Peer Review File

Article information: https://dx.doi.org/10.21037/tcr-23-1123

Reviewer A

Comment 1: Since the risk score model in this study uses gene expression associated with brain metastasis, it should also indicate whether it can predict brain metastasis.

Reply 1: Firstly, we would like to express our gratitude for your feedback and attention to our risk scoring model. While our research primarily focuses on predicting the prognosis of breast cancer patients based on the expression of Brain Metastasis-Related Genes (BMRGs), considering whether the model can also predict the occurrence of brain metastasis is indeed an important question. In this study, our main objective is to develop a prognostic model that assesses the overall survival outcomes of breast cancer patients based on BMRGs. However, we acknowledge the clinical significance of extending the practicality of our model to predict the risk of brain metastasis development. To address this issue, further research and validation of our risk scoring model's predictive capability for brain metastasis occurrence are necessary. We have introduced this point in the Discussion section of the manuscript to underscore its significance. Once again, we appreciate your insightful suggestions.

Changes in the text: we have modified our text as advised (see Page 18, line 346-348)

Comment 2: P7, line 104. It is not clear why the BM related genes listed were selected.

Reply 2: Thank you for your feedback regarding the selection of Brain Metastasis-Related Genes (BMRGs), and indeed, the rationale behind the choice of these genes should be clarified. The BMRGs listed in our study were selected based on their previously reported associations with breast cancer brain metastasis. We conducted an extensive literature review and identified these genes from prior research and publications. According to reports, these genes are involved in various aspects of the brain metastasis process, including tumor cell migration, invasion, and interactions with

the brain microenvironment (PMID: 33387511). We recognize the importance of providing clear justifications for gene selection, which will ensure a better understanding of the scientific basis for their inclusion in our study. Once again, we appreciate your valuable insights.

Changes in the text: We have provided an explanation for the reviewer's concern and have not made any modifications to the text.

Comment 3: P12, line 203. It is not clear which of the three groups was compared for the P value of the log-rank test.

Reply 3: We appreciate your feedback on the clarity of the log-rank test comparisons. To provide further clarity on this matter, we have enhanced the details on page 12, line 205, to make it more explicit. The log-rank test was utilized to compare the survival outcomes among the three identified subtypes of breast cancer (C1, C2, and C3). This clarification aims to ensure that readers can readily comprehend the outcomes of this analysis. Once again, we sincerely appreciate your valuable insights and feedback.

Changes in the text: we added specific grouping information. (see Page 12, line 205)

Comment 4: P12, line204 'demonstrating significant variations in all identified BMRGs ($p<0.05$).' Statistical methods are not shown.

Reply 4: Firstly, we appreciate your feedback. We have conducted appropriate statistical tests, such as t-tests, on each individual gene to determine whether there are significant differences among subtypes. The significance level was set at $p < 0.05$. In the revised manuscript, we have provided statistical information regarding the assessment of the significance of gene expression variation(see Page 12, line 207). This clarification aims to ensure transparency and enable readers to understand the statistical basis for the observed variations in BMRGs among different subtypes. Once again, we express gratitude for your invaluable insights.

Changes in the text: The statistical method has been added.(see Page 12, line 207)

Comment 5: P16, line304. Multivariate analysis should include molecular subtype

(luminal, HER2, or triple-negative), which is an important prognostic factor.

Reply 5: We appreciate your suggestion to include certain molecular subtypes as variables in the multivariate analysis. Indeed, molecular subtypes are valuable prognostic factors in breast cancer. However, in this study, our primary focus lies in developing and validating a risk-based predictive model using Brain Metastasis-Related Genes (BMRGs) and their associations with clinical and pathological features. While we acknowledge the importance of molecular subtypes, our research aims to explore the prognostic utility of the risk scoring model, which integrates BMRGs, stage, and age. Given the scope and objectives of our study, we have not incorporated molecular subtypes into the multivariate analysis.

There is already a substantial body of literature on breast cancer subtypes (luminal, HER2, or triple-negative), including numerous basic and clinical research articles employing bioinformatics approaches, totaling up to 4232 articles, such as:

Chen YL, [Application of next-generation sequencing in detection of BRCA1/2 and homologous recombination repair pathway multi-genes germline mutation and correlation analysis]. Zhonghua Yu Fang Yi Xue Za Zhi. 2022 Mar 6;56(3):302-311. Chinese. doi: 10.3760/cma.j.cn112150-20211208-01132. PMID: 35381651.

Groheux D. [FDG-PET/CT for Primary Staging and Detection of Recurrence of Breast Cancer]. Semin Nucl Med. 2022 Sep;52(5):508-519. doi: 10.1053/j.semnuclmed.2022.05.001. Epub 2022 May 27. PMID: 3563697

Gupta A. [Association of lipid profile biomarkers with breast cancer by molecular subtype: analysis of the MEND study]. Sci Rep. 2022 Jun 23;12(1):10631. doi: 10.1038/s41598-022-13740-x. PMID: 35739205.

Thus, to avoid redundancy, we chose to abstain from studying this specific subtype.

However, we are well aware of the significance of molecular subtypes as prognostic factors. In our future research, we intend to collect samples from our institution, classify them into luminal, HER2, or triple-negative subtypes, and subsequently conduct high-throughput sequencing. We firmly believe that this approach will yield valuable insights. Presently, we have added this limitation of our study by supplementing it in the Discussion section (see Page 18, line 348-349). Thank you once again for your

invaluable insights.

Changes in the text: we have added this limitation of our study by supplementing it in the Discussion section.(see Page 18, line 348-349)

Comment 6 : Figure 5 G, H. Text is small and illegible.

Reply 6 : These two figures pertain to signaling pathways and are not closely related to the main focus of the article. As a result, we have removed them.

Changes in the text: Figure 5 G and Figure 5H have been removed (see Page 34 line549, and Page 39 line 584-586).

Reviewer B

1. There are some minor typographical errors that can be easily amended.

Reply: We would like to express our gratitude for your feedback and attention to our manuscript. We deeply apologize for the occurrence of these basic errors. We have rectified them as per suggestion.

Comment 1 : Line 3 and 16 - should read "A model for", not "A model of"

Reply 1 : "A model of" was changed to "A model for"

Changes in the text: we have modified our text as advised (see Page 1, line 3 and 16)

Comment 2 : Line 6 - please capitalise the first author's name "Jiangwei"

Reply 2: "jiangwei" was changed to "Jiangwei".

Changes in the text: we have modified our text as advised (see Page 1, line 6)

Comment 3 : Line 46 - "observances" should be "observations"

Reply 3: "observances" was changed to "observations"

Changes in the text: we have modified our text as advised (see Page3 , line 47)

Comment 4 : Line 63 - include space between "number" and "of". Similar error

seen in line 67, 92, 129, 133, 243, 291 and 327

Reply 4 : I apologize for the spacing issue caused by compatibility problems with the Word version. The problem has been resolved now.

Changes in the text: we have modified our text as advised (see line 64, 68, 94, 131, 135, 245, 293,and 329)

Comment 5: Line 89 - should read "alive and dead at follow up, respectively"

Reply 5 : Thank you for your advice. This error has been corrected.

Changes in the text: we have modified our text as advised (see Page 7, line 91)

Comment 6: In terms of terminology, line 215 - stage I-III cancer is synonymous with M0 disease, similarly stage IV cancer is synonymous with M1 disease. Therefore, this is tautology.

Reply 6 : The descriptions regarding M0 and M1 have been removed.

Changes in the text: we have modified our text as advised (see Page 12, line 217-218)

Comment 7: References: a more relevant article should be cited for prognosis of brain metastases in breast cancer patients (instead of ref. 31, which is related to gastric cancer)

Reply 7: The references have been replaced.

Changes in the text: The references have been replaced. (see Page22 , line 434)

2. Now, of greater concern:

Comment 1: The BMRGs is derived from a review article that examined genes expressed within brain metastases. The authors in the current study use publicly available datasets for breast cancer patients with early stage disease, i.e. no brain metastases, and from which the data is derived from the primary tumour. This may account for the unusual pattern of BMRG expression between the tumour and normal samples (as seen in Fig 2B, e.g. lack of significant difference in expression of TP53), which is different from what was reported by Morgan et al .

The authors need to acknowledge this fact as a confounding issue and will adversely impact on the interpretation of the predictive model.

Reply 1: Thank you for the reviewer's suggestion. We understand your concerns regarding the dataset, and we plan to address this issue with clarity. In our study, we utilized a public dataset from early-stage breast cancer patients. However, we cannot speculate whether these patients will experience brain metastasis in the future. We acknowledge the distinction between this dataset and the source of the Brain Metastasis-Related Genes (BMRGs), which primarily stem from a retrospective study on brain metastasis. This difference may lead to varying expression patterns of BMRGs between tumor and normal samples compared to the findings of Morgan. We recognize that this discrepancy could introduce a potential confounding factor that might impact the interpretation of our predictive model. Nonetheless, due to the support from existing literature, we still maintain that the expression of BMRGs in breast cancer is relevant to the occurrence and prognosis of brain metastasis. Although such differences exist, our objective is to explore the potential role of these genes in different breast cancer subtypes and prognoses, aiming to offer improved prognostic predictions for breast cancer patients. We have included this aspect as a limitation of our study in the Discussion section. Once again, we express our gratitude for your valuable feedback.

Changes in the text: This limitation has been incorporated into the Discussion section.(see Page18, line 343-346)

Comment 2: Line 101 quotes GSE41998 as a breast cancer immunotherapy dataset. However, on further examination this is a dataset for early stage breast cancer in patients who have received chemotherapy only. Therefore, the authors need to justify why they have included this dataset.

Reply 2: Thank you for bringing up this important question regarding our research. You have accurately pointed out the nature of the GSE41998 dataset, which primarily covers early-stage breast cancer patients undergoing chemotherapy. This is an insightful observation, and we greatly appreciate your attention to detail. We acknowledge that this dataset is centered around chemotherapy rather than immunotherapy. However, it

provides valuable clinical information about early-stage breast cancer patients, particularly those related to chemotherapy, which is crucial for understanding the immune status and treatment experiences of breast cancer patients. Although the dataset is focus on chemotherapy, we believe it can contribute vital background information about early-stage breast cancer patients to our study. We understand the concerns raised by the reviewer and thank you for the reminder. We have made appropriate revisions in the manuscript to more clearly articulate this point. Once again, we sincerely appreciate your valuable input.

Changes in the text: " immunotherapy " was changed to " chemotherapy ".(see Page 7, line103)

Comment 3: Lines 245-246 quote a newly derived gene signature which has been used in a brain metastasis risk score model based on 12 differentially expressed genes. How can the authors say that there is any relation to brain metastasis, when the datasets from which the signature has been derived does not inform whether the patients subsequently developed brain metastases? From what I can tell, the breast cancer population has been stratified by high/low risk of BMRGs, but not by actual risk of developing brain metastasis.

Reply 3: First of all, we want to express our gratitude to the reviewer for their attention to our research. You have correctly pointed out the source of the gene signature, which is derived from datasets of early-stage breast cancer. We acknowledge this fact. However, the aim of our study is to establish a clinical model that predicts the prognosis of breast cancer patients, incorporating these genes associated with brain metastasis. While we cannot directly determine whether patients will develop brain metastasis, these genes have been implicated in breast cancer research as being related to the occurrence of brain metastasis. Therefore, our model is designed to assess the impact of the gene signature on the overall prognosis of breast cancer patients, rather than directly predicting the risk of brain metastasis. We understand the concerns raised by the reviewer and will clearly articulate this point in the revised manuscript to ensure that readers correctly understand the objectives and limitations of our study. Thank you

again for your valuable suggestions.

Changes in the text: This limitation has been incorporated into the Discussion section.(see Page 18, line 346-348)

Comment 4: The authors used the IMvigor210 dataset to classify according to immunotherapy response. Note that IMvigor210 relates to a study investigating an immune checkpoint inhibitor in metastatic urothelial cancer, a very different patient population to metastatic breast cancer. In fact, immunotherapy has been shown to be effective in a small subset population of breast cancer patients, i.e. PD-L1 positive triple negative breast cancer. Therefore, it is unclear why the authors are choosing to investigate this parameter specifically and how it relates to the remainder of the manuscript, i.e. how is it relevant to brain metastasis management when it is not relevant to the primary tumour? This section of the manuscript distracts from the overall message of the paper.

Reply 4: Firstly, we would like to express our gratitude for the reviewer's insights. Regarding the issue with the dataset, we are willing to provide further clarification. You correctly pointed out that the IMvigor210 dataset pertains to metastatic urothelial carcinoma patients, not breast cancer. It is true that the efficacy of immunotherapy in breast cancer patients is typically confined to a small subset, namely PD-L1-positive triple-negative breast cancer. We acknowledge this viewpoint.

In this study, our objective is to establish a clinical model for predicting the prognosis of breast cancer patients, especially with risk of brain metastasis. Regrettably, there is a lack of immunotherapy-related datasets for breast cancer patients with brain metastasis. Therefore, we utilized the IMvigor210 dataset to explore the potential impact of the gene signature on immunotherapy response, aiming to extend our assessment of breast cancer patient prognosis. We seek to understand whether these genes associated with brain metastasis are also related to immunotherapy response.

Once again, we sincerely appreciate your valuable insights and feedback.

Changes in the text: We have provided an explanation for the reviewer's concern and have not made any modifications to the text.