Reviewer's report

Title: Gene expression profiles of lung adenocarcinoma correlate with histopathological grading and survival but not with EGF-R status: a microarray study

Version: 1 Date: 10 July 2009

Reviewer: Stearman Robert

Reviewer's report:

Please number your comments and divide them into
- Major Compulsory Revisions

The authors have presented a small study using gene expression microarrays to examine the relationship of expression to the associated clinical information and EGFR status for their patients. In addition, they tested whether a published gene expression signature (Potti et al.) that is correlated with patient survival would correctly sort their independently derived gene expression dataset.

1. The authors have not clearly delineated a research question in this manuscript. On the one hand, they have shown some success with their independent validation of the Potti et al. (2006) signature. If this is the focus of the work, why were not additional expression signatures tested and reported, especially given the rather limited success this approach has had in a large multi-site study (see Shedden et al Nat Med 14, 822 (2008)). Based on Shedden et al., the few expression signatures that had some utility were significantly augmented by incorporating clinical parameters, yet this manuscript only produce some correlation with tumor grade, not any of the other typical parameters (age, stage, sex etc.) which is not the typical findings in this type of study. On the other hand, the authors may have a potentially interesting question to follow up on in the cases of tumor...
grade 2 samples clustering with the grade 3 samples. This would require additional samples for refinement and validation of a gene classifier.

2. In reading the Potti et al. (2006) paper it is not clear that they have implemented their gene expression signature correctly. In this paper, they have “lumped” all the genes found in the different metagene signatures of Potti et al. It is not clear that is how Potti et al. intended their gene signature to be used (individually, in series, or lumped together) because it is not clear in the NEJM published paper.

3. The EGFR status is an interesting addition though they were unable to show any correlation to the gene expression pattern. Pathway analysis (e.g., Ingenuity) could be a great help here as well. They did not demonstrate high copy number/amplification in any of their patients which is surprising given that EGFR copy number increases are reported in 30-50% of adenocarcinoma patients with ~10% amplifications (see Hirsch et al. cited in the paper ref 36, as well as Hirsch et al J Clin Oncol 26, 3351 (2008)). There was no indication of a set of positive and negative controls for this experimental approach so it leaves the possibility of a technical problem.

4. Overall, the approaches used to analyze the microarray data are rudimentary and could be vastly improved. For instance, there are freeware tools available (like BRB-Arraytools from NIH/NCI) which allow categorical (stage) and continuous data (survival time) to be included in the analysis. I would suggest examining the microarray data in a more thorough manner than presented in this manuscript.

5. The manuscript concludes with the usefulness of publicly available datasets and expression signatures. There is no indication that the authors have taken any
steps
toward depositing their data (data and .cel files) at GEO or another public repository.

Acceptance of this paper should be dependent on making the dataset publicly and easily available.

- Minor Essential Revisions
1. There is confusion in the Table 1 and Figure 1 and 2. Table 1 refers to “T” of TNM classification as “extent of tumor.” My understanding “T” stands for Tumor Stage. If extent is really their meaning, than this is not the proper use of TNM classification. In addition, Table 1 has an entry of “2” for one of the “Node” values not described in the legend (typo?). Figure 1 and 2 legend describe the labeling of EGFR expression in the figures which doesn’t agree with the samples as described in the text. (Text says 8 samples negative; 6 samples complete; therefore 14 samples are partial).

2. At least by the end of the reference list (high 30’s) there is inconsistencies in the topic of the paper cited and the paper listed in the references. Reference 40 is in the text and there is no reference 40 listed.

3. The “Cluster” shown in Figures 2 and 3 should be consistently labeled as Cluster 1, 2, 3….Figure 2 should have the Clusters indicated. The blue vertical lines dividing the heat map in Figure 2 do not correspond to anything and should be eliminated or fixed.

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

Quality of written English: Needs some language corrections before being published

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

Declaration of competing interests:
"I declare that I have no competing interests."