Reviewer's report

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

Version: 3  Date: 16 February 2015

Reviewer: Tao Wang

Reviewer's report:

In this paper, the authors made effort to identify BMP2 induced gene expression signatures in human lung fibroblasts that can be potentially used to predict outcome in lung adenocarcinomas. The authors mainly used in vitro microarray experiments to identify induced common genes and specific genes between different cells. Then the fibroblasts specific BMP2 induced genes comprised an interested gene signature and it was compared to publicly available in vivo lung adenocarcinomas gene expression data sets. The results showed the majority genes overlapped with available data sets and have a significantly higher risk when they have higher expression levels. Overall, I think this paper have merits. Although there are many available works to study lung cancer using large scale microarray gene expression data, there’s not much to study transcriptional effects on fibroblastic stromal cells. The author’s work will further broaden our knowledge on complicated gene expression patterns according to different tumor stages. However, the paper has issues need to be improved and I provide some detailed comments below:

Major issues:

(1) It’s good that BMP4 was performed a time-series gene expression experiments. Why BMP2 and BMP7 didn’t do such kind of experiments? I think it’s important to make sure BMP2 induced genes are consistently expressed and thus will be more reliable for downstream analysis. Please provide additional data or explain it in manuscript.

(2) The authors stated that the response to BMP2 observed in CCL-171 cells was very similar to that observed in response to BMP4 and BMP7. I think such statement needs further evidence or clarification. Because the commonly induced gene number is 115 and only BMP2 induced number is 171. There’s quite large difference. Moreover, how many genes were overlapped between BMP2 induced 171 genes and 115 commonly induced genes?

(3) The manuscript should be improved. There are many redundant phrases. The figure captions have too much text. Basically they can be denoted only once somewhere in the manuscript. E.g. “Pearson correlation”; meanings of the rows and columns of heatmap; and other statements already explained in previous sections, etc. The authors may want to proofread the manuscript and double check.

(4) The authors firstly identified up-regulated genes and down-regulated genes...
by expression fold changes. Then later on SAM analysis was applied to identify
genes with significant changes. It seems they are duplicate work. For the SAM
analysis identified 10 down-regulated genes, what are the fold changes? (p. 7,
second para.) Please clarify.

Minor essential issues:
(5) The paper title can be more accurate, what outcome will be predicted?
(6) How many biological duplicate samples were used? Please clearly state that
in the manuscript.
(7) Does Excel-SAM-Package refer to SAM package Excel plug-in?
(8) Data with a regression correlation greater than 0.6… (p.5, data analysis
section). It’s not clear and more details are needed. And does the correlation
refer to Pearson correlation?
(9) “A list of 156 unique genes comprising fibroblast specific BMP2 induced gene
list was extracted from three datasets. However, such gene list only contains 67
genes in table 7. Please clarify and see if they are consistent about the
interpretation.
(10) 206 genes showed a more than 1.5-fold decrease in the expression level
(mean: 0.386, standard deviation: 0.12) (Results section, para. 1). This is
confusing because the cutoff is 1.5 and why the mean expression level is only
0.386?
(11) How was hazard ratio calculated? More details are needed.
(12) The statement of Garber’s data set is not consistent. In “Human cancer
datasets” section it was said 67 human lung cancers, but in “prognostic
significance” section it was said 41. Please double check.
(13) The False discovery ratio sometimes was set to 1%, sometimes set to 0.8%.
Is there any specific reason? Why not just make them consistent?
(14) In the experiment protocols, BMP4 concentration was 24 while BMP2 and
BMP7 were 200. Why’s that and any specific reason?
(15) The overlaps between 67 “Fibroblast specific BMP2 induced gene list” and
Garber dataset, Lee dataset are 37 and 48, respectively. So the in vitro and in
vivo genes are not high concordant. What’s the overlap between 37 and 48
genes?
(16) In the figure 2, BMP-6 should be BMP-7?
(17) Figure 5 should be improved. X-axis labels overlaps with those numbers.
(18) Please double check the gene symbols. They are not consistent, such as
BMP-2, BMP2 among the manuscript.

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

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