

Peer Review File

Article information: <http://dx.doi.org/10.21037/tcr-22-2166>

Reviewer A

Comment 1: There are some minor English-related edits that could benefit the paper - any that I've noticed while reviewing I've marked in the attached PDF.

Reply 1: We have revised the sentences in the article.

Changes in the text: see Page 3, line 49; Page 3, line 52; Page 3, line 64; Page 4, line 73; Page 4, line 75; Page 4, line 79-80; Page 4, line 83; Page 5, line 101; Page 9, line 211; Page 10, line 226; Page 11, line 244-247; Page 12, line 270; Page 13, line 292-294

Comment 2: Line 173-180: I don't think you need to include all of this in the text since you are referring to a figure with this information. It makes it hard to read and doesn't really add useful data that you haven't included elsewhere.

Reply 2: We have deleted the descriptions of HR, CI, and P values.

Changes in the text: see Page 8, lines 168-169.

Comment 3: Line 263: It might be useful to define this since I don't think it is intuitive yet in the field. If it is defined earlier in the paper I missed it, but even re-defining it here might make it easier for researchers that look at the discussion first so they understand you're referring to regression analysis.

Reply 3: We have made a specific statement on LASSO.

Changes in the text: see Page 11, lines 251-252.

Comment 4: Line 298-299: I'm not exactly sure what is meant here. I think it is 'may have a similar effect on tumor development' but it is possible that the authors mean something different.

Reply 4: We have revised the sentence in the article. The details are as follows: "It seems that they share a similar biological process and may have a similar effect on tumor development. "

Changes in the text: see Page 13, line 286.

Comment 5: Line 308-309: I don't think you need this - it is not really a limitation of the study and I think doesn't need to be stated. It is true that the study needs to be replicated to have more support/evidence but that is true of any paper.

Reply 5: We have revised the sentence in the article. The details are as follows: "Finally, in further study, we need to increase the number of clinical samples to verify."

Changes in the text: see Page 13, lines 296-297.

Comment 6: Line 317-318: I don't understand how this last line fits in with the

conclusion, maybe an additional explanation is required.

Reply 6: We have revised the sentence in the article. The details are as follows: “According to the co-expression network of the 5 CENPs, 441 hub genes were screened and their main function is cell mitosis. Future research will focus on the detailed mechanisms of the CENP family.”

Changes in the text: see Page 14, lines 305-307.

Reviewer B

Comment 1: Line 32: ‘Results: There were 15 differentially CENP proteins expressed highly in LUAD.’ should be indicated in comparison to what.

Reply 1: We have revised the sentence in the article. The details are as follows: “Expression of 15 differentially CENP proteins is higher in LUAD when compared with normal lung tissues on mRNA level.”

Changes in the text: see Page 2, lines 32-33.

Comment 2: Line 37-38: ‘The qRT-PCR results showed that the 5 CENPs were upregulated in LUAD.’ Again it should be pointed out in comparison to what.

Reply 2: We have revised the sentence in the article. The details are as follows: “The qRT-PCR results showed that the 5 CENPs were upregulated in LUAD compared with normal lung tissues.”

Changes in the text: see Page 2, line 38.

Comment 3: Already in the Abstract, in the Methods section, one is struck by the small number of samples used to study expression at the RNA level in the tissue.

Reply 3: We use fresh tissue samples to extract RNA. Not all patients are willing to provide it, so it is difficult to obtain. We plan to collect more specimens to make tissue microarrays for further research.

Changes in the text: none

Comment 4: Line 49: ‘...patients are diagnosed when the disease is in the advanced or metastatic stage.’ Are the patients at a stage with metastasis, not at an advanced stage? Or do they mean locally advanced and metastatic lung cancer? It should be clarified what the Authors mean.

Reply 4: We have revised the sentence in the article. The details are as follows: “Most patients are diagnosed at an advanced stage.”

Changes in the text: see Page 3, line 50.

Comment 5: The Authors used interesting statistical methods.

Reply 5: Thanks for the comments.

Changes in the text: none.

Comment 6: Unfortunately, for the umpteenth time, it should be noted that there were very few samples for qPCR compared to the number of data downloaded from the databases. In addition, there is no demographic and clinical information about the

patients, except that they were diagnosed with resectable adenocarcinoma of the lung.
Reply 6: Patients are hospitalized for the first time for surgical treatment, and it will take several years to obtain their prognosis information. We plan to collect patient clinical information in the next in-depth study.

Changes in the text: none.

Comment 7: In silico testing is rather a way to validate the results obtained from wet lab results. Here the Authors used the reverse order.

Reply 7: We learned from the literature that some CENP family proteins have an impact on tumors, which prompted us to fully understand the role of this protein family in lung adenocarcinoma. Bioinformatics analysis is an easy way to achieve our research objectives.

Changes in the text: none.

Comment 8: Line 154-158: This is a repetition of the information from the Methods section. Is there a need to repeat it?

Reply 8: We have deleted “To investigate the expression differences of CENP family members in LUAD, we used TCGA and GTEx databases to compare the expression levels of CENP family members in LUAD tumors and normal tissues. Among them, 526 tumor tissues were obtained from the TCGA database, 288 of 347 normal tissues were obtained from GTEx, and 59 from TCGA.”

Changes in the text: see Page 7, line 154.

Comment 9: Line 246-277; 250, 253: The p-value notations should be standardized: p-values $p < 0.05$ so how many?, then $P < 1e-13$ and $P < 1e-07$. Perhaps it would be worth creating p-value ranges (e.g. $p < 0.01$, $p < 0.0001$, $p < 0.0001$, and so on) to make statistical significance a little clearer.

Reply 9: The definition of these P-values refers to other similar literature (PMC5668367, PMC3658166). We chose the accepted P-values for defining differential genes.

Changes in the text: none.

Comment 10: Line 265: ‘Differential 265 expression of the five CENP proteins was validated by qRT-PCR.’ Slightly unfortunate formulation, because it looks a bit as if the level of protein expression was validated by examining expression at the mRNA level (qRT-PCR). One might guess that if the authors write protein expression, they mean expression at the mRNA level because that's what they pulled from the databases, but protein is not a mRNA. There is this kind of misleading wording in the article. One would have to make it more precise is the phrase 'expression on mRNA level' and then there is no doubt whether they mean protein expression or mRNA expression. And it is necessary to return to the first point of this review: Line 32: 'Results: High expression of 15 CENP proteins was found in LUAD' Proteins or CENP expression at the mRNA level?

Reply 10: We have revised the sentence in the article. The details are as follows:

“Expression of 15 differentially CENP proteins is higher in LUAD when compared with normal lung tissues on mRNA level” and “Differential expression of the five CENP proteins was validated by qRT-PCR on mRNA level.”

Changes in the text: see Page 2, lines 32-33; Page 12, line 254.

Comment 11: Line 281-282: ‘In our study, similar results were founded’ probably should be ‘found’ instead founded, perhaps even: ‘results have been obtained.

Reply 11: We have revised the word in the article. The detail is as follows: “found”.

Changes in the text: see Page 12, line 270.

Comment 12: The article lacks clinicopathological information on the data extracted from the TCGA and GTEX databases. The authors performed analyses concerning age, sex, or T, N, and M stages, but they did not provide the size of each group. They do not state how many patients were in the early stage and how many were in the advanced stage. Possibly a table summarizing these data can be made?

Reply 12: We have created Table 2. In addition, we find annotation errors in Figure 4C. The number of “Risk” and “Age” is not correct. So we have re-annotated them. And the corresponding part of the text has been revised.

Changes in the text: see Page 8, line 183.

Comment 13: Some figures are not very legible. It should be corrected.

Reply 13: We have adjusted Table 1, Figure 3, Figure 4, Figure 6, and Figure 7.

Changes in the text: none.