Peer Review File

Article information: https://dx.doi.org/10.21037/tcr-23-1299

Reviewer A

Comment: The study is interesting and the bioinformatic analyzes conducted on the HNSCC samples present in the databases (TGCA) seem correct. The strong limitation is that in addition to the TGCA and the GEO, another cohort of data validation is missing for a comparison. By virtue of this, I would add to the title at least if it is a bioinformatics analysis conducted on the TGCA and with a GEO validation set.

Reply: Thank you for this suggestion.We have made corresponding changes in the title to change the original title "CXCL family-related classification predicts prognosis and response to immunotherapy in patients with head and neck squamous cell carcinoma "to" CXCL family-related classification predicts prognosis and response to immunotherapy in patients with head and neck squamous cell carcinoma Based on TCGA and GEO databases"

Changes in the text: we have modified our text as advised (see Page 1, line 1-3)

Reviewer B

The authors have presented an interesting study looking at specific proteins in the TCGA database for HNSCC to stratify for prognosis.

However, the reader has several concerns:

Comment 1: On what basis was the decision to use CXCL genes to perform clustering of these tumors?

Reply 1: Thank you for this suggestion. The CXCL family gene set has not been studied in head and neck squamous cells. Currently, existing articles have only studied the impact of a single CXCL gene on the prognosis of head and neck squamous cell carcinoma patients , and there has been no systematic study on the CXCL family gene set. Based on this, we conducted cluster analysis and lasso regression analysis to investigate the impact of CXCL family set genes on the prognosis and immune cell infiltration of head and neck squamous cell carcinoma, in order to understand the role of the entire family genes on head and neck squamous cells

Comment 2: Was there a difference in clustering based on whether the sample was taken from the primary or nodal disease?

Reply 2: Thank you for this suggestion. In order to include more clinical sample information and make our clustering model more accurate, we included all patients with head and neck squamous cell carcinoma in the public database we downloaded for clustering analysis. Based on this, in the future, we will use our clustering model to analyze tumor patients with self-limiting and lymph node diseases collected by our hospital, in order to better promote our predictive model

Comment 3: The authors designed a prognostic score but used an unknown tool to determine if ICI would be useful, more detail needs to be provided in the manuscript on the tool that was used.

Reply 3: Thank you for this suggestion. We have provided a detailed description of the tool. The specific description is “To create a risk prediction model for patients with HNSCC, We used the R software package glmnet to integrate survival time, survival status, and gene expression data, and conducted regression analysis using the lasso box method. In addition, we also set a 10 fold cross validation to obtain the optimal model. We set the Lambda value to 0.0282483463007567 and ultimately obtained four genes.The model formula constructed is: RiskScore=0.0262776278846678 * CXCL1+0.022842939330026 * CXCL8-0.071226203344226 * CXCL13-0.0465917389070204 * CXCL17. We also evaluated the relationships between survival status, risk scores, and risk genes. With an increase in risk variables, the survival rate of patients dramatically decreased. CXCL1 and CXCL8 were risk factors, whereas CXCL13 and CXCL17 were protective factors.We used the R software package maxstat (Maximally selected rank statistics with sever p-value approximations version: 0.7-25) to calculate the optimal cutoff value for RiskScore. We set the minimum group sample size to be greater than 25% and the maximum group sample size to be less than 75%, and ultimately obtained the optimal cutoff value of -0.272929657980261. Based on this, patients were divided into high and low groups, and further analyzed the prognostic differences between the two groups using the survival function of the R software package, We evaluated the significance of prognostic differences between different groups of samples using the logrank test method, and ultimately observed significant prognostic differences (p=5.9e-5) and higher risk scores were linked to poorer prognoses in the patients with HNSCC

Changes in the text: we have modified our text as advised (see Page 12 -13, line 217-237)

Reviewer C

Comment : The study on the development of new classification to predict the prognosis of HNSCC is interesting and the amount of data to prove its significance is substantial. However, a few data was not well correlated with other data. For example, the authors suggest the importance of CXCL1 in determining risk scores, but the data on correlation of CXCL1 expression with survival (Fig. 6) was not represented. In addition, significance of clustering analysis is not prominent, though half of the data in this study was on the importance of C1 and C2.

Reply : Thank you for this suggestion. This performance is caused by different algorithms. In Figure 7, the Lasso regression algorithm used CXCL1, along with CXCL8, CXCL13, and CXCL17, as common factors that constitute the risk score to evaluate the prognosis of head and neck squamous cell carcinoma patients, and CXCL1 was used as a negative factor affecting the risk score. Figure 6 shows the use of univariate regression analysis to explore the impact of a single CXCL gene on the prognosis of head and neck patients. Although CXCL1 did not serve as a single factor affecting the prognosis of head and neck squamous cell carcinoma patients in univariate regression analysis, a P-value of 0.06 for CXCL1 still has the potential to be a negative factor affecting the prognosis of head and neck squamous cell carcinoma patients. However, the differences caused by these two algorithms will not affect the overall conclusion of the article. In addition, we are also continuously improving our model and planning to collect tumor patients from our hospital to analyze and verify prognosis through our clustering model.