Dear Julio Vera,

We would like to thank you and the reviewer very much for your work and for the constructive critism, which helped us to improve our manuscript. Below please find a detailed account of our responses to the reviewer’s questions and how we have dealt with them in the revised version of our manuscript.

**Reviewer #2:** The manuscript "Pathobiochemical signatures of cholestatic liver disease in bile duct ligated mice" by Abshagen et al. aims at identifying markers for chronic liver disease progression by measuring physiological parameters as well as the expression of selected genes in mice following disrupted bile ligation over time up to 14 days.

The authors performed extensive correlation analyses between the expression data and the various hematological measurements and combined them in a "consensus correlation" value to predict, which genes and factors explain best the different disease phases.

Finally, they propose a decision tree based on the main markers they found to predict the disease stages. The manuscript is a nice example of a collaborative effort in experiment, analysis and interpretation, yielding interesting new insights into the chronic liver disease. Overall, the results are well explained and compared to the literature. In particularly the decision tree approach is a good idea to translate the experimental data into new biological insight of chronic liver disease progression.

**Nevertheless, the manuscript needs a major restructuring of the data analysis and a clear focus in its presentation. The reader is flooded with gene lists, information and interpretation, while the motivation and explanation of the analysis is underrepresented in the main text. All information is there, but scattered throughout the main text, methods section and supplement.**

*Answer: we thank the reviewer for the objective estimation of the manuscript. Our main focus in this study was to generate time resolved multi-level knowledge on cholestasis disease progression in a holistic manner, aiming to define robust biomarkers representative for the different disease stages from initiation to severe fibrosis as candidates for diagnosis.*

*Changes in the manuscript: to make our motivation more clear, we added corresponding text in the abstract, background and main body parts marked with red font and underlined (see page 3, lines 5 and 17; page 5, line 21; page ).*

**In particular Figs. 8 and 9 are overloaded with information and are hard to understand.**

*Answer: Figures 8 and 9 were redrawn……*

**The manuscript would win substantially, if the authors focused on selected genes and results, better explained their analysis approaches in the main text and redrew their figures in a more intelligible format with readable fonts and structured information.**

*Answer: Due to new mathematical analyses and application of dimension reduction methods….we focused on much less genes….*

*Nevertheless, as already mentioned above, it was our specific intention to represent the time-dependent disease progression following bile duct ligation in a comprehensive holistic manner. Several studies are already available which describe the expression behavior of the selected parameters. Thus, we explicitly wanted to systemize existing and newly acquired knowledge on the molecular biomarkers of cholestasis*

**Major Points**

**- The authors should consider dimension reduction methods such as multidimensional scaling or principal component analysis, when discussing the samples' expression patterns, use statistical tests, when assessing the significance of gene regulation over time and display ROC curves, when discussing their decision tree model. This would reduce lengthy description of the data and provide better overview on the dynamic behavior of the system.**

*Answer:* *As suggested by the reviewer, we considered dimension reduction methods….*

*Changes in the manuscript: …*

**- For the qPCR data the authors used a single Gene, Gapdh, for normalization, which possibly results in noisy dCT (delta CT) values. There is no guarantee that the expression of this gene remains constant across the samples, given the severe impact of BDL and the measurement time of 14 days. Usually, dCT values are normalized to two control genes, also e.g. 18S and/or Hprt1. The authors need to check and show the behavior of Gapdh and that their normalization approach does not affect their results.**

*Answer:* *We totally agree with the reviewer. Parallelization of qPCR such as by the microfluidic Taqman Fluidigm Biomark Platform enables evaluation of multiple transcripts in the samples treated under various conditions. Despite of advanced technologies, correct evaluation of the measurements remains challenging. We wanted to apply ΔΔCt method and therefore assessed expression of two reference genes in order to find the stable one. We accurately controlled the expression levels of Gapdh and Actb over the time period and their variability between different Fluidigm runs and chose Gapdh for normalization due to its stability. We assure that the normalization approach does not affect the main results of our evaluation.*

*Changes in the manuscript: none*

**- The authors state a delta delta CT of approx. 17 between IL28b and Gapdh (Fig. 6K, 5d). This is a rather large difference and would mean that IL28b appears around the qPCR cycle 5 already. This should be checked. What is the primer efficiency for IL28b? Maybe it is unusually high.**

*Answer:* *Indeed the Ct value of the IL28b expression is very low. All gene expression assays used in this study were purchased commercially as validated assays by Applied Biosystems. We analyzed the amplification efficiency of the IL28b assay on our own using the dilution series of a control cDNA sample and could reproduce the data provided by the manufacturer (E ranged 100± 5%).*

*Changes in the manuscript: none*

**- The benefit of the consensus score needs to be better discussed, e.g. the choice of different weights seems rather arbitrary. In supplement2, 1.2 Consensus correlations the authors state that "the correlation of time average has 4 times higher weight". Where does this number come from? Why are the correlations separated in positive and negative parts and then the smaller part are ignored? A lot of information is lost this way. The score should be normalized to be able to compare in between factors.**

*Answer:*

*Changes in the manuscript: …*

**- Why do the authors use a consensus measure to perform clustering (Fig. 7)?**

**Why do they not use all correlation data for each factor as a matrix,**

**and then apply clustering method on this matrix instead?**

*Answer:*

*Changes in the manuscript: …*

**- The authors used the Pearson Correlation, which can be easily influenced**

**by outliers. In particular in combination with the noisy qPCR data this can lead to**

**spurious correlations. The authors should check for consistency of their results by**

**using either more robust correlation measures like Spearman correlation**

**or by low-pass filtering their data before performing the analysis.**

*Answer:*

*Changes in the manuscript: …*

**- The description of the separator approach in the manuscript is very wordy and formulas are necessary to understand what has been done.**

*Answer:*

*Changes in the manuscript: …*

**- Figures 8 and 9: Significance and consensus scales are not really readable.**

**It is hard to distinguish between the not significant ones (0.1) and the significant ones (<0.05). For example, few different colors would be more efficient than yellow gradient.**

**Also the correlation scale should be from -1 to 1.**

*Answer:*

*Changes in the manuscript: …*

**- Overall, figure legends need to be improved, as they lack sufficient annotation**

**to understand what is displayed.**

*Answer:* *agree*

*Changes in the manuscript: All figure legends have been carefully revised to improve readability and to supplement missing indications. Changes are marked with red font and underlined.*

**Minor Points**

**- All gene symbols should be written consistently with small letters and a capital**

**first letter throughout the text and figures, e.g. use Gapdh instead of GAPDH.**

*Answer:* *As recommended, the manuscript text has been corrected for the consistently notation of all gene symbols with small letters and a capital first letter.*

**- in the derivation of the consensus correlation the authors use a rather unusual -log100 transformation. is this true or a typo?**

**- The relative expression of the genes in Fig. 6 should be displayed in log2 scale and the domain ranges should be the same for all genes to make the changes in gene expression comparable.**

*Answer:* *As suggested by the reviewer, we have modified figure 6 and now display gene expression in log2 scale.*

**- Most likely, the authors used a log2 scale in Fig. 5, but annotation of the color bars is missing and needs to be added.**

*Answer: Annotation of the color bars in Fig. 5 was already included in the original version of the manuscript, but has now been improved.*

**- Page 9 lines 17 and 19, Page 14 line 19: Figs. 7 and 8 are actually Figs. 8 and 9**

*Answer:* *We corrected these mistakes.*

**- Place Figure 7 after Figure 8, as the former is based on the latter**

*Answer:* *To follow the reviewers suggestion, we rearranged these figures and placed figure 7 after figure 9.*

**- The abbreviations in the leaf names of the circular tree in Fig. 7 are nowhere explained in the main text and need to be added.**

*Answer:* *We now included a more detailed description of the used abbreviations in the figure legend.*