Reviewer’s report

Title: Integrated mutation, copy number and expression profiling in resectable non-small cell lung cancer

Version: 1 Date: 16 October 2010

Reviewer: Andrea Staratschek-Jox

Reviewer’s report:

Newnham et al., Integrated mutation, copy number and expression profiling in resectable non-small cell lung cancer

The authors analysed a series of frozen 69 lung cancer tissues (AC, SCC, LCC) by using gene expression profiling as well as BAC array based genomic profiling. Moreover they investigated the p53 mutation status for all of the samples as well as the K-Ras and EGFR mutation status for AC and LCC. The genomic profiles of AC were first compared to the SCC profiles and revealed remarkable differences with regard to gains and losses in particular genomic regions. Likewise the K-Ras mutant tumors differed from the nonmutant tumors. The number of tumors with EFGR mutation was too small to allow for a representative analysis. Genomic alterations could not be correlated with tumor recurrence or survival. When analysing the gene expression profiles of TP53 positive tumors and wild type tumors differential expressed genes were identified that were most probably also related to the different underlying histological subtypes. Likewise differentially expressed genes were identified in the comparison of K-Ras mutated tumors to wild type. Finally a feature list was established to discriminate metastatic and non-metastatic tumors based on hierarchical clustering. This result could be validated using an independent data set. Last correlations were seen of transcriptional date to genomic profiling data.

Overall the study is well performed and the aim of the study is interesting. However I have some suggestions which might improve the manuscript as follows:

Major

1. The hierarchical clustering approach is probably not the best approach to predict tumor survival in a patient cohort based on a feature list. Rather the author should use a machine learning algorithm to build up a real classifier based on their own data set and to test this classifier by using the external data set. Since the authors intend to identify a prognostic signatures this needs to be done.

2. Are the expression data submitted to the Gene expression omnibus database? Submission is needed prior publication. Please provide the GSE number and www link.

Minor
1. In figure 2 and 3 the annotation of the genome should be more precise. The numbers of the chromosomes should be given either at the region of the centromer (dashed line) or at the beginning. It is hard to identify each of the chromosomes by only using the labels at the left side. Is there no data for the X chromosome?

2. Legends for tables and figures are missing. Please supply.

3. Add fold changes and p values to Table 3 and 4

4. Add p values to table 5

5. Page 6 last line: Figure 1 is cited. Most probably Figure 2 is meant.

6. Tables and figures should be rearranged so that they can be printed upright.

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

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests.