Peer Review File

Article information: http://dx.doi.org/10.21037/ tcr-20-845

Comment 1: The author did not mention the number of patients enrolled in this study.
Reply 1: Thank you for raisng this question. We showed in lines 94-95 that the GSE44076 dataset contains 98 colon cancer tissue samples and 98 adjacent normal tissue samples, and the GSE47074 dataset contains 4 colon cancer tissue samples and adjacent normal tissue samples, respectively. But we really did not mention the number of samples in the TCGA dataset, so we added the number of samples in the TCGA COAD dataset in lines 100-101.
Changes in the text: In text 4 line 100-101, we added the number of samples in the TCGA COAD dataset.

Comment 2: What are the inclusion and exclusion criteria?
Reply 2: Thank you for your questions about inclusion criteria and exclusion criteria for the dataset used in this study. We believe that this issue does need to be corrected to enhance the rigor of our article. Actually, we have wrote the screening criteria for the GEO database for chip filtering in text lines 82-87. But for clarity, we re-describe the screening criteria as inclusion criteria and exclusion criteria. In addition, since the data in the TCGA database has been collated according to the unified standards, inclusion and exclusion criteria have been included in its inherent logic. And we downloaded and analyzed the colon cancer dataset (COAD) from it, which we considered to meet the inclusion criteria of this study, so there was no additional explanation.
Changes in the text: In text page3 line 82-87, we change the original screening criteria for GEO database to inclusion criteria and exclusion criteria.

Comment 3: What is the method of selection and duration of the study?
Reply 3: Thank you for raising this issue. In the datasets screening of GEO database, we strictly followed our inclusion criteria and exclusion criteria to select the study in the last 10 years. We qualified the search strategy as "colon AND normal" in the GEO Datasets sub-database, selected "entry type" as "series" and limited "Publication dates" to after January 1, 2010. The retrieval results were screened one by one, and finally 2 datasets that most conform to our standards were selected for further analysis. In view of the question you have raised, we think it is necessary to describe the relevant retrieval methods in the text, so we make corresponding modifications. However, the TCGA COAD dataset was directly downloaded and analyzed using R software, and the software packages used have been described in line 88-95 in the text, so we didn't add the selection and duration of the study from TCGA.

Changes in the text: We describe the retrieval methods of screening datasets in GEO database. (See page 3, line 88-95)

Comment 4: What is the demographic data of the involved cases?

Reply 4: We are sorry that we didn't list the specific demographic data of involved cases in our text. The number of dataset samples from the GEO and TCGA database for screening differentially expressed genes has been explained in the reply 1. In our study, the data used to verify the hub gene expression and hierarchical clustering in UCSC database and used to obtain the hub gene co-expression network in cBioportal database were all obtained from TCGA database. The number of samples used for survival analysis in Starbase database was shown on the survival graph (Fig 6). The Oncomine database used to verify the characteristics of hub genes and colon cancer contains multiple sample data sources. The analysis results in figure 7 and figure 9-10 show the sample size of the relevant cases we selected for analysis. However, due to the limitations of each database, the specific demographic data of each study were not further obtained. This is a very good question, and in future research we can try to obtain demographic data from relevant databases for analysis.

Changes in the text: The modification of the sample size has been explained in reply 1. Due to the limitations of each database, no specific demographic data for each study was obtained. So, no other modifications were made in the text on this comment.

Comment 5: The authors analyzed the expression levels of hub genes between colon cancer and normal colon tissues, using 3 groups of the colon cancer tumor cells (HCT-8, SW620, and SW480) and a normal colon cell (HT-29). This was mentioned by the author in the methodology section, however, in the discussion section, the author said (Furthermore, we observed that the expression levels of these eight genes in patients with grade 1 cancer were higher than those in patients with grades 2-3 cancer) but in the result section, the author mentioned that they used both cell line and tissues from a colon cancer patient with different grades and normal colonic tissue. What a contrast!!! I cannot understand Is in the current study qRT-PCR was performed using cell line or cancerous and normal colonic tissue? Please could you clarify.

Reply 5: Thank you very much for pointing out our mistake. As a result of our mistakes in writing, the result, discussion and method were not consistent. In the qRT-PCR experiment, three groups of colon cancer cells (HCT-8, SW620, and SW480) and one group of normal colon cell (HT-29) were used to evaluate the expression level of hub genes. Therefore, we modified the expression of qRT-PCR experiment in the results and discussion section. In addition, we added where the colon cancer cells and normal colon cell ontained from in the materials and methods section.

Changes in the text: In page 16 line 233-230 and page 18 line 312-316, we modified the expression of qRT-PCR in the results and discussion section. And in page 5 line 153-155, we added that three groups of colon cancer cells (HCT-8, SW620, and SW480)

and one group of normal colon cell (HT-29) purchased from Cell Bank of the Typical Culture Preservation Committee of the Chinese Academy of Sciences (Shanghai, China).

Comment 6: The discussion section is too short to clarify the large amount of data obtained in this research.

Reply 6: We are sorry that we have not explained all the analysis results clearly. We strongly agree with your opinion that the discussion section should be expanded to enrich our research and clarify the data of our research. We think that our analysis of the results of the UCSC and Oncomine databases in the text is relatively brief, which may lead to difficulties in understanding. Therefore, we modified this part in the discussion section.

Changes in the text: We modified the analysis of the results of the UCSC and Oncomine databases in the discussion section. (see page 17-18, line 262-294)

Comment 7: High resolution of graphs should be provided.

Reply 7: Thank you for suggesting that we need to improve the clarity of the images in this article. To this end, we adjusted the arrangement of the images in figure 6, figure7, figure 9, and figure 10.

Changes in the text: We replaced the figures 6-7, and figures 9-10 in the text with our modified figures.

Once again, thank you very much for your comments and suggestions. We tried our best to improve the manuscript and made some changes in the manuscript. We appreciate for Editors and Reviewers' warm work earnestly, and hope that the correction will meet with approval.