Peer Review File

Article information: http://dx.doi.org/10.21037/tcr-20-2436.

Reviewer comments:

The incidence and mortality of lung cancer rank first among various malignant tumors, mainly for lacking clearly molecular typing and efficient individual therapy which greatly limited the therapeutic benefit of patients. Long non-coding RNAs (lncRNAs) have been demonstrated widely involve in tumor progressing, and been proved easy to detect for occupying majority in transcriptome. In the manuscript "A Six-long Noncoding RNA Model Predicts Prognosis in Lung Adenocarcinoma", authors analyzed the expression of all lncRNAs by WGCNA to determine prognosis related module.

Couple questions are required to be answered before accepted.

(1) There were several similar reports (Oncol Lett. 2020 Apr;19(4):2793-2800) and (Cell Physiol Biochem. 2017;42(5):1857-1869) in the PubMed. What is the novel idea in the paper? Please elaborate in the introduction.

Reply 1: Thank you very much for your question. Oncol Lett's article directly divided patients into high and low risk groups for model construction, and finally got 44 lncRNAs models. Because there are too many lncRNAs, the cost of clinical testing of this model is too high and the feasibility is too small. The Cell Physiol Biochem's article only studied the role of a single lncRNA SOX21-AS1 in the prognosis of lung adenocarcinoma. In this study, a 6lncRNAs model was finally determined by constructing a weighted co-expression network. The novelty of this study lies in the systematic analysis of the correlation between lncRNA and the prognosis of lung adenocarcinoma. At the same time, the clinical detection of the constructed 6lncRNA model is also feasible.

(2) There were several grammar errors in the text. Such as the first sentence of the background of abstract.

Reply 2: Thank you very much for the grammar error you found. We have modified it.

Changes in the text: we have modified our text as advised (see Page 2, line 3-4).

(3) In the introduction, please enrich the progress of the treatment for LUAD. Please supplement the research progress of the prognostic biomarkers for LUAD in the introduction.

Reply 3: Thank you very much for your question. By searching the literature, we have supplemented relevant research progress of lung cancer treatment in the introduction. And for the research progress of LUAD prognostic biomarkers, we have described in the third paragraph of the introduction.

Changes in the text: we have modified our text as advised (see Page 3, line 12-18).

(4) Why to focus on lncRNA in the paper?

Reply 4: Because about 70% of the human genome is transcribed into RNA, more than 95% of the transcripts are non-coding RNA. At the same time, lncRNA has fewer exons, so there is a higher evolutionary conservation. These characteristics make lncRNA easy to be detected in body fluids including blood and urine. For these reasons, we mainly focus on the role of lncRNA in tumor prognosis.

(5) What is the meaning of "WGCNA" in the introduction?

Reply 5: Thank you very much for your question. WGCNA is the abbreviation of Weighted Correlation Network analysis. We added the full name where WGCNA first appeared in the introduction.

Changes in the text: we have modified our text as advised (see Page 5, line 9).

(6) Why to determine the abline=650, β=5? Why not to validate the model by the real-world data in your hospital?

Reply 6: By clustering the samples, we can see that the samples are clearly separated by abline=650. Therefore, we choose abline=650 to eliminate samples (Figure 2A). Looking at Figure S1A, we can see that the abscissas of the two figures are both soft thresholds. The ordinate of the first picture is R-squared, and the ordinate of the second picture is the average of connectivity. We can see that when the soft threshold β=5, R-squared reaches 0.9 for the first time, and the average connectivity is close to 0, so 5 is determined as the soft threshold.

The problem you mentioned is exactly what we are doing. Because we did not collect enough patient samples from the hospital in the early stage, it cannot be used to verify the predictive effect of the 6 lncRNAs model on the prognosis of patients with lung adenocarcinoma. We are still continuing to collect samples, and we will use these samples to verify the reliability of this model in the future.

(6) The calculation formula of risk score was developed based on what?

Reply 7: The calculation formula of the risk score is developed based on the 123 lncRNAs of the red modules related to the prognosis obtained by weighted co-expression network analysis. In the training group, through univariate and multivariate Cox proportional hazards regression, we constructed a risk prediction model.

(7) What is the meaning of "(28 4142)" in the validation and analysis of risk score model?

Reply 8: Thank you very much for your question. (28 4142) represents references 28, 41 and 42.