Author's response to reviews

Title: Clinicopathologic and Gene Expression Parameters Predicts Liver Cancer Prognosis

Authors:

Ke Hao (ke_hao@merck.com)

Version: 2 Date: 22 August 2011

Author's response to reviews: see over
Reviewer# 1

The authors used gene expression profiles of normal and tumor tissues to improve the prognosis power of clinical biomarkers in liver cancer. They showed expression profiles can further divide good and poor survival groups into subgroups with differential survival outcomes. The resulting gene signatures overlapped significantly with previously published survival signatures, and with the SNPs from eQTL studies of liver cancer.

Major Compulsory Revisions

1. The improvement of expression profiles was assessed within the good and poor prognosis groups of clinical parameters (stratified model). Alternatively, one can build a single prediction model using both clinical biomarkers and expression profiles (both normal and tumor), and compare it to the stratified model and the models using only clinical parameters or only expression profiles.

The reviewer's comment is well-taken. We mainly focused on stratified models for at least two rationales. First, it is a nature extension of our previous work on using clinicopathological parameters alone to predict HCC prognosis. In this paper, on top of the strata defined by clinicopathological parameters, we showed gene expressions further enhance the prediction. Second, in each strata, the gene expression profiles offered various information content. For example, many genes in the tumor tissue were associated with survival in the good-survival group but not in the poor-survival group (Figure S2), suggesting that additional tumor gene expression may further enhance prognosis prediction in the good-survival group but not the poor-survival group.
We followed the reviewer's suggestion and built a single prediction model using both clinical biomarkers and expression profiles (both normal and tumor), and compare it to the stratified model and the models using only clinical parameters or only expression profiles. In the revised manuscript, we added, "As we previously showed, several clinicopathologic parameters 1) provided excellent predictive power for outcome in HCC [10] 2) were easily and routinely measured, and 3), resulted in predictions that were readily applicable to clinical practice. Given this, it is natural to attempt to further enhance the clinicopathology-based prediction model by adding gene expression data. We conducted a head-to-head performance comparison between gene expression predictors derived from normal and tumor tissue (denoted as $h_{\text{gene-expression}}$) vs. predictors derived solely from clinicopathology ($h_{\text{pathology}}$; Materials and Methods) and benchmarked them in a LOO framework (Figure 1 and Figure S4). Please note, the genes used in the prediction models might be different regarding to normal vs. tumor tissue expression, as well as in each LOO iteration. Overall, $h_{\text{gene-expression}}$ and $h_{\text{pathology}}$ performed similarly. The $h_{\text{gene-expression}}$ of tumor tissue beat $h_{\text{pathology}}$ in predicting DFS, but $h_{\text{gene-expression}}$ slightly underperformed $h_{\text{pathology}}$ in all other scenarios. Taken together, gene expression was not superior to clinicopathology in predicting prognosis. One reason might be that gene selection primarily identified genes correlated with clinicopathologic parameters (e.g. cancer stage)." Afterwards, we investigated the single prediction model incorporating both clinical biomarkers and expression profiles. We added, "The main focus of this study is stratified modeling since it was a natural extension of our previous work. Alternatively, we can build a single model incorporating clinicopathological parameters and gene expression profile simultaneously (Figure S5). The prediction framework was identical as above analyses except the multivariate Cox model included both the clinicopathological parameters and the top 6 PCs. It is noteworthy that, in the gene selection step, we also included clinicopathological parameters in the Cox model and then picked 100 genes with smallest pvalues. The overall prediction was better than using gene expression alone (Figure S4), indicating clinicopathology captured valuable information beyond
gene expression. In the other hand, comparing to Figure 1, adding gene expression only enhanced prediction in one scenario (tumor gene expression improved the prediction of survival). A possible explanation would be that different gene sets were associated with prognosis across the strata defined by clinicopathology, and these gene sets offer various prediction value. For example, shown in Figure 2, normal tissue expression was used in prediction in the good-survival stratum but not the poor-survival stratum. In the single model approach, genes with little prediction value also entered the model, bringing noise and reducing the performance. Lastly, we also evaluated a single model incorporating clinicopathological parameters, and expression profiles of both normal tissue tumor tissue (Figure S6). Herein, expression profile of each tissue were reduced to 6 PCs, therefore, a total of 12 PCs entered the model. Overall, such models did not greatly outperform the models based on clinicopathology alone (Figure 1) or models based on clinicopathology + expression profiles (Figure S5). Again, this lack of improvement could be attributable to noises introduced into the prediction."
Figure S6

Leave-one-out Analysis using Normal and Tumor Tissue Gene Expression (crude) and Clinico-pathological Parameters as Predictors
Minor essential revisions

Fig S1 not included.

We thank the reviewer for pointing out this error. In the original submission, we wrongly labeled the Figure S1 as Figure S2C. This error has been fixed in the current version. We also carefully checked and confirmed that all Figures were correctly numbered and referred.

**Level of interest:** An article of importance in its field

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

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

I declare that I have no competing interests
Reviewer's report:

The authors combined the clinicopathologic characters and gene expression signatures to predict the outcome of Liver cancer patients. The work is rather meaningful. However, some points need to be revised.

- Major Compulsory Revisions

1. In the part of "Further Classification of good and poor prognosis groups using gene expression...", the authors should explain the principle components (PCs) in detail.

   In the Methods section, we added the details in dimension reduction using PCA and described the algorithm explicitly, "In each LOO iteration, we reserved one patient for testing and used the remaining patients (say N-1 patients) for training. On these N-1 training patients, we run univariate Cox model on each gene and derive the pvalue for association between prognosis (e.g. survival) and this gene's expression value. Then we picked the 100 genes of the smallest pvalues. On these 100 genes over N-1 patients, we constructed the 6 PCs, which then served as independent variables in a multivariate Cox model. This model and its coefficients captured the association between HCC prognosis and the PCs. Next, we projected the gene expression values of the reserved patient to the PC space (defined on the training patients) to obtain the coordinates on the first 6 directions. We plugged these 6 coordinates into the multivariate Cox and calculated the relative hazard for this reserved testing patient. By these means, after N LOO iterations, we derived the relative hazard for every patient, which was actually the linear predictor. Lastly, we used the log-rank test and Kaplan-Meier plot to examine and visualize the performance of the linear predictor. This scheme is similar to a previous report..."
on using clinicopathologic parameters to predict HCC outcome [10], except for the extra step of PCA dimensionality reduction."

2. As the authors have mentioned, there were really relationships between gene expression signatures and clinicopathologic characteristics. If mixed the two parameters in one model, the relationships should be carefully discussed.

The reviewer's comment is well-taken. We mainly focused on stratified models for at least two rationales. First, it is a nature extension of our previous work on using clinicopathological parameters alone to predict HCC prognosis. In this paper, on top of the strata defined by clinicopathological parameters, we showed gene expressions further enhance the prediction. Second, in each strata, the gene expression profiles offered various information content. For example, many genes in the tumor tissue were associated with survival in the good-survival group but not in the poor-survival group (Figure S2), suggesting that additional tumor gene expression may further enhance prognosis prediction in the good-survival group but not the poor-survival group.

We followed the reviewer's suggestion and built a single prediction model using both clinical biomarkers and expression profiles (both normal and tumor), and compare it to the stratified model and the models using only clinical parameters or only expression profiles. In the revised manuscript, we added, "As we previously showed, several clinicopathologic parameters 1) provided excellent predictive power for outcome in HCC [10] 2) were easily and routinely measured, and 3), resulted in predictions that were readily applicable to clinical practice. Given this, it is natural to attempt to further enhance the clinicopathology-based prediction model by adding gene expression data. We conducted a head-to-head performance comparison between gene expression predictors derived from normal and tumor tissue (denoted as h\textsubscript{gene-expression}) vs. predictors derived solely from clinicopathology (h\textsubscript{pathology}; Materials and Methods) and benchmarked them in a LOO framework (Figure 1 and Figure S4). Please note, the genes used
in the prediction models might be different regarding to normal vs. tumor tissue expression, as well as in each LOO iteration. Overall, \(h_{\text{gene-expression}}\) and \(h_{\text{pathology}}\) performed similarly. The \(h_{\text{gene-expression}}\) of tumor tissue beat \(h_{\text{pathology}}\) in predicting DFS, but \(h_{\text{gene-expression}}\) slightly underperformed \(h_{\text{pathology}}\) in all other scenarios. Taken together, gene expression was not superior to clinicopathology in predicting prognosis. One reason might be that gene selection primarily identified genes correlated with clinicopathologic parameters (e.g. cancer stage)." Afterwards, we investigated the single prediction model incorporating both clinical biomarkers and expression profiles. We added, "The main focus of this study is stratified modeling since it was a natural extension of our previous work. Alternatively, we can build a single model incorporating clinicopathological parameters and gene expression profile simultaneously (Figure S5). The prediction framework was identical as above analyses except the multivariate Cox model included both the clinicopathological parameters and the top 6 PCs. It is noteworthy that, in the gene selection step, we also included clinicopathological parameters in the Cox model and then picked 100 genes with smallest pvalues. The overall prediction was better than using gene expression alone (Figure S4), indicating clinicopathology captured valuable information beyond gene expression. In the other hand, comparing to Figure 1, adding gene expression only enhanced prediction in one scenario (tumor gene expression improved the prediction of survival). A possible explanation would be that different gene sets were associated with prognosis across the strata defined by clinicopathology, and these gene sets offer various prediction value. For example, shown in Figure 2, normal tissue expression was used in prediction in the good-survival stratum but not the poor-survival stratum. In the single model approach, genes with little prediction value also entered the model, bringing noise and reducing the performance. Lastly, we also evaluated a single model incorporating clinicopathological parameters, and expression profiles of both normal tissue tumor tissue (Figure S6). Herein, expression profile of each tissue were reduced to 6 PCs, therefore, a total of 12 PCs entered the model. Overall, such models did not greatly outperform the models based on clinicopathology alone (Figure 1) or models based
on clinicopathology + expression profiles (Figure S5). Again, this lack of improvement could be attributable to noises introduced into the prediction."
- Minor Essential Revisions

Some of the tables in the submitted work were not referred in the work and not well explained.

Reviewer's comment is well taken. In the original submission, we wrongly labeled the Figure S2C, therefore, it was not properly referred in the text. This error has been fixed in the current version. We also carefully checked and confirmed that all figures were correctly numbered and referred. In the revision, we provided in depth explanation of the tables.

- Discretionary Revisions: None

Level of interest: An article of importance in its field

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

The authors thank reviewer's comment. We have carefully checked the manuscript and corrected language errors.

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.