Reviewer's report

Title: BMP-2 response pattern in human lung fibroblasts predicts outcome in lung adenocarcinomas

Version: 2  Date: 4 January 2015

Reviewer: Henry Yangh

Reviewer's report:

The authors used BMP2 stimulation to obtain a so called "Fibroblast specific BMP2 induced gene list" by stimulating CCL-171 cells with BMP2 and treating two breast cancer cell lines with BMP2. Using the gene list, they found that the list could be used as prognostic markers for segregation of good and poor survival outcomes. However, both results in identification of the Fibroblast specific BMP2 induced genes and unsupervised clustering for classification of patient samples into two groups are confusing, and inconsistent.

There are multiple serious discrepancies in the results. Two major discrepancies are

1. Table 1 shows the genes with significant change greater than 1.5 fold after BMP2 stimulation of CCL-171. To obtain a common response gene set to BMP-2 shared by CCL-171 and other cells, the authors compared/integrated this data with the data generated from breast cancer cells treated with BMP2. The derived gene list (Table 3) from the integration should be a subset of Table 1. However, only a handful of them overlap with Table 1. Instead, the authors obtained a lot of genes not significantly changed in CCL-171 cells upon BMP2 treatment and claimed a novel discovery of e.g. POSTN (Table 3). This raises a critical issue of the correctness of the analysis method or procedure thus the reliability of the list in Table 3.

2. Assuming that the list in Table 3 is correct, there are 67 Fibroblast specific BMP2 induced genes (Table 3). In Figs 5-6, the authors try to demonstrate the prognostic significance of these genes in three lung adenocarcinoma datasets. However, the heatmaps show that in each of three datasets, a different gene subset was used in hierarchical clustering to segregate the patient samples into two groups, raising serious doubts about the claim of the prognostic power of the gene list (Table 3).

The other major comments are as follows:

3. The description of data analysis is not up to a professional (journal) standard. A lot of important information is missing. For example, how the data were normalized with which methods (lowess, quantile, or ...) and how many replicates were performed. Although it stated in the legend of Figure 1 "The gene expression levels were normalized to the non-stimulated specimens as described in the Methods section". However, such description can be found nowhere in the
Methods section. From heatmaps, one can guess that two replicates were performed, but what kind of replicates, technical or biological?

4. A 10-nearest neighbor imputation engine was applied to estimate missing data. It is well known that similar expression level of a set of genes in one case may not be preserved in other cases. In general, we should discourage data imputation. Nevertheless, the authors should at least show the statistics of the manipulated data such as how many spots or what percentage of genes were missing for each array. In particular, how many of the significantly changed gene list derived from SAM were computed using the imputation engine. In the heatmaps, all computed gene expression levels across all replicates should be indicated in order for the readers to judge the reliability of the gene expression levels.

5. Table 1 shows genes with a change in expression level greater than 1.5-fold after stimulation with BMP2. Why are 40+ genes with fold change between 0.678 to 1.479 in the table?

6. How can one understand the additional file 1 and extract the enriched GO terms such as BMP2 signature from the un-annotated network? What statistics was used in the GO Term analysis and how was p value corrected?

7. How can one understand the multiclass SAM analysis and time profiling analysis from the SAM plots in Figure 2A and Figure 4B, respectively? I strongly doubt that any score-score (SAM) plots presented in Figs 1-4 give any informative view of the experiment or the gene set. In Figure 2B, the heatmap should include all replicates to demonstrate their consistency. In all analyses (Figs 1-4), sample clustering with all genes should be performed and the sample clustering trees should be presented to show the replicate consistency.

**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.