Author's response to reviews

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

Authors:

Michal Rajski (michal.rajski@access.uzh.ch)
Annika Saaf (asaaf@uchicago.edu)
Martin Buess (Martin.Buess@claraspital.ch)

Version: 4 Date: 1 March 2015

Author's response to reviews:

Dear Editor

Thank you for further considering our manuscript for publication in BMC medical genomics. Please find attached our manuscript revision in response to reviewer 3. We thank the reviewer for the careful review of our manuscript. Specifically we replied to or changed the following points (changed parts in the manuscript are underlined):

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.

We have performed a time course experiment for BMP4 to exemplarily show the development of the gene expression changes in response to BMP stimulation over time. That the expression of genes after stimulation with BMP4 over 24h and BMP2 over 24h is very similar is shown by the Pearson correlation in figure 2 (r= 0.892 p>0.001) as we have elaborated in response to issue 2. We therefore judge our gene list of BMP2 reliable for further downstream analyses. In our opinion further data using time course experiments with BMP2 or BMP7 would not add relevant additional information. We have adapted paragraph “BMP4 stimulation of lung stromal fibroblasts induces a time-dependent gene expression signature”.

(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?

As expected the different analysis methods of the “CCL-171 derived BMP2 signature” and the “Common BMP induced list” give similar but not congruent results. In our opinion describing the overlap is not the method of choice to judge the similarity. We therefore checked the correlation of the expression of the BMP2 induced genes with the expression upon stimulation with BMP4, BMP7, Noggin and Gremlin by Pearson correlation. We displayed the data as a scatter plot which we integrated in figure 2 together with the Pearson correlation coefficient r. From this analysis the strong and highly significant correlation of BMP2 with BMP4 and BMP7 can be seen and the inverse correlation as expected for antagonists Noggin and Gremlin. We have adapted the methods section, the paragraph “A global picture of genes that are differentially expressed in response to BMP stimulation or inhibition in lung stromal fibroblasts” in the results section and Figure 2 with its legend.

(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.

We have eliminated redundancy in the results section and in the legends of figures 2-4.

(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.

As expected the selection of genes by fold change cut-off and by SAM gives similar but not congruent results. In our case the fold change was the more rigorous selection method than SAM. For the additional genes identified to be significantly down-regulated by SAM we found the following fold changes (decrease): GNG11 (2.29), ABHD5 (2.23), PHF17 (2.23), GAS1 (2.59), ARSI (2.29), JUN (2.25), KLHL29 (2.20), CHEK2 (2.14), ST6GAL1 (2.25), PDGFRA (2.23), H1F0 (2.21), BTG1 (2.19) (fold changes of these genes: 2.15-2.58 fold decrease).

Minor essential issues:

(5) The paper title can be more accurate, what outcome will be predicted?

We prefer to leave the title as it is.
(6) How many biological duplicate samples were used? Please clearly state that in the manuscript.

Two biologically independent duplicate samples were used. This is explicitly mentioned in the results section page 8.

(7) Does Excel-SAM-Package refer to SAM package Excel plug-in?
Yes it is the Excel plug-in. This package was obtained from the webpage of the Tibshirani Lab: http://statweb.stanford.edu/~tibs/SAM as described in the methods section SAM analysis.

(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?

We were interested in high quality data. As a measure of spot quality for a specific probe the Pearson regression correlation of the pixels within the spot was used and the cut-off value set to 0.6. We have changed the paragraph in the analysis section.

(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.

156 unigene clones comprising 67 genes. We clarified this issue in the methods section paragraph “Human cancer datasets”.

(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?

We realize that the presentation of the decrease in expression might be confusing and therefore changed this paragraph and the presentation in table 1. For purpose of consistency we also adapted the expression of fold changes in Table 4.

(11) How was hazard ratio calculated? More details are needed.

All statistical tests were performed using the R statistical software (version 2.10.1). Survival curves were obtained using the Kaplan-Meier estimator and univariate Cox proportional hazards regression models were fitted (R package “survival”) R Development Core Team. A language and environment for statistical computing. Available: http://www.R-project.org. Accessed 2013 December 2. We have clarified this in the methods section.
(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.

Garber published data of 67 human lung cancers of different histologies. 41 were adenocarcinomas and we focused on these. We specified this point in the methods section.

(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?

The false discovery rate is an arbitrary cutoff value. There is no specific reason for the small differences in the chosen cut-off values and we decided to leave it as described.

(14) In the experiment protocols, BMP4 concentration was 24 while BMP2 and BMP7 were 200. Why’s that and any specific reason?

The physiological potency of BMP4 is higher than that of BMP2 and BMP7 e.g. Herrera, Blanca, and Gareth J. Inman. "A rapid and sensitive bioassay for the simultaneous measurement of multiple bone morphogenetic proteins. Identification and quantification of BMP4, BMP6 and BMP9 in bovine and human serum." BMC cell biology 10.1 (2009): 20. Therefore the indicated concentrations were chosen for the experiments and mentioned in paragraph “A global picture of genes that are differentially expressed in response to BMP stimulation or inhibition in lung stromal fibroblasts”.

(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?

The previously published datasets of Garber and Lee are missing several genes of the “Fibroblast specific BMP2 induced gene list”. Because different platforms were used for the generation of our in vitro datasets, the Garber dataset the Lee dataset and the Bhatacharjee dataset they have different missing genes. The overlap between the Garber and the Lee datasets is 33 genes. Based on the overlapping or missing genes, which have technical issues, one cannot conclude anything about in vitro and the vivo concordance of the signatures. To overcome the problems with missing values we have cross-checked the prognostic values of the signatures based on a continuous score as described in paragraph “Prognostic significance of the “Fibroblast specific BMP2 induced gene list” in human lung adenocarcinomas” and shown in Figure 7.

(16) In the figure 2, BMP-6 should be BMP-7?
We fixed the typo.

(17) Figure 5 should be improved. X-axis labels overlaps with those numbers.
We thank for the indication and changed the figure.

(18) Please double check the gene symbols. They are not consistent, such as BMP-2, BMP2 among the manuscript.
Gene symbols were checked and made consistent throughout the manuscript.

We hope that with this revision the concerns could be eliminated and that the manuscript is now acceptable for publication in BMC medical genomics.

Best regards

Martin Buess