Author's response to reviews

Title: Cancer-testis gene expression is associated with the methylenetetrahydrofolate reductase 677 C>T polymorphism in non-small cell lung carcinoma

Authors:

Kerem M Senses (kerem@bilkent.edu.tr)
Mithat Gonen (gonenm@MSKCC.ORG)
Ahmet R Barutcu (rasim.barutcu@umassmed.edu)
Zeynep Kalaylioglu (kzeynep@metu.edu.tr)
Murat Isbilen (murat.isbilen@bilkent.edu.tr)
Yao T Chen (ytchen@med.cornell.edu)
Nasser K Altorki (nkaltork@med.cornell.edu)
Ali O Gure (agure@bilkent.edu.tr)

Version: 2 Date: 9 August 2013

Author's response to reviews: see over
We thank the reviewers for their comments. Our responses are below. We have altered the title of the manuscript as the newer analyses requested by the reviewers show an association with the MTHFR 677 C allele, and not just the genotype with CT gene expression. We have also added a geneticist (Dr. Ozlen Konu) who is knowledgeable in various genetic association analyses.

Responses to comments of the 1st reviewer (Dr. Mokarram):

Comment: In Table 1 There is no explanation why the author divide age based on age 60.
Response: The division is somewhat arbitrary but was chosen because the average age at which patients are diagnosed with lung cancer generally falls between 60 and 69. However, in this revised version of the manuscript, tables 3 and 4, where results of MVA are presented, include age as a continuous variable.

Comment: In Table 2, the author should explain different genotype distribution among CT+ and CT-. For example, how many samples have CC(MTHFR)/AA(MTHFR)/ AA (MTR)/......... genotype combination.
Response: We realize the importance of assessing the status of SNPs analyzed in this study in combinations. This is important to determine if some alleles have additive/cumulative effects on CT gene expression and for model generation. However, we don’t feel it is appropriate to report statistics of combinations simply because the sample size of this study is limited. We, therefore, removed the sentence referring to the diplotype associations with CT gene expression and discuss the issue about sample size in the discussion section in the revised manuscript.

Comment: In table 3, There is no information about “OR” and how ORs were adjusted.
Response: The original table was a logistic regression analysis of CT gene expression, adjusted by MTHFR 677, MTHFR 1298, tumor stage and histology. We realize this might have been incorrect, as MTHFR 1298 genotype is not associated with CT gene expression by univariate analysis. We have repeated the logistic regression analysis using age, sex, MTHFR 677 status, tumor stage, and histology. The ORs reported in the additional table (table 4) are now adjusted by the same nongenetic parameters, as indicated (page 19).

Responses to comments of the 2nd reviewer (Dr. Prokunina-Olsson):

Comment: Please explain definition of low/high expression of CTs.
Response: The basis by which we categorized samples based on CT expression levels has been further elaborated in the methods section (page 7), and CT expression levels for patients are given in supplementary table 1.

Comment: Was the gene expression used as a single factor for selection of these patients?
Response: Yes. We now state this in the methods section (page 6).

Comment: In a supplementary table please provide a list of all 50 patients with their metrics (age, sex, race, diagnosis) and all 9 CT genes with indications on which genes were expressed at high level in each patient. Ethnicity of these patients should be indicated and possibly taken into consideration. Is there known difference in allelic distribution of these genes in the populations used in the study?
Response: The information is now provided in supplementary tables 1 and 2. Race data was available for only 29 patients of whom 25 were non-Hispanic white, 3 were of mixed race and 1 was a non-Hispanic black individual. Since race is unknown for 21 patients and since the overwhelming majority is white we have not included race in our analysis. This information is now given in page 9.

Comment: The reason for using NCI-60 panel is unclear. This is a highly heterogeneous population, even between cell lines within the same tissue origin. NCI-60 is really not a good place to look for genotype-phenotype correlations. Each cell line may have or not have expression for totally different cell-specific reasons unrelated to genotypes of these markers. So, the part on NCI-60 can be removed completely. However, the part on the AML set has to be extended by a better presentation of the results, including the PCA and K-clustering. Without any data presented, this part doesn't bring much to the table.
Response: The analysis with the NCI-60 data has been removed in this version. We now provide details as to how PCA and K-clustering was performed. We have also included a figure to help clarify the cluster analysis (supplementary figure 1).

Comment: All variants should be accompanied with corresponding rs numbers, especially in the table 2. In the text, the linkage disequilibrium between the 677 and the 1298 markers should be presented as D’ and r2 values determined in your samples and not by X2 and p-values.
Response: We have added the reference SNP ID numbers to all tables. LD values are now given as D’ and r2 (page 8).

Comment: Table 3 has to be explained better. The magnitude of effect sizes accompanied by the hardly significant p-values doesn't make much of a sense.
Response: The original table was a logistic regression analysis of CT gene expression, adjusted by MTHFR 677 and MTHFR 1298 genotypes, tumor stage and histology. We realize this might have been incorrect, as MTHFR 1298 genotype is not associated with CT gene expression by univariate analysis. We have repeated the logistic regression analysis which now includes age, gender and the variables significant by univariate analysis in tables 1 and 2 (histology and t stage) (page 19). We feel the ORs obtained with the currently reported analysis match corresponding p values. The upper limits of some of the confidence limits are very high due to the small number of patients in the corresponding categories but we do not see this as a shortcoming as the more important components of this analysis are the lower confidence limits, which are appropriately estimated, and provide the significance information.

Comment: The very small data set (n=50) should be discussed as a strong study limitation. It would be informative to have a power calculation on the effect size expected to be detected in such a data set. Some markers have very large differences in genotype distributions but this is clearly not enough in this sample set. The more correct way would be to present the per-allele ORs and p-values adjusted for relevant variables (age, sex, race and stage).
Response: The discussion section now includes an acknowledgement of the limitations due to sample size and is supported by power calculations. We performed additional per-allele based multivariate logistic regression analysis, the results which are presented in table 4 where all ORs are adjusted by age, sex, histology and t stage (page 19).

Responses to comments of the Editor:

Comment: The paper has an interesting observation on the association between the MTHFR variant and gene expression in lung cancer and this part should be highlighted.
Response: We have altered the title of the paper to emphasize this fact and also emphasize this in the introduction. We separated the results and discussion sections in this version and the first paragraph of the discussion is on the association between MTHFR 677 and CT gene expression.

Comment: The NCI-60 part could be removed altogether as it's not a good model to look for genotypic effects. The in silico analysis in AML samples is not a validation and can be called "in silico expression analysis" instead.
Response: We have removed the NCI-60 related analyses from this version of the manuscript. The section on in silico analyses of AML datasets is now titled "in silico association analysis" as this is a study of the correlation of CT gene expression and MTHFR genotype.

Comment: A part of the text on HWE and allele frequencies should be included as a part of Table 2 and not presented in the text.
Response: The text description has been removed. However, we feel the inclusion of the HWE data into table 2 convolutes the data. We, instead, present the HWE analysis independently as a supplementary table (suppl. table 3).

Comment: Rs numbers of variants should be used, especially in the tables, in addition to other labeling type.
Response: All relevant tables now contain rs numbers.

Comment: Table 3 presents the results of multivariate logistic regression, not the Cox regression model.
Response: The labeling has been corrected.

Comment: ORs and confidence intervals could be presented with 2 decimal values. It should be explained how the ORs in Table 3 and p-values in Table 2 were calculated and adjusted. Based on table 1 results, they should be adjusted for histology and T stage.
Response: P values in table 2 were calculated by chi-square or Fisher’s exact test as indicated. Table 3 reports ORs adjusted for all variables shown in the table. We have included a 4th table with this version (per-allele associations with CT gene expression) and the ORs presented here are adjusted by age, sex, histology and t stage, as indicated.

Comments: Please use page numbering. The paper would benefit from a proofreading. The paper analyzed 5 markers, not 5 alleles. The sentence in background section (Introduction?) reads "Most of the CT genes consist of families of closely related members". Does that mean "belong to families of close related members"? There are some other places like this that require attention.
Responses: Pages are now numbered and the manuscript has been proofread by an additional author. The sentence regarding 5 markers has been corrected. We have rephrased sentences that were not clear in the first version.

Comment: Please discuss limitations of your study in the discussion.
Response: We now discuss limitations that are primarily due to the limited sample size of the study. Results from power analysis are used to emphasize this fact. We additionally mention the fact that the data is derived from only a single cohort of patients with lung cancer. Limitations of the in silico data analysis and future perspectives on analyses of other microarray datasets, together with predicting the additive effect or risk alleles are also discussed.