Peer Review File

Article information: https://dx.doi.org/10.21037/atm-22-277

First External Peer Review

Reviewer A:

At first glance it seemed a very interesting piece of research given the unmet need to tailor treatment in triple negative breast cancer. However, the manuscript has several major drawbacks listed below:

Major concerns

Comment 1: Does any of these patients received neoadjuvant treatment (tumor sizes of 12 cm reported on Table 1) and if they did: which proportion? what regimen? Do the authors considered using this as a variable to the model?

Response: Thank you for the sincere comment. We reviewed all the clinical data used in this study, and there were 18 patients (18/463, 3.9%) who had received neoadjuvant treatment before surgery per clinician choice. The case with a tumor size of 12 cm was diagnosed as carcinosarcoma; thus, she exempted neoadjuvant treatment. We also edited Table 1 to make it easier to understand. In addition, we highly agree with you that neoadjuvant treatment may influence the molecular features and clinical characteristics of tumors, for example, decreasing lymph node metastasis, which would be a confounding variable for model construction. Thus, we have removed the patients who received neoadjuvant treatment (n=18, 3.9%) from the final cohort (**Manuscript Tracked Changes, page 5, lines 98-99**). Meanwhile, we have edited the corresponding parts in the manuscript.

Comment 2: Lack discussion on how this approach will benefit clinical practice - reduce axillary dissection? tailor treatment? - nor about the feasibility - costs? time to get the results to clinical decision?

Response: Thank you for the suggestion. Currently, some patients with breast cancer are exempted from SLNB according to preoperative evaluation by various methods. This reduces the pain resulting from invasive procedures while retaining the risk of underestimation of axillary LNM. Therefore, more tools are needed to assess the risk of axillary lymph node metastasis and select the patient eligible for SLNB exemption more precisely. LN status in TNBC is difficult to predict using clinical factors only. We aim to supply more evidence for TNBC patients regarding whether they could be exempted from LN surgical excision (**Manuscript Tracked Changes, page 14, lines 294-307**).

Although the multi-omics model has robust predictive efficacy and is feasible in theory, it is difficult to implement in clinical practice. Neither genome nor transcriptome sequencing data can be obtained in a short time, and the cost is high. However, with the development of sequencing technology, it remains to be seen whether sequencing technology can be more easily accessible in

the future. Based on this model, we also plan to evaluate the level of these genes through other methods, such as qPCR or IHC (**Manuscript Tracked Changes, page 17, lines 370-376**).

Comment 3: English requires improvement - especially in the results and discussion sections.

Response: Thank you for your suggestion. According to the suggestion on the English writing of our manuscript, all authors have checked our language and employed AJE to polish the manuscript.

Reviewer B:

This is an overall interesting story with a large dataset.

Major concerns

Comment 1: I would recommend that the authors start with (i) delineating TNBC subtypes in the LN-pos and Ln-neg TNBC, e.g Burnstein et al., and Lehman et al., (ii) and then compare each TNBC subtype Ln-pos versus LN-neg. TNBC are too heterogenous and just comparing LN-pos versus LN-neg TNBC are too simplistic and will not lead to anything that can be used in the clinic, so clinical molecular factors have to be tightly matched.

Response: Thank you for the suggestion. In Lehmann's research, TNBC samples were divided into 7 subtypes. If we separated the cohort into several parts according to Lehmann's TNBC subtype, the number of patients in each subtype was too small to develop a robust model. We tried to construct a Lehmann subtype model based on Lehmann's subtype. In addition, we also built a FUSCCTNBC subtype model based on our previous publication (Jiang et al. Cancer Cell, 2019) (**Figure 4A, Supplementary Figure S6** and **Manuscript Tracked Changes, page 12, lines 286-291**). Finally, we attempted to add the information of TNBC subtypes into the multi-omics model, but it did not improve the performance of the multi-omics model.

Comment 2: The data analyses is clear, however the interpretation is superficial - how can the results be linked?

Response: Thank you for this comment. We have revised our manuscript according to your suggestion. In the discussion, we evaluated the performances of every model and analyzed their advantages and disadvantages. We discussed the molecular characteristics and biological features of the five predictors in the multi-omics model and attempted to explain their correlation with the risk of lymph node metastasis (**Manuscript Tracked Changes, page 14-17, lines 308-369**).

Comment 3: The data sets have to be made publicly available.

Response: All datasets have been published in our previous study (Jiang et al. Cancer Cell, 2019) and can be downloaded online. All sequence data and microarray data used in the study can be downloaded from the Sequence Read Archive (WES and RNA-seq; SRA: SRP157974) and NCBI Gene Expression Omnibus (OncoScan array; GEO: GSE118527) (**Manuscript Tracked Changes, page 5, lines 109-111**).

Reviewer C:

The authors attempt to find features to distinguish triple-negative breast cancer without axillary lymph node metastasis in order to exempt patients from sentinel lymph node biopsy and potentially having side effects from lymph node removal. This would provide benefit to patients in the treatment of their breast cancer. The authors use data collected from 463 TNBC patients and analyzed mutation, copy number, and gene expression data.

Major concerns

Comment 1: Overall, the method details are lacking which make it difficult to follow exactly how the data was analyzed. There are no methods detailing how WES, OncoScan, or RNAseq data was sequenced or processed for analysis. WES and OncoScan methods are only mentioned 1 time in the text.

Response: Thank you for the comment. Methods details have been reported in our previous publication (Jiang et al. Cancer Cell, 2019; see data generation in section Method Details). Here, we summarized the method of sequencing and processing and added it to the manuscript (**Manuscript Tracked Changes, page 5-7, lines 112-138**).

Comment 2: The statistical analysis is inappropriate for RNA-seq analysis. There is no mention of normalization and a student's t-test is not the best method for analysis. Multiple test correction should also be applied to the p-values.

Response: In this study, the differentially expressed mRNAs (DEMs) were analyzed by the package "limma" in R rather than Student's t test. The false discovery rate (FDR) correction was used to decrease false positive rates. We added the details in the manuscript and are sorry to make you confused due to our ambiguous statement before (**Manuscript Tracked Changes, page 7, lines 140-141**).

Comment 3: Details are lacking for the GO and KEGG analysis.

Response: We have revised our manuscript according to your suggestion. In this study, the differentially expressed mRNAs (DEMs) were analyzed by the package "limma" in R and met the standard of false discovery fate (FDR) < 0.05. Gene Ontology (GO) was used to investigate the biological processes of these DEMs. The metabolic pathways were analyzed with Kyoto Encyclopedia of Genes and Genomes (KEGG). GO and KEGG pathway analyses were both performed utilizing the R package "clusterProfiler" (**Manuscript Tracked Changes, page 7, lines 139-145**).

Comment 4: The integrated model was already narrowed down from each individual data set before applying it to the final model. This is not a full integrated approach. When applying integration methods, other factors may be of more importance than the individual analyses.

Response: Thanks for this comment. According to your suggestion, we tried to take all multi-omics factors into account in the integrated model to apply integration methods. In the fully integrated approach, we built a model with good performance (AUC of 0.83 in the training set,

AUC of 0.73 in the validation set), which further proved that predicting LNM using multi-omics was practicable (**Supplementary Figure S8** and **Manuscript Tracked Changes, page 13-14, lines 286-291**).

Comment 5: Other prediction models should also be considered besides LASSO regression, such as Random Forest and others.

Response: Thanks for this good suggestion. In fact, we have tried random forest to construct models before. It was performed by the R package "randomForest". However, the random forest model did not have good performance (AUC of 0.61 in the validation set). Thus, we finally chose LASSO regression. We declared that the figure followed was original and has not been published or appeared elsewhere.

Relevant results are shown as follows.

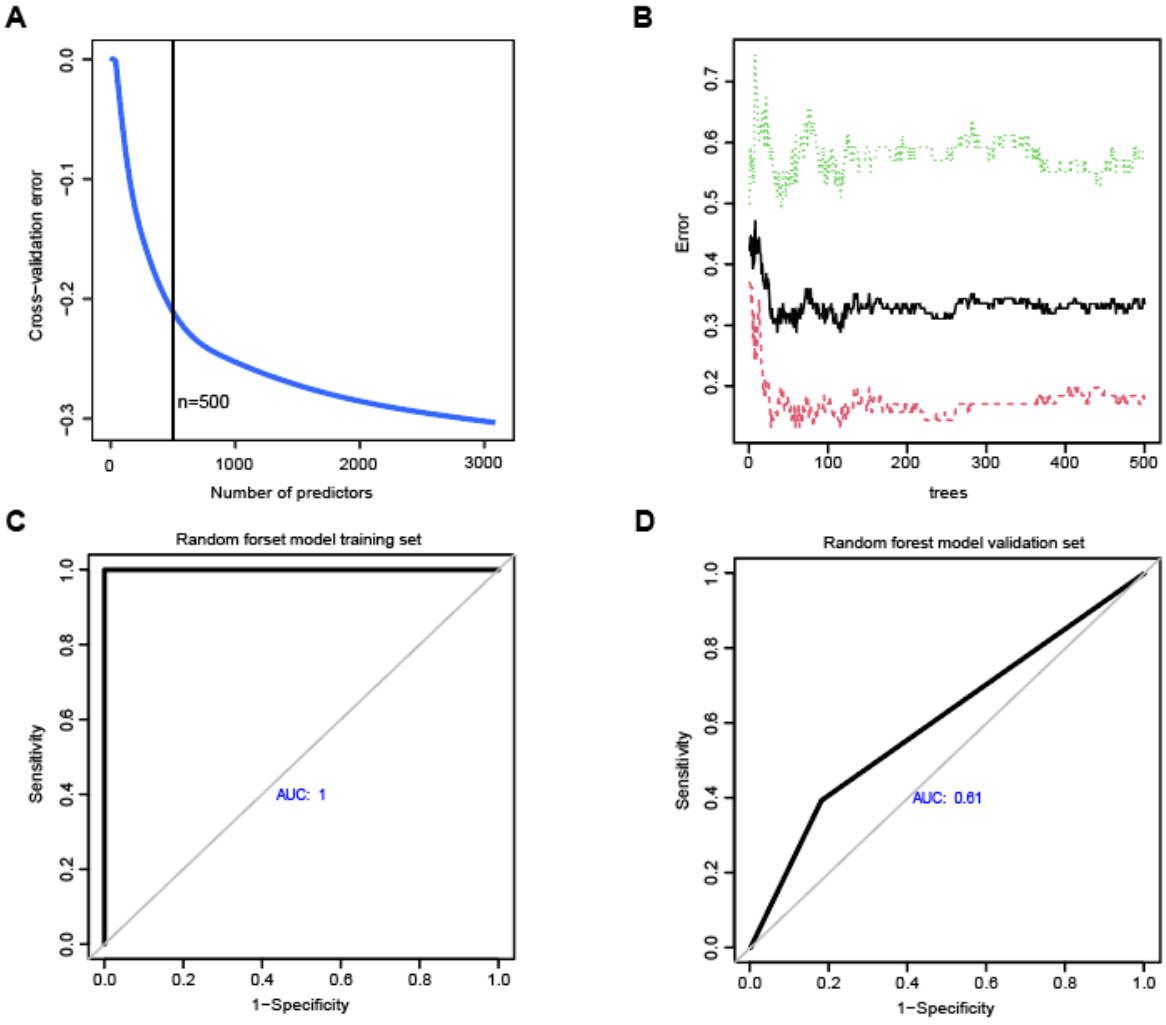

Figure R1. Details of the construction of the random forest model.

(A) Identify the appropriate number of predictors in the random forest model.

(B) Identify the appropriate number of trees in the random forest model.

(C) The AUC of the random forest model in the training set.

(D) The AUC of the random forest model in the validation set.

Second External Peer Review

Reviewer A:

The authors revised the manuscript well, but there are still some parts which need to be addressed before submission:

Major concerns

Comment 1: The normal samples need to be explained in more detail.

Response: Thank you for the sincere comment. In the research, the normal samples were used for RNA-sequencing. We add the detail "RNA sequencing was performed on 346 breast cancer tissues and 88 paired normal breast tissues" in the manuscript (**Manuscript Tracked Changes, page 7, lines 135-136**).

Comment 2: In line 190, 200 the statistical tests need to be added.

Response: Thank you for the suggestion. We have add the statistical test methods in the two sites (**Manuscript Tracked Changes, page 9, lines 194-195 & page 10, line 209**).

Comment 3: The result section still lacks scientific interpretation of the results and could be expanded.

Response: Thank you for your suggestion. We have further explained the results in the results section and discussion section (see in **Manuscript Tracked Changes**).