Peer Review File

Article information: https://dx.doi.org/10.21037/tcr-23-2175

Reviewer A

Reviewer comment 1: The authors use FPKM as a measure of gene expression, but FPKM is not suitable for the between-sample and group comparisons. Count normalization that is implemented in many R packages (DESeq2 and EdgeR) is much more appropriate doi.org/10.1186/s12967-021-02936-w FPKM or RPKM are both suitable only within-sample comparison if one likes to compare one gene to another gene WITHIN the same sample. Both FPKM and RPKM are NOT suitable for between-sample comparisons or DE analysis. This is my main concern as the primary analysis is affected by the unsuitable method of gene expression normalization.

Reply 1: Thank you very much for your important comments. Indeed, FPKM is not suitable for the between-sample and group comparisons. Therefore, we downloaded the gene expression counts data from the TCGA database and performed a subsequent analysis of these data. The main results of the subsequent counts-based data analysis remained consistent with the previous ones and did not show significantly different results. This indicates the reliability of the conclusions obtained from our analytical mining of public databases, and we remain grateful for your rigorous and constructive comments!

Changes in the text: Revised Figure 1A-1D, Revised Figure 2B-2H.

Reviewer comment 2: WNT/b-catenin signaling has been linked to OS in many previous studies, but the authors have not cited them at all. Here is one doi.org/10.7243/2052-7993-1-3 (https://www.hoajonline.com/genomics/2052-7993/1/3) based on whole transcriptome analysis and here is another doi.org/10.1186/s40246-014-0020-0 (https://humgenomics.biomedcentral.com/articles/10.1186/s40246-014-0020-0), based on comprehensive mutation screening and expression analysis. Both studies identified WNT signaling as a biomarker for the OS.

Reply 2: Thank you for the reminder that we did neglect to cite the existing studies about the Wnt/b-catenin signaling pathway playing an important role in osteosarcoma. Through reviewing the literature, we found that in osteosarcoma, the Wnt/β-catenin signaling pathway is closely

related to osteosarcoma progression and chemotherapy resistance, among others. Targeting this pathway can significantly inhibit the malignant progression of tumors and enhance the sensitivity of patients to chemotherapy. We have included citations to the relevant literature in the corresponding sections of the article.

Changes in the text: Page 3, lines 71-76.

Reviewer comment 3: Student t-test is not an appropriate statistical test for the RNA-seq based gene expression differences because these data are not normally distributed. RNAseq data fit to the negative binomial distribution and therefore student T-test is completely inappropriate test. Again, I suggest the authors to refer to other similar studies like this doi.org/10.3389/fgene.2017.00193 (https://www.frontiersin.org/articles/10.3389/fgene.2017.00193/full) and this doi.org/10.1177/1535370217736512 (https://journals.sagepub.com/doi/10.1177/1535370217736512). Bith of tehs studies use RNAseq data and apply correct statistical analysis and bioinformatics making their conclusions solid and reproducible.

Reply 3: We are very grateful for the critical and meaningful advice you have given. After studying the reference, you provided in detail and reviewing related materials, we found that indeed, most raw data obtained from RNA-seq high-throughput sequencing may not conform to a normal distribution, and it is not reasonable to use the student t-test for statistical analysis in this case. Therefore, we re-downloaded the count data and processed it with log2 (count+1). Next, we performed normality tests (Anderson-Darling test, Shapiro-Wilk test, and Kolmogorov-Smirnov test) on the log2(count+1) data in GraphPad Prism software, and the results showed that these log2 (count+1) processed data conformed to normal distribution. Based on this, we performed student t-tests on the log2(count+1)-processed data to verify whether NCAPG, MAZ, and E2F1 gene expression was differentially expressed between normal and tumor tissues, and between metastatic and non-metastatic groups. Although our previous analysis method was not reasonable, after we analyzed the data again from the beginning based on your comments, we came to the same conclusion as before. These conclusions are consistent with the results of subsequent cellular experiments and remain credible.

However, we would like to thank you again for your professional and constructive review comments! **Changes in the text**: Revised Figure 1A-1D, Revised Figure 2B-2H.

Reviewer comment 4: Statistical analysis requires quite a serious overhaul by the authors. First of all, the description of statistical analysis under the Methods section is too superficial. What are the sample sizes, and what is the power of the study? Figures 3,4 and 5 have error bars. How the authors did get them? How many replicates were in the study. How the replicates were defined - technical of biological replicates.

Reply 4: Thank you for your suggestion, and based on your comments, we have added a note in the Methods section. The sample sizes we used from the TCGA database are detailed in the corresponding places on the figures. The bar obtained from the cellular experiments performed in Figures 3,4, and 5 is the result of statistical analysis of the results obtained from three replicates of three different batches of samples (i.e., three biological replicates), not three technical replicates of samples obtained at one time (i.e., three technical replicates). This point has already been mentioned in the "Availability of data and material" section.
Thank you for your thorough review of our manuscript, and we are in awe of your professionalism!
Changes in the text: Page 7, lines 200-208. Figure1A-1F, Figure2B-2H.

Reviewer comment 5: Kaplan-Meier estimator is used correctly, but what is GEPIA? In addition, log-rank test is very commonly used to compare two K-M estimators, but regression analysis would be much more robust for clinical data.
Reply 5: Please excuse our oversight, the GEPIA tool we used in the manuscript is an online platform tool. GEPIA's full name is Gene Expression Profiling Interactive Analysis, which was developed by Prof. Zhang Zemin's lab at Peking University in 2017. This platform includes RNA sequencing data from 9,736 tumor tissues and 8,587 normal tissues from TCGA and GTEx databases, and mainly provides functions such as gene expression analysis, gene correlation analysis, survival analysis, similar gene prediction, and downscaling analysis.
Indeed, as you say, regression analysis has some advantages in handling clinical data. Undeniably, the log-rank test is very commonly used to compare two K-M estimators (https://molecular-cancer.biomedcentral.com/articles/10.1186/s12943- 022-01533-9). Since the default of the GEPIA

platform is a log-rank test to compare two K-M estimators, we also did not continue to use linear regression for our analysis. However, the comments you provided further expanded and deepened my understanding of regression analysis. I appreciate your comments!
Changes in the text: Page 7, lines 194-196.

Reviewer comment 6: The English language needs some editing. Some parts of the manuscript were difficult to understand.
Reply 6: In response to your comments, we have re-examined our manuscript carefully, cutting and revising inappropriate and unsuitable words and language. We hope this will enhance the reader's understanding of our manuscript, thank you. Strikethrough is used to remove unnecessary or redundant content. Text in red is added content.
Changes in the text: Strikethrough is used to remove unnecessary or redundant content. Text in red is added content.

Reviewer comment 7: Data should be made public and deposited in public repositories as per journal requirements.
Reply 7: Thanks to your reminder, we have uploaded the content covered in the manuscript following the magazine's requirements.
Once again, we appreciate the amount of time and effort you put into reviewing our manuscript.

Reviewer B

1. Figures
   - The citation of Figure 4F was **missing** in the text. Please check and revise.

   Reply: Thank you for your reminder and please forgive any mistakes made due to our negligence. At present, we have corrected the citation of Figure 4F.
   - Please provide Figure 1 and Figure 4-5 in **higher resolution** if possible.

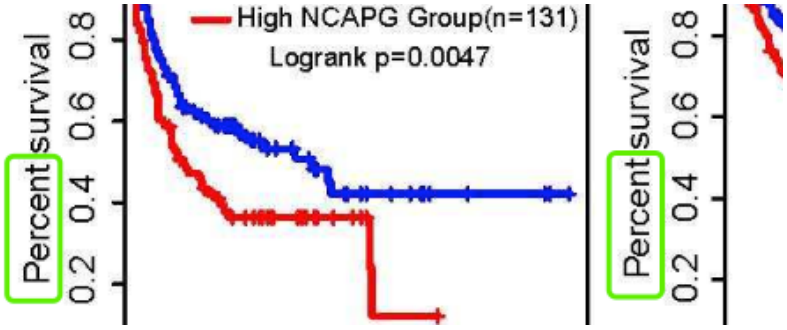
   Reply: To improve the reading experience of our readers, we have sent all the higher-resolution images as attachments.
   - Figure 1: Please indicate the meaning of *, ***, **** and ns.

Reply: We indicated the meaning of *, **, ***, ****, and ns in all figure legends, including Figure 1.

- Figure 2: Please indicate the meaning of **, **** and ns. **Please also check the symbols left unexplained in all your figures and revise.**

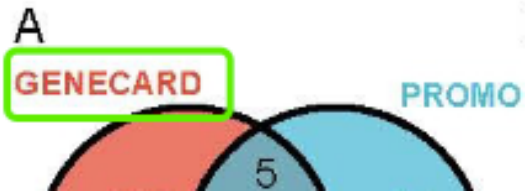
Reply: We indicated the meaning of *, **, ***, ****, and ns in all figure legends, including Figure 2. And, we added the missing symbol in Figure 2H that has not been labeled previously. We remain very grateful for your kind reminder.

- Please remove "Percent" from the y-axis in Figure 1E-F, since the rate is 0-1.



Reply: Unfortunately, our carelessness caused much trouble for your review. The Y-axis labeling of Figures 1E and 1F should be disease-free survival and overall survival, respectively, and we have corrected them in the figures, respectively.

- Figure 2A: the word is "JASPAR" in the legend. Please unify.



Reply: The "GENECARD" in Figure 2A is mislabeled, it should be "JASPAR", please forgive our oversight, thank you.

- Please indicate the staining/observation method and magnification in Figure 3I, Figure 4A-B and Figure 4E legend.

Reply: Thanks for your reminder, we've added the staining method and magnification to the corresponding legend.

- Figure 4I, 5E: Please check if the following word should be shscramble.

Reply: We have corrected the mislabeled “shscrsmble” to “shscramble” and emailed the revised figure as an attachment, and we admire your careful review of our manuscript.

2. The author’s name you mentioned in the main text is **inconsistent** with that in the reference list. Please check and revise.
    - Mechanistically, Chen et al. demonstrated that NCAPG could promote the progression of lung adenocarcinoma via the TGF-β signaling pathway[8].
    - 8. Wu, Y. *et al.* NCAPG promotes the progression of lung adenocarcinoma via the TGF-beta signaling pathway. *Cancer Cell Int* **21**, 443, doi:10.1186/s12935-021-02138-w (2021).

Reply: We are sorry for the confusion we caused you due to our carelessness. We have corrected the corresponding errors.

3. Please check if any citations of references should be added since you mentioned ***studies***.
    - Many ***studies*** have suggested that NCAPG is abnormally up-regulated in various malignant tumors.

Reply: Indeed, adding the appropriate citation to this place you mention would make our manuscript more rigorous, and we have added and supplemented it following your comments.