Reviewer's report

Title: Expression of DNA repair and replication genes in Non-Small Cell Lung Cancer: a role for Thymidylate Synthetase.

Version: 1 Date: 1 May 2012

Reviewer: Giannis Mountzios

Reviewer's report:

This is a very well designed, conducted and presented project of translational research regarding the use of mRNA expression of four specific biomarkers used as prognostic and predictive (for treatment response) tools in NSCLC. The study attempts to address the important gap that exists today in standardization of mRNA expression levels using a rational design and is therefore clinically relevant and interesting. The scientific hypothesis is robust, design is accurate, execution is sound and data analysis and interpretation are adequate and informative. I would like to emphasize on the exhaustive quality control that seems to have been performed by the researchers in their effort to optimize the selection of standard biomarker levels and their RQ values. There are, however, some minor concerns:

Major Compulsory Revisions: NONE

Minor Essential and discretionary Revisions:

ABSTRACT: Please provide explanation for acronyms in their first appearance in both the abstract and the manuscript text

INTRODUCTION: In the last sentence, the expression: "...in order to avoid changes from field cancerization" may be somewhat confusing for the inexperienced reader. I would suggest change to "...in order to mitigate the effects of the field cancerization phenomenon"

PATIENTS, MATERIALS AND METHODS:

Page 5, second paragraph, please change "quitted" to "quit"

Page 7: It is not clear to me why subgroup B was not evaluable for overall survival and why such an analysis was not applicable (as the authors state in Table 4). Please explain

RESULTS:

Page 8: Do the authors have any explanation on why the expression of all genes was higher in stage IIIA? Could this be a statistical artifact due to the multiplicity of analysis and excessive clustering of the study population?

Page 8: The authors state clearly in the last paragraph that they opted to assess
tumor B, E, R, T mRNA expression in reference to available paired tumor-normal samples because the commercially available RNA reference was excessively high for all genes. Given the fact that paired tumor-normal lung samples were available for ONLY 12.5% of the study population (35 samples), it may be possible that the limited number of paired samples used for the identification of reference values deprives the study from originally planned statistical power and renders some of the comparisons non applicable, especially in cases of multiple case clustering. This should be addressed in the discussion session as a potential drawback in the study.

Page 9, first line, perhaps the authors meant: "consistent" instead of "constant"

Page 9, the beginning of the last paragraph ("As described....appear as a handicap...real life scenario...") is rather lengthy and aberrant from strict scientific data presentation, as it should be in a Results session of a manuscript. Please edit/shorten.

Page 10, last paragraph: Although the authors have explained that they opted to group together both very high and very low marker levels (namely: "deviating") because the proportion of low values was very small to be studied separately, this strategy provokes some scientific drawbacks: For example, when the authors state that "patients with tumors expressing deviating ERCC1 having received platinum without taxanes performed better in comparison to patients that received all other combinations...", this is biologically justified for tumors with low ERCC1 but not justified for tumors with high ERCC1 that are known to be resistant to platinum-based chemotherapy (References 5-8). In other words, grouping together tumors with both excessively high and excessively low mRNA levels may obscure the analysis and does not help to clarify the biological hypothesis made by the authors.

In the same paragraph (and the first of page 11), the multiplicity and complexity of analysis leads to excessive clustering (patient groups with N=5) and renders the results difficult to interpret and biologically questionable. I would suggest omitting this paragraph except from the results regarding normal ERCC1 and treatment with taxanes/platinum that are the only statistically significant and biologically robust ones. I would also recommend the use of the term "abnormal" instead from "deviating" that clearly shows the difference from normal levels.

DISCUSSION: Please see previous comment for the limited number of paired tumor-normal lung samples.

Page 12, last paragraph: The sentence: "Although lower...22% of these tumors" is rather incomprehensible. Please rephrase.

Page 13, last paragraph: The finding that patients with adenocarcinoma histology had a prolonged survival than all other types is in contradiction with literature, where squamous cell carcinoma bears the better prognosis. is there a possible explanation for that?
Again, since deviating ERCC1 levels contain mostly high ERCC1 levels (and very few low ones), the finding that deviating ERCC1 levels correlated with a favorable outcome after platinum-based chemotherapy is not biologically justified (references 5-8).

Page 14, last paragraph: Please change "...may be avoided receiving taxanes" with "...may be spared receiving taxanes".

Overall, I recommend acceptance of the manuscript pending the minor revisions suggested.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests