Peer Review File

Reviewer A

The authors used data from TCGA and conducted bioinformatic analyses to identify differentially expressed genes (DEGs) predictive of survival. Some comments and questions about the study:

1. The major limitation of this study is the lack of external validation. Only internal validation was conducted.

- Reply 1: Thank you for illustrating this point. We were limited by the low sample size and availability of datasets with overall survival data. Thus, we validated the models internally using the training testing split of 80:20 ratio.

2. Please explicitly state the "machine learning model evaluation score) in table 1. Please confirm if the score for KIF20A is really 68.

- Reply 2: Yes.

3. It is not clear to me how the machine learning model predicted survival and how was the accuracy of this prediction measured? Was survival treated as a continuous variable with censoring, or was it dichotomized?

- Reply 3: Good point. Yes, the overall survival was used in a dichotomized form to be classified using the random forest model. In addition, we also used both the overall survival time (continuous) and event (dichotomized) in the Cox model.

4. Table 4, the multivariable adjustment should also adjust for key clinical prognostic factors such as gender, age, stage, treatment, etc.

- Reply 4: The univariate and multivariate Cox model were used to evaluate the role of the transcriptomic markers alone, and whether their expression is associated with survival independently or all together. Thus, we did not include any clinical variables as the model were solely a transcriptomic model.

5. Figure 4, why TPX2 was not shown, and why was MYEOV shown?

- Reply 5: This was a typo and will be corrected.

- Changes in Text: Figure 4 labels have been edited.

Reviewer B

This is well designed and planned study. I have a few minor comments:

1) The list of genes: TPX9, KIF20A, CCNB2, and NCAPH should be described in Abstract.

   - Reply 1: Thank you for illustrating this point. The listed genes will be added to the abstract.

- **Changes in Text:** We modified the abstract to include the following genes (See page 2 line 11).

2) Sample sizes should be described in the abstract.

- **Reply 2:** Thank you for illustrating this point. The sample size will be added to the abstract.

- **Changes in Text**: The abstract has been modified to include the sample size (see page 2 line 7).

3) Key statistical results should be described in the abstract.

- **Reply 3:** The accuracy scores of the model are present in the abstract.

4) A paragraph describes limitations should be included in the discussion.

- **Reply 4:** There is a limitation paragraph in the discussion page 8 lines 18-23.