# Peer Review File

**Article information: http://dx.doi.org/10.21037/gs-21-3**

**Reviewer A**

**Comment 1: This is a good work. However, the results and corresponding discussion are too early to accept. I strongly suggest authors to extend their experimental results and discussion sections.**

**Reply 1:** Thank you for your advice. We had enriched the results and discussion sections according to your suggestion.

**Changes in the text:** We added the survival data using the KM plotter database and added the OS curves in LncRNA LOC100505851 expression and survival. See Page 8, line 1-5. Page 9, line 7-10.

**Comment 2: There are many similar studies available in the literature. The author should compare their data with those of the existing data. A comparative discussion on predictive value of LncRNA LOC100505851 in breast cancer is required (I have seen some similar studies in the literature).**

**Reply 2:** Thank you very much for your nice suggestion. We have compared our results with some database and listed all the literature we could found associated with LOC100505851. There was an abstract reporting a gene fold change of LOC100505851 in aorta smooth muscle cells after iron TGFβ stimulation (Nephrology Dialysis Transplantation, Volume 29, Issue suppl_3, May 2014, Pages iii330–iii338, https://doi.org/10.1093/ndt/gfu161). However, LOC100505851 relevant result was not showed in the published article afterwards. Thus, we did not citate this article in our manuscript.

**Comment 3: In Figure 2- where are the error bars? Same is true for Fig 6.**

**Reply 3:** The results of Figure 2 are mixed with Figure 4 according to the opinion of ***Reviewer B; thus, we have deleted Figure 2.***

Figure 6A was generated by the STRING (https://string-db./org) while Figure 6B by the DAVID (https://david.ncifcrf.gov/) online tools, the X-axis represented the number of proteins involved in these pathways, so there were no error bars.

**Changes in the text:** We deleted Figure 2.

**Comment 4: The author should improve the English (Grammar) of the manuscript.**

**Reply 4:** Thank you for your comment. We have had our manuscript polished by a highly qualified native English-speaking editor. Attached is the certificate.

**Comment 5: The reference section is also very poor. There are many similar studies available in the literature. These papers should be acknowledged in the introduction and R/D parts of the paper.**

**Reply 5:** Thanks for your advice. We have gone through a systemic and detailed search of the articles associated with LOC100505851; however, we have not found any article focused on LOC100505851 itself. Most of the relevant information were from databases such as GeneCards, Ensemble, NONCODE, NCBI et al., and we had reported the results from GEPIA database in Supplementary Figure2&3.

**Reviewer B**

The study aims to evaluate the expression of LncRNA LOC100505851 in breast cancer in a setting from cases that received neoadjuvant chemotherapy, compared to adjacent breast tissues.

The authors used the qRT-PCR technique and correlated it with clinicopathological variables.

LncRNA LOC100505851 has been identified in low expression in breast tumors compared to adjacent tissue. The cases with high expression of LncRNA LOC100505851 were likely to achieve pCR, better relapse-free survival, and overall survival.

Below are some suggestions for changes and corrections, as well as questions about the manuscript.

**Comment 1: Introduction**

**Lines 2-5 – Provide current references on the percentage of advanced breast cancer. The cited references are from 2006 and 2007.**

**Reply 1:** Thank you for your suggestion. We have updated the percentage of advanced breast cancer from 2018.

**Changes in the text:** We rephrase the sentences as "Around 5% ~ 30% of the patients came with locally advanced disease[1, 2]" with new references. See Page 4, line 3-4. Page 15, line 14-17.

Results

**Comment 2: Figure 1 - Identify the total number of adjacent tissues, cancer and tumor subtypes in the figure bar, or in the legend. In Figure 1B, include the adjacent tissues in the evaluation. The authors perform a statistical analysis of the figure comparing the other molecular subtypes with the HER2 subtype, establishing this subtype as a control. However, the authors must first compare the expression of each subtype with the adjacent tissue (which is the control of the expression study) with their p values, and then correlate each tumor subtype with the HER2 subtype, showing their p values as shown in figure.**

**Reply 2:** Thanks for your advice. We have marked the specific number of patients providing cancerous and adjacent tissue as well as the number of patients in each subtype in Figure 1A and 1B. In addition, the expression of LOC100505851 in adjacent tissues were added as controls in Figure 1B, and the expression levels of LOC100505851 in each subgroup were compared with adjacent tissues first, and then each subgroup was compared with the HER2-positive subgroup.

**Changes in the text:** We changed the description of Figure 1 in the figure and in the

manuscript Page 6, line 3-6.

**Comment 3: The exploration of the data in Figure 2 is not clear, as these data are mixed with the data in Figure 4. The meaning of this analysis may not have been clear in the methodology. The authors show in Figure 2 that LOC100505851 is in high expression for tumors that have reached pCR; however, the data for the HER2 and basal subtype tumors, among other variables, need to be revealed in the figure.**

**Reply 3:** Thank you very much for your comment, we intended to highlight the subgroups with statistically significant difference in Figure 2, however, the data in Figure 2 was overlapped with Figure 4 indeed. And we had deleted Figure 2 according to your advice.

**Changes in the text:** We had deleted Figure 2 as well as the related description.

**Comment 4: "LncRNA LOC100505851 expression and pCR outcomes" - Provide in Figure 2 (or in the legend) and in the text the number of exact cases that represent the percentage (30.8%, 52%, 19.2%, 46.2%, 32.4%, 56.5%, 28.6 % and 56.5%).**

**Reply 4:** Thank you. According to your comment, we had deleted Figure 2. The exact pCR rate of the total patients was added in the text.

**Changes in the text:** We had deleted Figure 2 and the related descriptions. See Page 6, line 11-12.

**Comment 5: LncRNA LOC100505851 expression and survival - Line 14-17 - Provide global survival data for the cases analyzed, as in lines 10-13, or as in Figure 5, or if the authors wish, as a supplement.**

**Reply 5:** Thank you for your suggestion. When we were revising our manuscript, the data from the KM plotter database was newly updated on February 20, 2021. We have re-draw these figures using the latest data. Despite the consistent tendency, some figures have changed. The RFS curves were placed in Figure 4 and the OS figures were in Supplementary Figure 1 in our revised manuscript.

**Changes in the text:** We redraw these figures and the RFS curves were in Figure 4 and OS curves were in Supplementary figure 1. And we added the description of Figure 4 and Supplementary figure 1 in Page 7, line 20-21; Page 8, line 1-5.

**Comment 6: "LncRNA LOC100505851 expression and survival" - change the order of the figures. Figure 7 is cited before 6.**

**Reply 6:** We have changed the order of the figures and checked carefully of the order of each figure and table.

Materials and methods

**Comment 7: It is not clear which cases were considered to be adjacent. Are the 103 in the cohort who received neoadjuvant? Detailing better in "Patients and study design"'"**

**Reply 7:** Thank you for your suggestion. A hundred and six patients who ***did not have neoadjuvant therapy*** provided adjacent tissue. The details of the patients had been described in Page 10, Line 14-21.

**Comment 8: Cell culture and subcellular localization - provide which procedures (methodology, kits, etc.) are used to detect morphology, cell viability, DNA fingerprinting, isoenzimology, and mycoplasma.**

**Reply 8:** Thank you for your comment. Cell identification is a routine work in our laboratory, as there are several cell lines used in our laboratory currently. We examined the epithelial-cell-like morphology of ZR7530 cell lines under the microscope. The cell viability was detected by Edu proliferation kit (CAT# C00053; RiboBio, Shanghai). DNA fingerprinting was checked with western blot (WB) to see the expression of ER, PR and HER2 protein (CAT#8644, CAT#8757, CAT #2242; Cell Signaling Technolog, USA).

The detailed procedure of EDU and WB was shown as follows. Considering the lengths of the article we did not include this part in the manuscript.

WB: Cells were washed in PBS and suspended in 100 μL of RIPA buffer (Pierce, Dallas, TX, USA). Supernatant protein concentrations were determined using the BCA protein assay kit (Pierce). Supernatant samples containing 30 μg total protein were resolved by 10% or 12.5% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) depending on the molecular weights of the target proteins, and were transferred to immobilon-P polyvinylidene difluoride (PVDF) membranes (Millipore, Billerica, MA, USA) by electroblotting, and then probed with or anti-GAPDH (Cell Signaling Technology) antibodies. Membranes were incubated with horseradish peroxidase-conjugated secondary antibodies. Blots were developed using an ECL kit (Merck Millipore, Billerica, MA, USA).

EDU: Cells were cultured in medium containing 100μL 50μM EdU (catalog no. C00053; RiboBio) for 2h before fixation and Apollo staining (catalog no. C10371-1; RiboBio). Detection of the EdU signal was performed by fluorescent microscope.

**Changes in the text :** Page 11, Line 6-7.

**Comment 9: Statistical analysis - Line 19 and 20 - Although the reference was offered, put in brackets the positivity and negativity of the markers that the authors considered for molecular subclassification. Currently, there are other immunohistochemistry markers that differentiate a tumor from the basal type to a triple negative type.**

**Reply 9:** Thank you for your suggestion, we added details in the text and Table 4 for markers we used as molecular subclassification.

**Changes in the text: SEE** Page 13, Line 4-9 and Table 4.

Discussion

ok

**Reviewer C**

Sha et al. reported potential of LOC100505851 expression level as prognostic biomarker in breast cancer. The expression level of LOC100505851 was found to be down-regulated in tumor tissues as compared adjacent normal tissue in breast. In four subtypes, LOC100505851 expression is higher in Her2+ type. Except Her2+ type, pCR rates were high in high LOC100505851 expression groups. ROC analysis showed a better prediction of pCR including LOC100505851 expression. As expected, Kaplan-Meier analysis showed high expression of LOC100505851 showed a better prognosis, suggesting a tumor suppressive role of LOC100505851 in breast cancer. Finally, bioinformatics approaches predicted RNA binding proteins to LOC100505851.

Overall, the manuscript was clearly written and showed possibility of LOC100505851 for a prognostic marker. Several concerns need to be addressed. The comments are shown below.

**Comment 1: In Her2+ subtype, why is LOC100505851 highly expressed as compared with other subtypes? If possible, it would be better to describe the reason in the discussion.**

**Reply 1:** Thank you for your good question. This is also what we are trying to figure out. In the GEPIA database, there is a positive correlation between HER2 expression and LOC100505851 gene expression. Consistently in our study, LOC100505851 was also highest expressed in HER2 overexpression subtype. However, there is little research about the regulational role between HER2 and LOC100505851 until now. And we are trying to find the reason in the further study.

**Comment 2: In Figure 3, specificity and sensitivity need to be shown.**

**Reply 2:** Thank you, we have added the specificity and sensitivity in the current Figure 2.（which is Figure 3 in the original version）

**Changes in the text:** We have added the specificity and sensitivity in Figure 2. Also see Page 7, line 4-6 for description in the manuscript.

**Comment 3: In Figure 6A, GO enrichment analysis showed that nucleic acid binding is the top. However, in this study, starBase was used to predict RNA binding proteins, which could bind with LOC100505851. Using these RNA binding proteins, GO enrichment analysis was performed. In this case, "nucleic acid binding" should be on the top. If so, what is the meaning of this analysis?**

**Reply 3:** We extremely admire your advice. Thank you for pointing out this illogical argument circular for us, and we had deleted Figure 6A and Figure 6B.

**Changes in the text:** We deleted Figure 6A and 6B, consequently Figure 6C is renamed as Figure 6.

**Comment 4: In the text, Figure 6C is not described.**

**Reply 4:** Thank you for your comment. We added the description of Figure 6C (which

is currently named as Figure 6) in the result section.
Changes in the text: See Page 8, line 9-12.

**Comment 5: It is better to switch the number of figure 7 and figure 6 because figure 7 is shown first in the text.**
**Reply 5:** Thank you for your nice suggestion. We have changed the order of the figures and checked carefully of the order of other figures and tables.

**New references:**
1. Cardoso F, Spence D, Mertz S, et al. Global analysis of advanced/metastatic breast cancer: Decade report (2005-2015). Breast. 2018;39:131-138.
2. Cardoso F, Senkus E, Costa A, et al. 4th ESO-ESMO International Consensus Guidelines for Advanced Breast Cancer (ABC 4)†. Ann Oncol. 2018;29(8):1634-1657.