**Peer Review File**

Article information: https://dx.doi.org/10.21037/gs-21-594

**Reviewer A**

**Comment 1: With TMB, they classified into two groups in HER2+ TCGA. They found 17 TMB-related immune genes and 6 TMB mutations-related genes. Of these, the 6 hub genes were utilized for their principal risk scoring model. However, they did not explain how they determined 6 hub genes with these 25 genes.**

**For 25 immune genes, they accounted for their prognostic ability. But readers cannot find the reason why they finally selected 6 core genes.**

**Reply 1:** Thank you for the kind remind. We used the Cox regression analysis to construct an immune- and mutant-related risk scoring system among 23 genes. We supplemented the results of Cox regression analysis as supplemental materials.

**Changes in the text:** *we performed univariate Cox regression analysis on each of the characteristics and for factors with a P value < 0.10 (Table S3), then we further constructed the final risk model by multivariate Cox regression analysis. (see Page 12, line 235 - 237)*

**Comment 2: In the risk scoring model, some markers lower the risk and other elevate. The authors should discuss clinical meaning for at least these 6 genes.**

**Reply 2:** We appreciate your instructive suggestion. We supplemented the characteristics of these 5 genes as Figure 5.

**Changes in the text:** *The heat map shows the expression of the 5 hub genes and clinicopathological variables in the high- and low-risk groups (Figure 5). We verified that the expression level of ERBB3, DNAH11 in high-risk group was higher than low-risk group, while MAPT, FABP5, PGR were the opposite. (see Page 13, line 246 - 249)*

**Comment 3: Among 17 immune-related genes, they included FGFR4. However, FGR was included in core 6 genes. What is correct?**

**Reply 3:** Apologize for the confusing information and many thanks to your attention. We corrected it into PGR in revised manuscript.

**Comment 4: In figure 4F, this is very critical. In the validated cohort with 6 core gene risk scoring system, they found that pCR rate was higher in the high risk group. However, in their TCGA cohort, the high risk patients showed an inferior outcome and low immune signature.**

**In HER2+ breast cancer, it is well know that the pCR patients have a much better survival and tend to have high TIL score. However, they did not provide any comments regarding this controversial finding.**

**Reply 4:** We highly agree with your instructive suggestion. We further analyzed the relationship between the risk scores and pCR of patients who did not receive trastuzumab, and we found no difference in pCR rate among patients receiving chemotherapy alone, indicating that the difference in prognosis between the high and low risk groups was independent of treatment. At the same time, we provided comments regarding this controversial finding of pCR and prognosis in the Discussion section, which was supplemented in the revised manuscript.

**Changes in the text:** *Some studies now showed that trastuzumab in combination with ICBs may benefit some patients. Su et al confirmed that trastuzumab can enhance the expression of PD-L1 (40). The PANACEA study estimated that in the PD-L1-positive subgroup, the objective response rate (ORR) of trastuzumab combined with pembrolizumab was 15% (41); similarly, the KATE2 study showed that the 1-year OS rate of patients in the atezolizumab group was better (42). However, our study suggested that patients in the high-risk group were able to achieve a higher pCR trend with trastuzumab plus chemotherapy, and this is inconsistent with poorer outcomes in the high-risk group. Our results may indicate that patients in the high-risk group are more sensitive to chemotherapy plus trastuzumab, which may be not related to the treatment of trastuzumab, but be associated with the change of immune microenvironment. pCR rate is not the only prognostic factors, as clinical trial of NeoALTTO, for example, suggested no difference in OS and DFS between patients in the pCR group and those in the non-PCR group (43). Thus, more studies are investigating changes in immune markers after trastuzumab treatment to explore more methods and strategies for combined therapy. (see Page 17, line 334 - Page 18, line 358)*

**Comment 5: Figure 2G, they may evaluate 24 immune cell signatures using ImmunCellAI database in the study population. However, they only displayed 6 signatures in their results. Maybe readers wonder the relationship between CD8 cell and TMB**

**Reply 5:** We highly agree with your warm suggestion. We detailed the supplemental materials section by adding more information regarding other immune cells and TMB in the revised manuscript.

**Changes in the text:** *However, no significant differences were observed in the infiltration levels of other immune cells between the two TMB groups (Figure S2). (see Page 11, line 213 - 215)*

**Comment 6: Figure 1D, 1E: X-axis labels should be provided.**

**Reply 6:** Thanks for your attention. We revised it in Figure 1D, 1E.

**Comment 7: Among 17 immune-related genes, they included FGFR4. However, FGR was included in core 6 genes. What is correct?**
**Reply 7:** Apologize for the confusing information and many thanks to your attention. We corrected it into PGR in revised manuscript.

**Comment 8: Last part with Figure 5 seems unnecessary.**
**Reply 8:** Thank you for your attention and warm advice. The function of these immunocytes and their relationship with prognosis have been gradually known. It seems unnecessary to include them in the figure again, and we have removed this part from our manuscript.

**Comment 9: Method: Kmplot.com part should be cited with the reference.**
**Reply 9:** Thanks for your suggestion. We have added it in M&M part.

**Reviewer B**
This study investigates the role of TMB for immune gene characteristics and clinical outcome, and presents a risk score model consisting of immune-related and mutated genes in HER2-positive breast cancer patients. The authors further show that the groups divided by the gene risk score are related to the degree and type of infiltrating lymphocytes.

The manuscript is well written and presents interesting and novel data, yet a couple of weak characteristics and potential pitfalls related to study design and methods should be addressed.

**Comment 1: The fact that this study is entirely based on in silico databases that suffer to various degrees from missing critical variables (e.g. treatment information, recurrence endpoints, short follow-up) makes it highly susceptible to inter-tumor heterogeneity and sample bias. Hence the clinical relevance of the 5 hub gene is unclear until independent validation in better-documented clinical trials is provided. This aspect must be elaborated in the discussion.**
**Reply 1:** Thanks for your kindly pointing out. We revised the limitation in the Discussion section.
**Changes in the text:** *There were some limitations to our study. First, this is a retrospective study, and the results should be further confirmed by prospective studies to exclude sample selection bias and inter-tumor heterogeneity. Additionally, the clinical relevance of the 5 hub genes needs more independent validation in better-documented clinical trials in the future. (see Page 18, line 366-371)*

**Comment 2: Patient inclusion/exclusion criteria are missing in the M&M section e.g. it is not clear why 63 patients with neoadjuvant trastuzumab treatment were selected in the validation (page 11)**

**Reply 2:** Thank you for your attention and warm advice. We have added it in M&M part.

**Changes in the text:** *In the GSE50948 dataset, we found that only 63 patients who had received a combination of trastuzumab and chemotherapy in neoadjuvant therapy, while the other patients received chemotherapy alone. Patient inclusion criteria: Immunohistochemistry (IHC) or fluorescence in situ hybridization (FISH) was used to determine the status of HER2 in breast cancer. If one of these was positive, we define these samples as HER2 + regardless of estrogen receptor (ER) or progesterone receptor (PR) status. Patient exclusion criteria: a) both IHC and FISH were negative or uncertain HER2 status; b) patients with incomplete clinicopathological information, such as TX stage (the primary tumor could not be assessed), NX stage (regional lymph node involvement could not be assessed), and MX stage (the metastatic status could not be assessed) in the TNM staging system. (see Page 7, line 112 - 132)*

**Comment 3: The use of average TMB for a cutoff appears arbitrary, as other approaches such as conditional inference tree classification or distribution modeling for the presence of clusters appear more statistically sound. The subsequent gene filterings depending on this cutoff may be biased in consequence.**

**Reply 3:** We appreciate your instructive suggestion. We calculated the cutoff values of TMB according to the literature of "*Thomas, Alexandra et al. Oncoimmunology vol. 7,10 e1490854. 30 Jul. 2018*", which had verified that use of average TMB for a cutoff value is an acceptable threshold choice by alternative cutpoint testing. However, this also needs to be verified by larger samples, which we will improve in future work.

**Comment 4: Prognostic value of TMB: was the association of high TMB with worse outcome confirmed in multivariate cox regression? This asked given its correlation with age and hormone receptor status**

**Reply 4:** Thanks for your attention. Multivariate cox regression analysis showed a trend between higher TMB and poorer prognosis, but unfortunately the difference was not significant ($P>0.05$). We look forward to the validation of a larger sample in later work.

**Comment 5: The construction of risk scores derived from individual cox hazard ratios are not clearly described in the Material & Methods, e.g page 11, lines 200-203 are lacking critical information.**

**Reply 5:** Thank you for the kind remind. We used the Cox regression analysis to construct an immune- and mutant-related risk scoring system among 23 genes. We

supplemented the results of Cox regression analysis as supplemental materials.
**Changes in the text:** *We performed univariate Cox regression analysis on each of the characteristics and for factors with a P value < 0.10 (Table S3), then we further constructed the final risk model by multivariate Cox regression analysis. (see Page 12, line 235 - 237)*

Minors
**Comment 6: Page 4, References 14 and 15 could be complemented by newer studies, e.g. immunogenic subtypes/subpopulations were even found in hormone-sensitive breast cancer (PMID: 33008814)**
**Reply 6:** Many thanks to your attention. We added it as Refs 16 in Instruction part.
**Changes in the text:** *Meanwhile, Werner et al suggested that BRCA-related DNA repair deficiency and suppressed tumor immune responses may be clinically relevant predictors of endocrine therapy complementing treatment options in subgroups of hormone-sensitive early breast cancer (16). (see Page 5, line 78 - 81)*

**Comment 7: Page 5, the refs regarding predictors of immune therapy response can be complemented by a mRNA expression based, pan-cancer predictor including TNBC, PMID: 28650338**
**Reply 7:** Many thanks to your attention. We added it as Refs 23 in Instruction part.
**Changes in the text:** *Currently, programmed cell death-1/ programmed cell death-ligand 1 (PD-1/PD-L1) expression on tumors (20), DNA mismatch-repair deficiency, neoantigen load (21), and tumor-infiltrating lymphocytes (22) and IFN-γ-related mRNA profile (23) are well-recognized molecular determinants. (see Page 5, line 90 - 94)*

**Comment 8: Page 6: the number of analyzed samples from TCGA and GEO should be given here**
**Reply 8:** Many thanks to your attention. We added it in M&M part.
**Changes in the text:** *We prepared somatic mutation data for 216 HER2+ BC from the "Masked Somatic Mutation" category in TCGA database via the GDC data portal (https://portal.gdc.cancer.gov/). 114 HER2+ samples from GSE50948 and 323 HER2+ samples from GSE96058 in The Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) were used as the validation datasets. (see Page 7, line 119 - 122)*

**Comment 9: The endpoint disease-free survival should be described**
**Reply 9:** Many thanks to your attention. We supplemented the definition of disease-free survival in M&M part.
**Changes in the text:** *The definition of disease-free survival: in cancer, the length of*

*time after primary treatment for a cancer ends that the patient survives without any signs or symptoms of that cancer. Also called DFS, relapse-free survival, and RFS. (see Page 8, line 146 - 149)*

**Comment 10: Table 2: two genes have been included in the risk score model, that were not significant. How is this justified?**

**Reply 10:** Thanks for your kindly pointing out. 5 signatures with $P < 0.10$ were screened by univariate Cox regression analysis, and the risk model was established by multivariate Cox regression analysis. Generally speaking, the lead-in inspection level is 0.05 and the rejected inspection level is 0.10. We set the lead-in inspection level at 0.10 to minimize the number of missing covariates. We can even set the lead-in inspection level at 0.20 when the study requirements are relatively loose. In our subsequent study, PGR and FABP5 were observed to be closely related to immune infiltration, and it would be a pity if we excluded them from the risk model at the beginning.

**Comment 11: Page 12, top: an AUC of 0.61 cannot be called "good prognostic"**

**Reply 11:** Many thanks to your attention. We corrected it into "relatively good prognostic" in revised manuscript.

**Reviewer C**

The authors conducted statistical analyses on the mutation profiles of patients with HER2+ breast cancer. The results showed that the tumor mutational burden is correlated with the overall survival rate, the risk of tumor recurrence, the hormone receptor status, and the patients' age. The authors also identified differentially expressed genes between the TMB-high and the TMB-low groups, which were used to develop risk scoring models to predict patients' 5-year overall survival status. Interestingly, the model showed that the patient group with low risk scores has a significantly higher survival rate.

My detailed comments on the article are as follows:

**Comment 1: In line 69-75, it may not be clear to the audience how the two cited studies led to the conclusion that immunotherapy is a viable treatment type for patients with HER2+ breast cancer. Rephrasing this part or adding more details would be helpful. For instance, what genes are found by Safonov et al. to be highly expressed in HER2+ BC and potentially lead to promising results of immunotherapy? How does FcγR3α affect the ADCC effect of trastuzumab (facilitate or inhibit)?**

**Reply 1:** Thank you for your attention and warm advice. We corrected it in the revised manuscript.

**Changes in the text:** *Safonov et al found that increased immune metagene expression associated significantly with lower clonal heterogeneity in all subtypes of BC and with a trend for lower overall mutation, neoantigen, and CNV loads in TNBC and HER2+ cancers. While in ER+ cancers, mutation load, neoantigen load, and CNV load weakly but positively associated with immune infiltration, which reached significance for overall mutation load only (15). ...Additionally, Musolino et al confirmed that the biological characteristics of Fc Fragment Of IgG Receptor IIIa (FCGR3A) can facilitate affect the efficacy of trastuzumab by enhancing the antibody-dependent cell-mediated cytotoxicity (ADCC) effect (17). (see Page 4, line 71 – Page 5, line 84)*

**Comment 2: In line 91-92 and 316-317, the authors mentioned that they identified biomarkers to predict the efficacy of immunotherapy, but no evidence was provided in the Results section, which shows only the prognostic value of the identified biomarkers. The differences between predictive and prognostic biomarkers need to be clarified.**

**Reply 2:** Apologize for the confusing information and appreciate your instructive suggestion. We corrected it in line 91-92, 316-317 and other confusion parts in the revised manuscript.

**Changes in the text:**

1. *Thus, many efforts have been devoted to identifying more immunotherapy targets and clarifying molecular mechanism of immunotherapy responsiveness (19). (see Page 5, line 88 - 89)*
2. *Our study suggested that TMB is associated with poor prognosis of HER2 positive breast cancer, and established a prognostic model for the combination of mutation-related genes and immune genes. Moreover, our results suggested that high-risk group are more likely to benefit from ICIs treatment, which provide a new strategy for evaluating the efficacy of immunotherapy and a new insight for finding HER2 positive breast cancer patients who are not sensitive to immunotherapy. (see Page 19, line 377 - 383)*

**Comment 3: The ROC curves in the manuscript do not seem to be correct. At least for binary classifiers based on the risk score, the true positive rate can only increase as the false positive rate increases. And the curves should always start at (0,0) and end at (1,1).**

**Reply 3:** Thank you for the kind remind. We recalibrated the ROC curve line thickness and set the diagonal to a gray dotted line to facilitate understanding.

I also recommend a few minor edits to give better clarity to the audience as follows:

**Comment 4: Line 31: "remained controversial" or "remained a controversy".**
**Reply 4:** Thank you for the kind remind. we corrected it in the revised manuscript.

**Comment 5: Line 136: since 178 patients were selected for the subsequent analyses, it may be more helpful to include only these patients in Table 1 instead of the total 216 patients.**
**Reply 5:** Thank you for the kind remind. we corrected it in the revised manuscript.

**Comment 6: There is a redundancy of conjunctive adverbs (particularly in the third paragraph of the Introduction) that should be removed or rephrased.**
**Reply 6:** Thank you for the kind remind. we corrected it in the revised manuscript.

**Comment 7: What statistical test was used to compare the two pCRs in Figure 4F?**
**Reply 7:** Thank you for the kind remind. We use the Chi-square test and supplemented the statistical test in M&M part.
**Changes in the text:** *The Chi-square test was used to compare composition ratios or rates between the two groups. (see Page 10, line 185 - 186)*