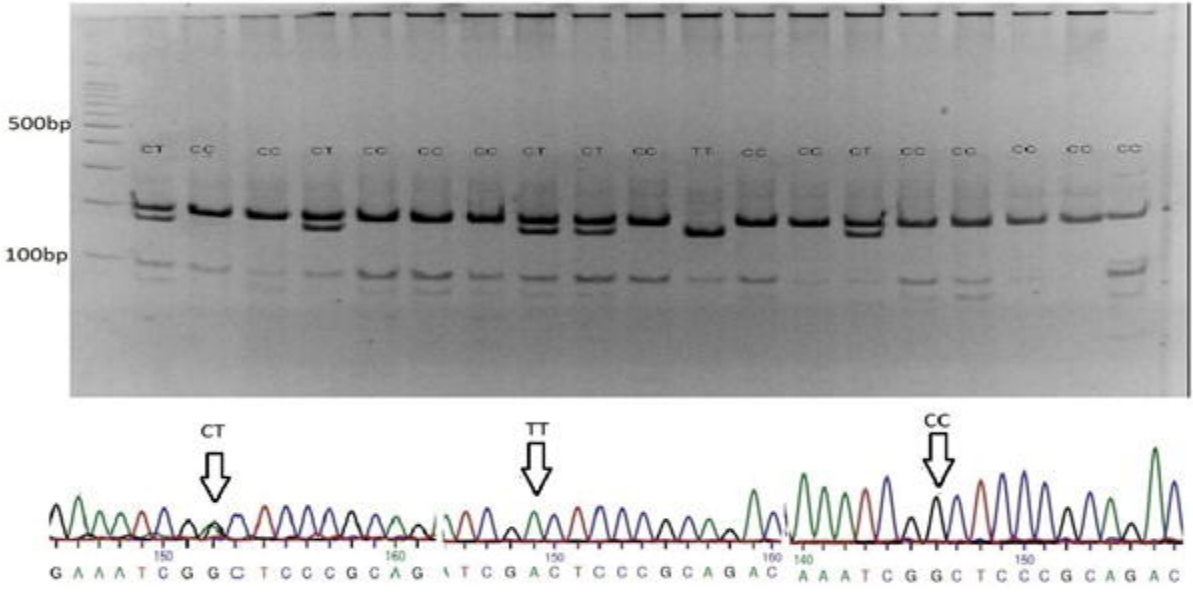
500bp
100bp
CT
CC
CC
CT
CC
CC
CC
CT
CT
CC
TT
CC
CC
CT
CC
CC
CC
CC
CC
CT
TT
CC
150
160
150
160
140
150
G A A A T C G G C T C C C G C A G
T C G A C T C C C G C A G A C
A A A T C G G C T C C C G C A G A C

Figure 2

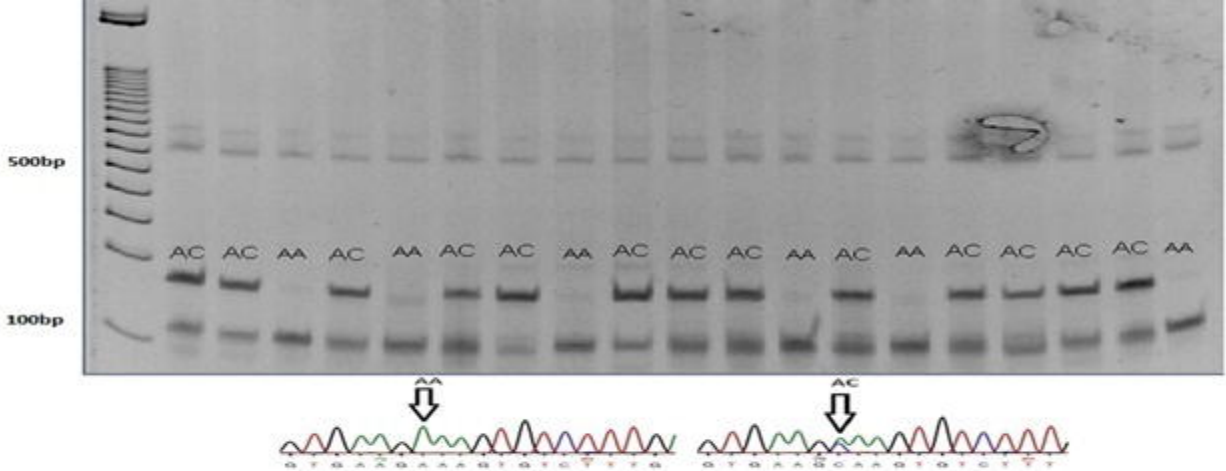
500bp
100bp
AC AC AA AC AA AC AC AA AC AC AC AA AC AA AC AC AC AC AA
AA
AC

Figure 3

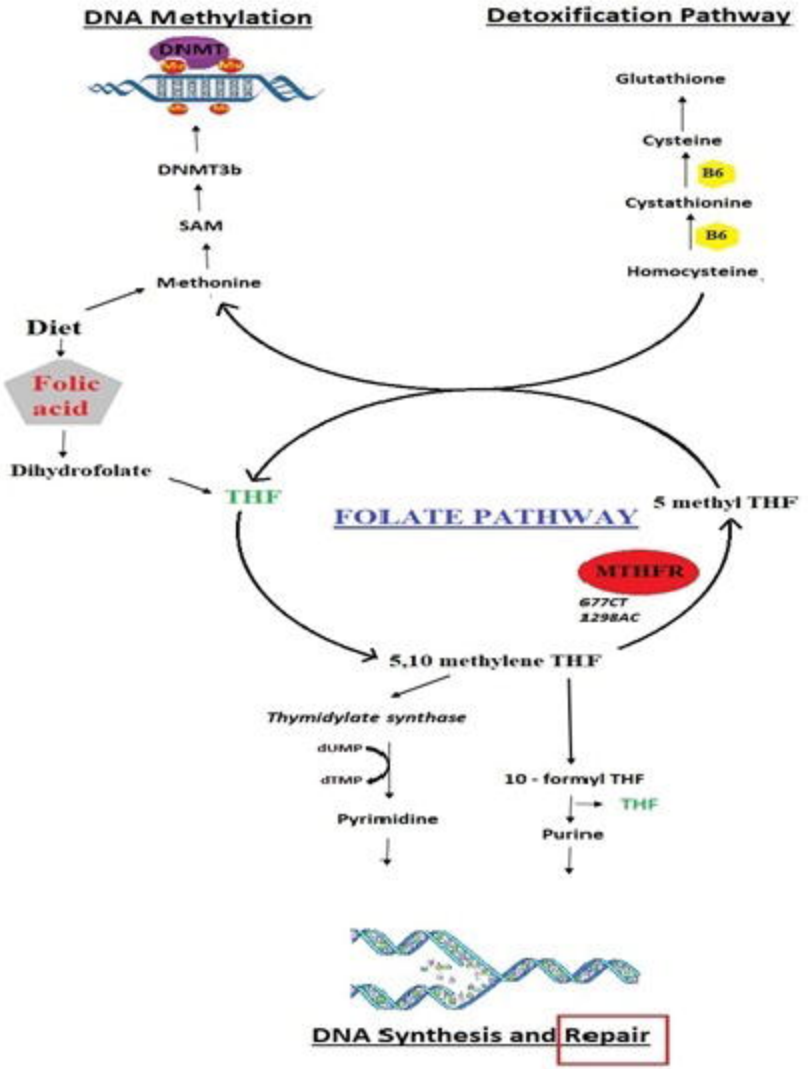
DNA Methylation
Detoxification Pathway
DNMT
Glutathione
Cysteine
B6
Cystathionine
B6
Homocysteine
DNMT3b
SAM
Methonine
Diet
Folic acid
Dihydrofolate
THF
FOLATE PATHWAY
5 methyl THF
MTHFR
677CT
1298AC
5,10 methylene THF
Thymidylate synthase
dUMP
dTMP
Pyrimidine
10 - formyl THF
THF
Purine
DNA Synthesis and Repair

Figure 4