Title: Association of MTHFR Gene Polymorphisms with Breast Cancer Survival

Authors:

Damali N Martin (martinda@mail.nih.gov)
Brenda J Boersma (boersmab@mail.nih.gov)
Tiffany M Howe (howet@mail.nih.gov)
Julie E Goodman (jgoodman@gradientcorp.com)
Leah E Mechanic (mechanil@mail.nih.gov)
Stephen Chanock (chanocks@mail.nih.gov)
Stefan Ambs (ambss@mail.nih.gov)

Version: 2 Date: 3 October 2006

Author's response to reviews: see over
Dear Dr. Deborah Saltman, Editorial Director:

We are submitting our revised manuscript entitled “Association of MTHFR Gene Polymorphisms with Breast Cancer Survival” for publication in *BMC Cancer*. We appreciate the reviewer’s comments. We addressed their concerns point by point and made the appropriate changes to the manuscript. We think these changes have substantially improved the quality of our manuscript. We look forward to the evaluation of our revised manuscript. Thank you for your consideration and time.

Sincerely,

Stefan Ambs, Ph.D., M.P.H.
Principal Investigator
National Cancer Institute
Building 37/Room 3050B
Bethesda, MD 20892-4258
Phone: (301) 496 – 4668
Email: ambss@mail.nih.gov
Response to Reviewer’s comments

Reviewer Dr. Martha Shrubsole:

1. Presumably few eligible women died prior to recruitment if recruitment occurred within 6 months of diagnosis. For those cases that were deceased but otherwise eligible, were they included in the study or current analysis? If not, how many were excluded?

Response: In our study, all patients were identified through surgery lists. Thus, they were scheduled for surgery in the near future when we identified them. We determined preliminary eligibility after reviewing the pathology report, and immediately contacted eligible patients by phone. If they agreed to participate, we usually obtained the consent of the patients within the next few days, which allowed us to collect frozen tissue samples at the time of surgery. The latter was a major focus of our protocol. Because of this design, almost all patients who were found to be eligible for the study survived prior to recruitment. Only one eligible patient died prior to surgery and is not included in the study population. We added this information to the manuscript and modified our description of the study population on page 4.

2. The mean follow-up time is given. With the long recruitment period (10 years) it would be helpful to include some more information about the distribution of follow-up time. Please also give the median follow-up time.

Response: We modified the manuscript accordingly, and report now the median and mean follow-up time for the patients, under Methods, in the statistical analysis section, on page 7.

3. The number of deaths (59) is given in Table 1, but the number of breast cancer deaths, the analysis events according to the methods section, is not. It would be helpful to have the person-years and number of events added to Tables 3-5. Having the sample size available would allow the reader to evaluate the context of the findings. Also, is information on disease-free survival available in this population?

Response: We followed the advice of the reviewer and describe now the number of deaths (n = 5) that were censored in our analysis. This information has been added under Methods, at the bottom of the statistical analysis section on page 7 as follows: “The cause of death was not related to breast cancer for 5 patients.” We also included this information into the footnote of Table 1. The time at risk in person-years and the number of events for all strata has now been added to Tables 3 and 4, and is referred to in the footnote of Table 5. We do not have information on disease-free survival.

4. It is possible that the relationships with ER (-) cancers are due to differences in chemotherapy treatment or stage, both of which are strongly related to survival. The authors stated they evaluated interactions of ER status with chemotherapy. Although the sample size is small to evaluate any interaction, did the authors evaluate an interaction of
MTHFR with TNM stage? Likewise, the authors state that few in their study population used MTX/5-FU, so it seems that they have specific information on chemotherapy modalities. If these data are available, the information should be added to Table 1. Was the null interaction between ER status and specific chemotherapy regimens or was it only evaluated for any chemotherapy?

**Response:** We addressed the reviewer’s concern and examined the interaction between the two MTHFR SNPs and TNM stage. We did not observe an interaction. We did not perform interaction analysis after stratification by specific treatment modalities because of the small sample size in the subgroups. We followed the reviewer’s advice and added the available treatment information to the manuscript. It reads as follows on page 7:

“Most women were of African-American descent, and 132 patients received chemotherapy. Among those, 15 patients received a 5-fluorouracil/methotrexate-based therapy, 71 patients received a doxorubicin-based combination therapy, and the remaining 46 patients received either other agents or the exact treatment modality was not reported.”

**Reviewer Dr. Donghui Li:**

1. The mean follow up time should be changed to median follow up time.

**Response:** We made this change. Please see our response to point #2 of the first reviewer.

2. Did chemotherapy include both neoadjuvant and adjuvant therapy? Was there a difference in the type of therapy each group received?

**Response:** Chemotherapy treatment included both neoadjuvant and adjuvant therapy. This is now indicated in the footnote of Table 1. In our population, 53 patients received neoadjuvant chemotherapy. This information was added to Table 1. For some patients (n = 17), we do not know whether they received neoadjuvant or adjuvant chemotherapy. The treatment modalities of these two therapy groups are similar.

3. Page 8, first line, the C allele lead to an increased survival hazard is easily misunderstood as increased survival. Increased risk of poor outcome or increased risk of dying is probably a better expression.

**Response:** We changed the wording per reviewer’s suggestion and replaced the term “increased survival hazard” with “increased risk of dying”.

4. In the Statistical Analysis section, it was stated that a statistical test for interaction was performed in Stata to determine if the effect of the two MTHFR genotypes on breast cancer survival was modified by other factors. The authors need to disclose more details on this test so others can replicate their study.
Response: The analysis to determine an interaction on breast cancer survival was coded as “\textit{xi: stcox i.Var1*Var2}”. The interaction term was coded with \textit{MTHFR} genotypes as Var1 and other factors (e.g., race, alcohol, and smoking) as Var2. We added this explanation under Methods, statistical analysis section, on page 6: “A statistical test for interaction was performed in Stata to determine if the effect of the two \textit{MTHFR} genotypes on breast cancer survival is modified by other factors. In the analysis, an interaction on breast cancer survival was coded as “\textit{xi: stcox i.Var1*Var2}”. The interaction term was coded with \textit{MTHFR} genotypes as Var1 and other factors (e.g., race, alcohol, and smoking) as Var2. Var1 and Var2 were coded as binary variables. For this test, the result was coded as zero (or referent) for the common \textit{MTHFR} genotype (C/C for SNP677 and A/A for SNP1298) and one for the variant genotype (C/T or T/T for SNP677; A/C or C/C for SNP1298).”

5. We were told “The genotype frequencies for both SNPs were in Hardy-Weinberg equilibrium for a matched control population (n=317).” What about the distribution in cases?

Response: We did not test for deviations of the observed genotype frequencies from the Hardy-Weinberg equilibrium (HWE) among the cases. We evaluated only an existing, matched control sample (n = 317) as part of the QC process for the two \textit{MTHFR} SNP assays. The genotype distribution among cases could possibly deviate from HWE because of selection, however, we do not see such a deviation in an analysis that we performed in response to the reviewer’s question. We added this information to the manuscript, on page, and wrote as follows: “The observed genotype frequencies for both SNPs were in Hardy-Weinberg equilibrium in our case population, and for a matched control population (n = 317).”

6. Line 7 from the bottom of page 7, the variant allele at codon 677 (A/C or C/C) perhaps should be codon 677 (T/T or T/C).

Response: We made the change.

7. Was there an interaction between ethnicity and estrogen receptor status?

Response: We did not find a statistically significant interaction between estrogen receptor status and race on breast cancer survival (p = 0.13).

Reviewer Dr. Ji-Yeob Choi:

1. Study Population: comparably small eligible subjects (n=248) has been collected in several hospitals (n=5) for 10 years. Although authors excluded patients having some criteria, the participation rate still appeared low. How many subjects were listed on to participate to obtain the 248 women and what percentage was excluded? This should be provided in the Study Population section.
Response: We revised the section, in which we describe the study population under Methods, on pages 4/5. Our protocol focused on the recruitment of breast cancer cases that had surgery and could provide frozen tissue samples for research. In addition, our recruitment was aimed to maximize the participation of African-American patients. Although we recruited patients at five hospitals, only one hospital, the University of Maryland Medical Center, scheduled significant numbers of surgeries during that time. The participation rate among patients, whom we contacted, was actually quite good. Of the patients that were eligible under the original protocol and could be contacted by phone, 83% agreed to participate in the study. This is information is included in the revised description of the study population.

3. Genotyping: genomic DNA for genotyping was from several different sources including 28 tumors and non-tumor tissues. Tumor DNA might be different from non-tumor tissues or buffy coat DNA, thus need to discuss that.

Response: We used 28 fresh-frozen tumor samples from those patients where blood samples or non-tumor tissue was not available. There is the possibility that the genotype in a tumor sample is not identical with the germline genotype due to either a loss of heterozygosity or chromosome amplification. In a pure tumor sample, those changes could cause misclassification of the heterozygote genotype and could lead to an overestimate of either the homozygote wild-type or the homozygote variant genotype. However, this possibility is remote using bulk breast tumor samples because of the abundance of non-tumor cells in breast tumor tissue. However, in response to the concern of the reviewer, we microdissected 21 of the 28 tumors that were used for genotyping and obtained normal surrounding tissue from these samples. We determined the germline MTHFR C677T genotype in these samples by direct sequencing and found a 100% concordance between the genotype obtained from the bulk tumor and the genotype of microdissected normal tissue from the same patients. Thus, we strongly feel that the use of tumor DNA did not compromise the quality of our genotyping data. This information has been included in the revised manuscript on page 6. We wrote as follows: “In addition, we analyzed the MTHFR C677T genotype of microdissected normal tissue from 21 of those 28 bulk tumor samples that were genotyped at the NCI GCF. The MTHFR genotype of these samples was determined by direct sequencing. We had 100% concordance between the C677T genotypes of the bulk tumor samples and the non-tumor tissue from these samples.”

4. What kinds of method did you used to estimate haplotypes reconstruction and frequency estimate?

Response: We did not use haplotype reconstruction and frequency estimates. This was not required for our analysis. Instead, we categorized the genotypes into high and low MTHFR activity diplotypes based on previously published reports of the effect of the variant alleles on MTHFR activity level, and the information shown in Table 2b. Patients were classified as carriers of a high activity diplotype (n = 143) if they either had the CC (677) and AA (1298) or the CC (677) and AC (1298) genotype combinations (see Table
2b). The CC(677)AC(1298) diplotype has 80—90% of the enzyme activity encoded by the CC(677)AA(1298) diplotype. All other genotype combinations were considered low activity diplootypes (n = 100) with a predicted MTHFR activity of 30—70% of the CC(677)AA(1298) diplotype. This categorization is described on pages 9/10 of the manuscript. Categorization can be easily obtained with the existing genotype data and does not require haplotype reconstruction. We also believe that our categorization is most informative because it addresses the biological effect of the two functional MTHFR genotypes.

5. Each table should present how many events (death) there were in each stratum because small sample size might result in non-significant association. For example, the associations of MTHFR on breast cancer survival were statistically significant in only ER negative, which may be related to poor prognosis (more events) in ER negative group rather than ER positive group.

Response: As requested by the reviewer, we added the number of events to the strata in Tables 3, 4 and referred to this information in the footnote of Table 5.

6. In table 2, the frequencies of each race looked different significantly, and should add p-value for difference (such as by chi-square test). Authors showed estimated haplotype from genotype data, but the population frequencies of haplotypes, as well as the diplotype configuration of each subject, are estimated from a set of genotypes of the subjects in a sample from the population, therefore, the haplotype estimated from genotype data should be estimated after stratified by races. For estimation of haplotype, linkage disequilibrium between SNPs should be addressed.

Response: We added the p-values from the chi-square test to Table 2. As explained in our response to point #5, our grouping of patients by MTHFR activity does not require haplotype reconstruction. All information that is needed is shown in Table 2b. High MTHFR activity is represented by only two diplotypes and we have full ascertainment of this information based on the genotype information for each patient (see Table 2b). All other possible diplotypes will yield a decreased MTHFR activity.

7. If there are significant differences of allele frequencies between races, the analysis should stratify the data by races rather than add interaction term in the model, especially in case that the relationship between variables was in opposite direction (Szklo and Nieto, Epidemiology Beyond the Basics, Chapter 6-7). How could authors interpret the results of difference of allele frequencies with opposite direction of outcome?

Response: We show the results of our survival analysis stratified by race in Tables 3-5, and the results of the interaction analysis are consistent with our findings in the race strata. Any further sub-stratification by race did not yield statistically significant associations. We discussed this limitation of our study on page 15. We are aware that our results need further verification in other studies but we believe that our findings, although they appear to be paradox at first, are consistent with observations in other studies. This is discussed in details in our manuscript.
8. Discussion: As authors described the role of MTHFR in the last paragraph on page 10, a lot of factor could influence the outcome of breast cancer related to MTHFR genetic polymorphisms. It should be very hard to interpret the results of opposite direction on breast cancer survival by C677T and A1298C, that both were reported to decrease enzyme activity, in mixed races with different allele frequencies as well as different prognosis profiles.

Response: Indeed, it is difficult to explain why two genotypes that affect an activity in a similar manner have opposite effects. We are not the first who observed this paradox, and an opposite effect of the two genotypes on breast cancer risk has been reported (Cancer Res. 2005, 65, 1606-1614). There are also several studies that observed differing effects of the two polymorphisms on 5-fluorouracil and methotrexate efficacy and toxicity (Br. J. Cancer 2004, 90, 526-534; Pharmacogenetics 2002, 12, 183-190; Ann. Rheum. Dis. 2004, 63, 1227-1231). It is possible that the two SNPs affect MTHFR activity by very different mechanisms that are not revealed in an in vitro enzyme activity assay. There is evidence from one study that the two SNPs may not only affect enzyme activity but also stability (PNAS 2001, 98, 14853-14858). As such, the biological effect of the two SNPs may differ in a more complex biological system and could depend on cofactor availability or other intrinsic and lifestyle-related factors. We addressed this point in the discussion and wrote as follows on page 14, second paragraph: “There is evidence that the two SNPs may have different effects on the biochemical properties of MTHFR [9]. The A1298C leads to a substitution of a glutamine by alanine in the regulatory domain of the enzyme but does not cause thermolability of the enzyme or enzyme instability in the presence of low folate, like C677T [7,9]. Several reports showed that the two SNPs have different effects on the efficacy and toxicity of methotrexate and 5-fluorouracil. The 1298C variant allele was found to be associated with fewer adverse effects and higher efficacy and the 677T variant allele with an increased systemic toxicity in rheumatoid arthritis and cancer patients [20,21,23,38].”