

## Peer Review File

Article information: <http://dx.doi.org/10.21037/tcr-20-2110>

### Reviewer A

This article describes the identification of genes possibly involved in lung adenocarcinoma prognosis by integrated bioinformatical analysis. The authors analyzed publically available databases of lung adenocarcinoma gene expression data and several complementary bioinformatical techniques. Initially, differential gene expression analysis was performed. The candidate genes were further analyzed by gene ontology and protein-protein interaction techniques. The possible prognostic significance of the increased expression of the 10 genes that accomplished these criteria was studied. Finally, the genetic alteration of these genes in lung adenocarcinoma samples was determined.

These studies allowed the authors to identify 10 genes whose expression might have prognostic value for lung adenocarcinoma. The results presented are sound and support the conclusions of the article. The data are of general interest in cancer research.

The article could be improved in several aspects.

One of them is the title that remarks the importance of the expression of the CENPF gene. However, the authors identify 10 different genes that seem to have similar relevance, as the authors discuss in the article. A more general title such as "Identification of key candidate genes associated with prognosis...." might be more explanatory of the content of the article.

A general comment is that many verbs are not properly used and the manuscript should be carefully checked for grammar. A few examples are indicated as minor points.

Minor points:

1. Line 32, "by calculation" might be "by calculating"
2. Line 33, "and fund" could be "and found"
3. Line 36, "In these ten genes CENPF has" should be "Out of these ten genes CENPF had"
4. Line 38, "show" might be "showed that"
5. Line 93, "the data analysis was used" might be "for data analysis we used"
6. Line 97, "and identity" could be "and identification of"
7. Line 104, "which have" should be "which has"
8. Line 116 "use" could be "using"
9. Line 125, "four study" should be "four studies"
10. Line 182, "use" could be "using"
11. Line 206, "Recent years" might be "In recent years"
12. Line 208, "is still high" should be "are still high"
13. Line 225, "shown" could be "has shown" or "showed"
14. Line 249, "and identified" might be "and was identified"

15. Line 255, “also” could be “were also”
16. Line 260, “prote” probably is “protein”
17. Line 304, “is included” should be “are included”
18. Line 418, the meaning of “represent the gene number of enrichment” is not clear.

**Comment 1:** A more general title such as “Identification of key candidate genes associated with prognosis....” might be more explanatory of the content of the article.

**Reply 1:** Thanks very much for your constructive suggestion. According to your suggestion, we have modified the title to “Identification of key candidate genes associated with prognosis of Lung Adenocarcinoma by integrated bioinformatical analysis”.

**Changes in the text:** The title of our article (see page 1, line 1)

**Comment 2:** A general comment is that many verbs are not properly used and the manuscript should be carefully checked for grammar.

**Reply 2:** Thanks for your valuable comments. We have checked the manuscript carefully and correct the mistakes according to your suggestion.

**Changes in the text:** Shown in the manuscript.

#### **Reviewer B**

Despite vigorous bioinformatic analysis using three datasets, the authors do not provide clinically useful information. In addition, there are so many grammatic errors. The limitations of the study are described in the Conclusion section, Figure 5 contains figures of 18 genes while the legend describes 20 genes, and Figure 10 is not referred in the manuscript. The manuscript should be polished by English editing service and the main findings of the study should be discussed more comprehensively before publication.

**Comment 1:** The authors do not provide clinically useful information.

**Reply 1:** Thanks very much for your constructive suggestion. Lacking of clinical information is a limitation of GEO database. In our study, the GEO datasets GSE43458 only contain smoking status and pathological type, GSE32863 only contain pathological type, GSE10072 only contain sample information (such as sex, age, race, smoking status and pathological type). So, there was rare clinical information for us to use. In addition, we further explored the gene expression of *CENPF* in different subgroups of age, gender, race, smoking condition and cancer stage groups of LUAD patients in a TCGA database. Thanks for your suggestion, we realize that we should use clinical information in our future study to strengthen the reliability of our results.

**Changes in the text:**

**Comment 2:** There are so many grammatic errors.

**Reply 2:** Thanks for your constructive comments. We have checked the manuscript carefully and correct the mistakes.

**Changes in the text:** Shown in the manuscript.

**Comment 3:** The limitations of the study are described in the Conclusion section.

**Reply 3:** Thanks very much for your valuable suggestion. We have shifted the limitations in the discussion section.

**Changes in the text:** At the end of discussion section. (see page 11, line 297-300)

**Comment 4:** Figure 5 contains figures of 18 genes while the legend describes 20 genes, and Figure 10 is not referred in the manuscript.

**Reply 4:** Thank you for your correction. According to your suggestion, we have corrected “20 genes” to “18 genes” in Figure.5 legend. We're so sorry that we made a mistake to submit the Figure.10 legend, and we have deleted the Figure.10 legend. We showed the related miRNAs and genes of CENPF in figure 10, but we finally considered giving up putting this result in our article. Thanks to your help, we were able to avoid a serious mistake.

**Changes in the text:** The modification of Figure 5 legend (see page 15, line 455). The Figure 10 legend should be deleted.

**Comment 5:** The manuscript should be polished by English editing service and the main findings of the study should be discussed more comprehensively before publication.

**Reply 5:** Thanks for your constructive comments. We have improved the discussion section according to your request.

**Changes in the text:** Shown in the discussion (see page 10, line 288-296)