Reviewer's report

Title: Differential expression of THOC1 and ALY mRNP biogenesis/export factors in human cancers

Version: 2 Date: 19 August 2010

Reviewer: Melissa J. Moore

Reviewer's report:

Major Issues:

(1) Even though the authors now cite two papers about a gene expression repository and the human protein atlas, they did not make an effort to incorporate these in their analysis or in discussion. I don’t think the current manuscript will be of value to the scientific community without taking these previous results into serious consideration. Here are some specific questions/suggestions:

- How do the observed mRNA level changes in THOC1, ALY differ from what was previously observed in other cancers, conditions, etc?

- It is not clear to me what is meant by the sentence: “However, in tumor tissues the samples used are too small to establish a reliable comparison.” Do the authors mean that sample size is too small? Or are they referring to the image size depicted at http://www.proteinatlas.org/?

- Are there other proteins that are involved in mRNP biogenesis that exhibit similar patterns of IHC based on the data from the Human Protein Atlas?

- Similarly, based on data from the Human Protein Atlas, are there any other conditions/tissue types/cancer cell lines where THOC1, ALY expression was assessed using IHC that could help contextualize or generalize the results from the current study?

- As I mentioned before, the study is entirely observational in nature. In cancer tissues/cell lines hundreds of genes are known to be differentially expressed both at the level of mRNA and protein. Furthermore, cancerous cell populations are highly heterogeneous. In any particular instance, the population of cancer cells will vary in their overall expression profiles and specifically their THOC1 and ALY expression. This is a major caveat of the current work and this point should not be ignored. This issue should be thoroughly discussed in the manuscript.

(2) I appreciate the authors’ effort in providing appropriate statistical analyses to support their conclusions. As I mentioned in my previous review, the issue of the multiple hypothesis testing has to be considered. Unfortunately, the authors have decided to ignored this suggestion. Even though the two-tailed t-test is the appropriate statistical test, one cannot claim significance at the level of p < 0.05 given that at least ~19 hypotheses are tested for each protein of interest. At this significance threshold, one would observe on average one spurious statistically
significant p-value even if there were absolutely no difference in expression. There are many ways to correct this problem. One popular and conservative way is to use the Bonferroni correction. In this case, one should require approximately a $p < 0.003$ to claim statistical significance. Given the nature of the study, a less conservative correction could be employed. Furthermore, the legend of supplementary table 2 has to rewritten to eliminate typos and language problems.

**Level of interest:** An article of limited interest

**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