Reviewer’s report

Title: Cancer-testis gene expression correlates with the methylenetetrahydrofolate reductase 677CC genotype in non-small cell lung carcinoma

Version: 1 Date: 9 June 2013

Reviewer: Ludmila Prokunina-Olsson

Reviewer’s report:

Major compulsory Revisions:

1. Please explain definition of low/high expression of CTs. Was the gene expression used as a single factor for selection of these patients? 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?

2. 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.

3. 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.

4. 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.

5. 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).

Level of interest: An article whose findings are important to those with closely
related research interests

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

I declare that I have no competing interests