Peer Review File

Article information: http://dx.doi.org/10.21037/tcr-21-2184

Reviewer Comments

Colorectal cancer (CRC) is the third most common cancer and the second leading cause of cancer-related mortality worldwide. Therefore, it is crucial to identify new prognostic biomarkers for cancer patients.

In this study, the authors examined the relationship between metabolic-related gene expression and long non-coding RNAs (lncRNAs) expression using the Cancer Genome Atlas (ATGC) transcriptional data of colon cancer. The authors showed that four metabolism-related lncRNA signatures might be useful for predicting the prognosis of CRC patients.

While the authors' results may be important in developing new prognostic biomarkers of CRC, there are several issues that need to be addressed.

Major points:

Comment 1: It is still unclear how the authors identified metabolism-related lncRNAs. The authors should include a flow chart about how the authors extracted the four lncRNAs in figure 1. Especially, it is still obscurity how the authors performed the correlation analysis using differentially expressed genes (DEGs) and lncRNAs.

Reply 1: We added the flow diagram in Figure 1. The correlation analysis result had been shown in the supplementary materials table S1 labeled “correlation.result”.

Changes in the text: We have added the flow chart in our text, see Page 8, line 157-158.

Comment 2: The authors defined the four lncRNAs as being related to metabolism. However, there is a lack of description of how the four lncRNAs relate to metabolism. Do the four lncRNAs have any function in metabolism regulation? Are the four

lncRNAs located close to metabolic genes? The authors need to add explanations about the possibilities in their paper.

Reply 2: We added functional enrichment analysis from the gene list that statistically correlated with the four lncRNAs respectively. The statistically enriched terms has shown in supplemental Figure 1, and it is obvious that the genes all related to the metabolism of lipids, which supports that the lncRNAs relate to metabolism.

Changes in the text: we have added the functional analysis method and result in our text, see Page 7, line 140-143; Page 9, line 179-183.

Comment 3: Using the TCGA data set, the authors showed that the four-metabolism-related lncRNA signature is useful for predicting the prognosis of CRC patients. However, it is still unclear whether the readers can use the signatures for other datasets. The authors need to check whether the signature is also useful using the different data sets.

Reply 3: We try to use the data is in GEO database for the verification of the results. Due to the microarray data in GEO database derived from the different sequencing platform, we first yielded 6 searching result of platforms, based on the keyword of "colon cancer survival". Then data were searched again based on the organism selected homo sapiens, and the remaining 4 sequencing platforms were GPL570, GPL96, GPL571, GPL10558. We searched our four lncRNA in the 4 platforms datasets based on the gene symbol ID, unfortunately only PCAT6 included in the GPL570 dataset, the other three lncRNA were not in the database. So we regret that we cannot easily check whether the signature is also useful using the different data sets.

Changes in the text: No changes were made to the manuscript.

Comment 4: Figure 4 seems very incomplete. The authors indicate that there are differences in the gene sets of the Treg, iDCs, APC co-stimulation, DCs; however, the differences were very small, and there were no differences in CD8 T cells and check-point. The differences in the immune-related genes that the authors pointed out do not

seem important. In addition, the authors performed gene-set enrichment analysis (GSEA); however, the implications of the results are still unclear. How the pathway affects the prognosis of patients? Are the metabolic genes involved in the pathways which they found? The authors need to add the data or explanation in their paper.

Reply 4: Besides amending the description of the difference in immune function between high- and low risk group, we also made several textual adaptations in the main text, especially in the "discussion section". The reason why the differences were very small have been discussed. Based on the comments of the reviewers and editors, we have substantially revised the manuscript and the figures about the GSEA part. We added the enriched gene sets result description, then metabolism-related pathways were selected for further analysis and displayed in figure 4B. The genes involved in the representative pathways, which were shown in the figure 4B, were listed in the table S3. As for how the pathway affects the prognosis of patients, relevant discussion has been added to the revised manuscript. Meanwhile, we adjusted the corresponding references.

Changes in the text: We have modified the results description and discussion about the difference between high- and low risk group in terms of immune function and GSEA to make the conclusion clear. See Page 11, line 236-239; Page 12, line 241-255; Page 14-15, line 301-310; Page 15, line 311-325.

Minor points:

Comment 1: In lines 52-53, the authors mentioned that they developed a "six" metabolism-related-lncRNA signature. What does this "six" come from? Could you mention more specifics?

Reply 1: This is a clerical error, we revised the text.

Changes in the text: we changed "six" to "four", see Page 3, line 52.

Comment 2: The authors explained adipose tissue, insulin, and obesity in the Introduction. However, the authors examined the gene expression of colon cancer

tissue and did not examine whether their LncRNA signature is related to obesity or adipose tissues. It is better to focus on the metabolism in cancer tissue and the specific metabolisms related to the four-lncRNA signature.

Reply 2: To focus on the abnormal metabolism in cancer development, we have revised the relevant content in the Introduction section. Meanwhile, we adjusted the corresponding references.

Changes in the text: We have revised the Introduction accordingly, see Page 4, line 71-87.

Comment 3: In Material and Methods, the authors should explain how the authors obtained the R-packages and what version of R-packages the authors used.

Reply 3: We have added the software versions to the methods section of the manuscript.

Changes in the text: We added version of R software, and provide a link to this software, see Page 6, line 118. We added R package versions that used in this article, see Page 7, line 133-135; Page 7, line 145.

Comment 4: In line 224, the authors mentioned that they displayed the top 5 pathways. However, it is still unclear which five pathways exactly. The authors should mention them by name.

Reply 4: We have substantially revised the manuscript and the figures about the GSEA part, metabolism-related pathways were selected in figure 4B.

Changes in the text: We added the displayed 4 pathways by name, see Page 12, line 245-246; Page 12, line 252-253.

Comment 5: In Figure 4A, there is a lack of explanation about colors. In addition, the authors should add an explanation of the abbreviations. For example, what does CCR stand for?

Reply 5: We have added the requested annotation about colors, and the explanation of the abbreviations in figure legends.

Changes in the text: The annotation about colors was added to Figure 4A, see Figure 4A, right-top corner. The explanation of the abbreviations were added in figure legends, see Page 22, line 465-471.

Comment 6: In Figure 4B, the GSEA data is very confusing. For example, the authors should separate the high pathways data in high-risk and low-risk patients.

Reply 6: As suggested, we re-performed our data and re-name the pathways data.

Changes in the text: We removed the not mentioned data, and the pathways data was separated into “high risk group” and “low risk group”, see Table S2.