Reviewer's report

Title: The role of the REG4 gene and its encoding product in ovarian epithelial carcinoma

Version: 2 Date: 20 January 2015

Reviewer: Stefan Enroth

Reviewer's report:

Chen and colleagues present a well-written manuscript addressing the role of the REG-4 gene expression and protein abundance in ovarian cancer. The authors study this in a series of cell lines representing different stages of the disease and in tumor material from patients. Higher expression and protein abundance levels are found in some tumors and cell lines compared to normal tissue and higher expression is also found to be associated with poorer survival rate in patient material. The claims made are justified by the data presented.

Major Compulsory Revisions

1. Line 269-271 & Table 2: This is the first mentioning of how many samples that were collected from the patients and the grouping of them based on tumor subtype. Table 2 lists some of the properties of the samples, but the baseline characteristics are missing. How many of the samples were fresh-frozen and how many where fixed in paraffin? Where there any difference in proportion of the tumor types in either of these categories? Where there any differences in age distribution of the samples in the different tumor categories? Why was age = 56 chosen as a divider in Table 2? What does the P-value in the last column in Table 2 denote? Which test was used? How was the classification into the differentiation classes done?

Minor Essential Revisions

2. Line 114-115: The cell lines where kept in 3 different media with the two cell line with the lowest expression (Figure1A: ES-2 and SKOV3) where kept in the same media (McCoy) different from all others. The media used all have slightly different composition, for example in glucose concentration. Could the different media have an effect on the expression patterns seen?

3. Line 166: Is there anything known about the degradation of the proteins analyzed here over time? The samples collected spans 8 years and this could be an influencing factor. This could be checked by stratifying on tumor type and then performing e.g. regression analysis on the abundance levels with time (e.g. in years) since sample collection date.

4. Line 168: Serum levels of REG-4 have been showed to be dependent on the subject’s age in control samples (PMID: 22866070). Since you have a big age-span in your samples, you should check (same procedure as above with
sample date) if there is an influence on your abundance levels stemming from age alone.

5. Line 238: The description of the multivariate analysis is scarce. How many covariates were included? If more than one, the p-value cut-off should be adjusted for multiple hypothesis testing.

6. Line 263: Please refrain from using statements like “there was no correlation” without providing justification. E.g. calculate R^2 for the datasets using the proper method. Note for instance that the Pearson’s statistics is sensitive to outliers and not always appropriate.

7. Line 271: What does “expression was statistically higher” refer to? Please provide the proper statistics for this statement.

8. Line 276: Please refrain from stating that there are correlations without providing the statistics on it (including p-values).

9. Line 280: Which criteria were used for stratifying the REG-4 expression?

10. Line 291-292. These two sentences don’t make any sense. Please rephrase.

11. Line 479: The legend specifies “positive correlation”, although the ordering of the bars in the figure (from “Well” to “Poorly”) suggests a negative correlation in the sense that poor is low and good is high. Please also provide statistics and measures on the correlation.


13. Figures: It’s hard to read the labels inside the panels. The small text in Figure 2A is impossible to make out. Please use larger fonts when the text is essential and remove otherwise.

14. Throughout: Please write out the actual p-values instead of P < 0.05. Since many tests are being made it is of interest to the reader to know if the statistics are clear or borderline.

Discretionary Revisions

15. Line 33-34: last part of the sentence (starting with “subjected to phenotypes’ measurement”) doesn’t make sense to me and should be rewritten.

16. Line 78: REG-4 have also been shown to be up-regulated in HepG2 after 1h stimulation of TGF-B (PMID:24771338).

17. Line 81: There is a “:” right after “[8]”, typo?

18. Line 206: Consecutive. Where these samples collected using consecutive sampling? None was excluded?

19. Line 245: Space missing in “50nM/ml(Figure 1)"
20. Line 246: Verb missing. Perhaps “SKOV3 cells where chosen to be transfected”?
21. Line 465-466: There seem to have been a change in font size?

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

**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