Reviewer's report

Title: Comparison of metabolic ratios of urinary estrogens between benign and malignant thyroid tumors in postmenopausal women

Version: 1 Date: 22 May 2013

Reviewer: Thomas O Metz

Reviewer's report:

In their current manuscript, Moon and colleagues follow up on their 2011 BMC Cancer paper that described metabolic alteration of urinary steroids in women and men with papillary thyroid carcinoma (PTC) by focusing on the levels of specific estrogens in post-menopausal women with benign and malignant thyroid tumors. The overall goals of the current work were to evaluate the differential expression of estrogen metabolism between benign and malignant PTC tumors and to identify potential biomarkers for diagnosis and staging of PTC tumors.

The analytical approach taken was that described in the authors' 2011 Journal of Lipid Research paper. The overall experimental design was sound and had high potential of achieving the goals of the study.

However, I do have some concerns with the study as outlined below.

Major Compulsory Revisions:

1. In looking at the concentrations of the estrogens in Table 1, it is evident that there is very high variability in the levels of these urinary estrogens among the patients studied. The coefficients of variation of these concentrations are near 100% for most of the estrogens in all three groups studied. This is reflected in the fact that the levels of only 4 estrogens were found to be statistically different (3 of these have p values just below the cutoff of 0.05) between benign and stage I malignant PTCs. No estrogens were found to be statistically different between benign and stage III/IV malignant PTCs and between stages I and III/IV malignant PTCs. Indeed, I am surprised that there were no statistical differences between benign tumors and the more advanced stage III/IV tumors. I therefore wonder if the authors had sampled a larger population of individuals with thyroid tumors would they have observed then no statistical difference between any of the groups compared? Can the authors provide any additional discussion on the confidence of the data, given the relatively small sample size (and high p values for 3 of the estrogens), that leads to the identification of E1, 2-OH-E2, 2-MeO-E2, and 17-epi-E3 as statistically significant between individuals with benign tumors and those with stage I malignant PTC tumors. Can they also discuss why no statistical differences were identified between individuals with benign tumors and those with stage III/IV malignant PTC tumors? Is this due to the relatively low sample size and high variability in urinary estrogens among individuals or is there an expectation based on known biology that explains this?
2. The authors identify significant differences in the estimated activities of 6 enzymes (Table 2). These 6 enzymes catalyze reactions involving only 2 estrogens (E1 and 2-OH-E2) whose concentrations were found to be statistically different between individuals with benign tumors and stage I malignant PTCs (Table 1). It is unclear to me how the activities themselves for those enzymes that catalyze reactions for estrogens whose concentrations were not statistically different (e.g., E3 and 16alpha-OH-E1) can then be found statistically different. Are the average ratios of metabolites to substrates calculated based on the average concentrations of the estrogens reported in Table 1, or are they calculated based on the individual concentrations? Can the authors provide a supplemental table that shows the individual concentration for each estrogen in each individual? I worry here about the propagation of error in the calculations of the enzyme activities. Can the authors clarify exactly how the enzyme activities were calculated and also comment on how certain enzyme activities (e.g. 16alpha-hydroxylase as E3/E2 and 17beta-hydroxysteroid dehydrogenase as E3/16alpha-OH-E1) were found to be statistically different if their respective metabolite and substrate concentrations were not statistically different?

3. Related to points 1 and 2 above, the method that the authors used, which is based on that reported in their 2011 Journal of Lipid Research paper, has reported intraday coefficients of variation of 1.6 to 11.5% in analyses of the same sample. In the current manuscript, the authors do not mention if process or technical replicates were performed in the analysis of the urine samples. Therefore, the reader has no appreciation of the degree of replication in and therefore the confidence of the measurements. Can the authors add this information?

Minor Essential Revisions:

1. On page 3, second paragraph, the authors state that "However, these candidates have not been proved to have enough sensitivity and specificity for a screening test or for pathologic diagnosis." I suggest that the authors delete this statement, since they themselves have no data on the sensitivity and specificity of the candidate biomarker estrogens identified in the current study. Sensitivity and specificity are typically determined through the analysis of a blinded set of samples and through construction of receiver operator characteristic (ROC) curves.

2. The authors should modify the titles of the legends to Figures 1 and 2 to indicate that only the statistically different estrogens and enzyme activities, respectively, are plotted.

Discretionary Revisions:

1. Under the Conclusions section of the Abstract, the authors state that "The effects of increase of...". I suggest that the authors revise this statement to "The increase of..." since they are not measuring the effects of anything but are actually measuring the estrogens themselves.
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.