Reviewer’s report

Title: Histotype-specific copy-number alterations in ovarian cancer

Version: 1 Date: 8 May 2012

Reviewer: Terry Furey

Reviewer’s report:

The manuscript by Huang, et al. describes an analysis of copy number data for each of four histotypes in ovarian cancer. The goal of the study is to identify genomic regions with recurrent copy number changes (amplifications and changes) for each of the histotypes. In addition, they identified potential driver genes within these regions.

I think this is an important study that extends previous work in characterizing copy number changes in ovarian cancer. The subdivision into histotypes provide a novel analysis that describes one aspect of how these differ at a molecular and phenotypic level.

The following are questions and suggestions for this manuscript:

Major Compulsory Revisions

1. Obviously, based on the frequency of each histotype, different numbers of samples of each are available. While the authors briefly discuss the implications of sample numbers on their analysis, I think this should be explored more fully. The serous histotype is the main outlier, though Clear Cell has 50% more samples than the other two as well. In figure 1, you see part of the effect of this by the y-axis ranges in b), so more samples clearly allow for lower q-value scores. What would happen if you sub-sampled 20 of the serous and celar cell samples and re-calculated significance? This could be done several times to get an average overall effect and help better understand how sample size affects the analysis. This could be presented in the results section.

2. Related to the above, without understanding better the effect of the sample sizes on your ability to call copy number changes, then the discussion about how many “unique” alterations are present in each histotype is premature. It is not clear whether the enrichment in serous samples is simply due to greater power to detect these with more samples. The above might suggest what alterations our found in each at similar sample levels and whether one histotype appears to be more frequently altered.

3. The pathway analysis should probably go in the results section as this is a separate analysis, or possibly combined within the discussion of the driver genes. How this pathway analysis was done needs to be described, possibly in the Methods. In the paragraph prior to this discussion, it says “This implies several canonical cancer pathways are involved in the pathogenesis of serous
histotype.” What is the basis for this statement? Results of the pathway analysis?

4. I think it needs to be clarified what you mean by a “driver” gene. In general, a driver gene is one for which there is evidence that this gene plays a role in the onset or progression of the disease. Evidence of driver genes is often based on expression levels correlated with a clinical or molecular phenotype, such as survival or progression. Since these data are not mentioned, then care must be taken to indicate in all discussions that these are “candidate” driver genes, and it would be good to clarify your definition of what a “candidate” driver gene is in your study.

5. Along the same lines, the description of the driver gene analysis is confusing. I would assume that the first part results in a copy number status for a gene, though this isn’t completely clear. The ANOVA then was applied to find genes with histotype-specific alterations? What is being correlated for the Spearman test? I assume this was done for each histotype separately. In general, I think it would be best to split the Data Analysis section of the Methods into two parts – one focused on detecting copy number alterations and one on candidate driver gene analysis.

Minor essential revisions

1. On page 8, in the Consistency with other studies paragraph, it needs to be clarified how genes were considered to be “found” in your study for this comparison. Are these all genes in any significantly altered region in any histotype? Were overlaps determined by gene name or genomic region?

2. It would be good to note in the legend of Figure 1b that the y-axis is different in the histotype plots

3. Overall, the writing in the manuscript is good, but it would be good to have a native English speaker review this for grammatical corrections.

Level of interest: An article of importance in its field

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

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.