Peer Review File

Article information: https://dx.doi.org/10.21037/jgo-23-188

Reviewer A

This is an intriguing manuscript and concise. However, I have some questions for the authors.

1) I am concerned about the information regarding sample sizes that was mentioned in the discussion section. Specifically, I am worried about the adequacy of the sample size used in the study. In comparison to other studies on colorectal cancer patients, such as Uchiyama et al. (2015) (DOI: 10.4172/jpb.S5-005) and Wang et al. (2017) (DOI: 10.18632/oncotarget.19587), which utilized large sample sizes, the sample size in this study may not be sufficient.
Answer: Thank you very much for your comments. Here we used three CRC samples with three healthy tissue samples. In other studies like Uchiyama et al. (2015) (DOI: 10.4172/jpb.S5-005) and Wang et al. (2017) (DOI: 10.18632/oncotarget.19587), they detected serum peptidome profiling between CRC and normal controls. However, analysis data from serum can not directly represent the peptidome change occurred in cancer tissue. Our study focused on the tumor tissue which can also indicated possible therapeutic targets in CRC. As far as we searched from Pubmed, our study was the first study identifying peptidome profiling between CRC tissue and healthy tissue. However, we do not want to emphasize this in the manuscript as we know the limited sample size of our study. We also plan to enrolled more samples to detect peptidome profiling analysis in the future. Thank you very much.

2) In table 1, Why used only stage T3? Normally for screening method or detection should be include stage T1 or T2.
Reply: Thank you very much for your comments. Here we focused on the peptidomic changes between CRC tissue samples and normal controls. So, the more severe the disease, the greater the difference. Meanwhile, the overall survival time of patients in T3 stage patients is much lower compared with T2 and T1 patients with great significance. So, we chose T3 instead of T1 or T2 stage patients. Thank you very much.

3) Please describes the inclusion criteria and exclusion criteria to collect the tissue samples from patients. According to drug or some chemical compounds can be interfered in Bradford method.
Reply: Thank you very much for your comments. Here we described the inclusion and exclusion criteria for the enrolled samples.
Changes in the text: Methods, Page 4, Line 34-Page 5, Line 1-6.

4) Page 6, line 3; CAN or ACN?
Answer: Thank you very much. It is our fault to miss this error because of the Microsoft Word software. We have corrected this error. ACN indicates acetonitrile.

Changes in the text: Page 6, Line 1, Line 13

5) Page 7, line 17; I would like to clarify the term "intracellular peptides" that was used in your previous discussion. As whole tissue samples were extracted and analyzed for protein content, it is true that the samples would contain both intra- and extracellular peptides. However, the term "intracellular peptides" refers specifically to the subset of peptides that originate from within the cells, as opposed to those that are released into the extracellular environment. Therefore, when discussing the identification and analysis of intracellular peptides, we are referring to the subset of peptides that originate from within the cells and are present in the whole tissue samples.
Reply: Thank you very much for your nice comments. It gives us a great explanation for the intra- and extracellular peptides. Here we extracted all peptides from tissues including intra- and extracellular peptides. So, we should not only mention the intracellular peptides in our study. Here we have revised this in our manuscript.
Changes in the text: Page 7, Line 27.

Reviewer B

In this study, a comparative peptidomic profiling was analyzed in 3 CRC tissue samples and 3 adjacent intestinal epithelial tissue samples by liquid chromatography-tandem mass spectrometry (LC-MS/MS). The manuscript is straightforward, well written, and concise and has clear results. Definitely deserves to be published and is a valuable contribution to the "Journal of Gastrointestinal Oncology". Some minor comments need to be addressed before publication.

[1] "INTRODUCTION", Lines 5-6:
"However, each treatment has its own indications, benefits, side effects, and potential risks.".
The authors should make a comment about the population of elderly patients. Even though older patients are more prone to severe postoperative complications, there is no consensus that age affects survival outcomes. The prognosis of older patients may be confounded by differences in stage at presentation, tumor site, preexisting comorbidities, and type of treatment received.
Recommended reference: Osseis M, et al. The Developing Story of Predictive Biomarkers in Colorectal Cancer. J Pers Med. 2022;12(9):1534.
Answer: Thank you very much for your comments. Here we have made a comment concerning the population of elderly patients.
Changes in the text: Page 4, Line 7-9.

[2] "INTRODUCTION", Lines 12-13:
"However, there are still no reliable biomarkers for the diagnosis of CRC.".
A comment should specifically be made about miRNAs that have been found to induce chemoresistance. Indeed, FOLFOX-resistance in advanced CRC is significantly associated with upregulation and downregulation of several serum miRNAs. The differentiation of

FOLFOX-resistant from FOLFOX responsive patients by serum miR-19a had a reported sensitivity and specificity of 66.7 and 63.9%, respectively. In terms of treatment response to anti-EGFR inhibitors in metastatic CRC, overexpression of miR-31, miR-100, miR-125b, and downregulation of miR-7 is correlated with resistance to cetuximab, respectively.
Recommended reference: Boussios S, et al. The Developing Story of Predictive Biomarkers in Colorectal Cancer. J Pers Med. 2019;9:12.
Answer: Thank you very much for your comment. It is indeed that many predictive biomarkers concering CRC is emerging. miRNAs are very promising, however we still need more large study to confirm the results as potential biomarkers.
Changes in the text: Page 4, Line 14-16.

[3] "Discussion", Lines 13-15:
"Peptidomics, as an emerging branch of proteomics aimed at producing protein fragments (20), could be used to screen for amounts of disease biomarkers from serum (21), and tissues (22).".
At that point, the authors should mention the recommended screening for defective, DNA mismatch repair, which includes immunohistochemistry (IHC) and/or MSI test. However, there are challenges in distilling the biological and technical heterogeneity of MSI testing down to usable data. It has been reported in the literature that IHC testing of the mismatch repair machinery may give different results for a given germline mutation and has been suggested that this may be due to somatic mutations.
Recommended reference: Adeleke S, et al. Microsatellite instability testing in colorectal patients with Lynch syndrome: lessons learned from a case report and how to avoid such pitfalls. Per Med. 2022;19(4):277-286.
Answer: Thank you very much for your nice comment. Here we have mentioned the current screening methods in tissue or serum. We have revised this part into the manuscript.
Changes in the text: Page 10, Line 25-27.

Reviewer C

The article can be accepted after further revision. Also, too many questions remained unanswered. My first impression while reading the article was that it was written hastily without giving much thought to comprehension and data sufficiency.

General observation.
1. Overall, the manuscript is written very poorly, making most of the things incomprehensible and obscuring the meanings.
Reply: Thank you very much for your comments. Here we revised this manuscript with the help of a native speaker to make it clearer to read. Thank you very much.

2. There is no logical sequence in the workflow.
Reply: Thank you very much. Here we conducted a standard analysis regarding LC-MS/MS data.
First, we identified the 59 peptides derived from 38 precursor proteins had significantly

different expression.
Secondly, we analyzed the characterization and cleavage site of these peptides.
Next, GO and KEGG pathway analysis of the precursor proteins of the differentially expressed peptides were conducted.
Then, 10 top variable proteins were investigated in the Web of Science database.
After that, STRING analysis was performed in these 38 proteins.
Finally, the possible survival analysis were considered into the study. So we think our analysis has a logical sequence.

Critical observation

1. Section related to GO and pathway analysis: The authors have proceeded with GO and pathway analysis without any consideration. The pathways related to downregulated genes will be weekly functioning, and pathways associated with upregulating genes will be over-functioning. The same is valid for GO. If the objective of the study is to find the responsible perturbed pathways or ontology groups, then downregulated and upregulated genes should have been considered separately.
Reply: Thank you very much. Here we identified the 59 peptides derived from 38 precursor proteins had significantly different expression (FC >2, P<0.05), including 25 up-regulated peptides and 34 down-regulated peptides between CRC and adjacent epithelial tissues (Figure 1A). Normally we enrolled all the significantly different expressed proteins into GO and KEGG pathway analysis.

2. Whether the mentioned enriched pathways and GO terms are for over or underrepresented genes - is not mentioned.
Reply: Thank you very much. We conducted GO and KEGG pathway analysis with all the significantly different expressed proteins. Those proteins without difference should not be considered.

3. In entire articles, in the majority of the analysis part, which parameters are taken are not given. Simply a generalized term is used.
Reply: Thank you very much. Here we used the determined 38 precursor proteins with 18 up-regulated and 20 down-regulated targets classified in Table 2 into analysis.

4. The data is highly biased, and only a particular age group and gender are considered for analysis. To which stage the sample belongs is not specified. the analysis lacks information regarding other pathological parameters. In fact, some of the shortcomings are explained by the authors themselves.
Reply: Thank you very much for your comments. Here we focused on the peptidomic changes between CRC tissue samples and normal controls in T3 stage patients. The more severe the disease, the greater the difference. So, we want to find these proteins with the largest difference.

5. One analysis is disjoined by another secondary research. The objectives that the authors are trying to achieve are not comprehensible (rather they are confusing) in the present article.
Reply: Thank you very much. We have updated this revision with a proper workflow in a logical

way. Our manuscript is a typical data analysis using bioinformation methods.

6. String section: which parameters were taken is not given. The links between the nodes are of various shapes and color – what do they refer to?
Reply: Thank you very much. Here we used the determined 38 precursor proteins with 18 up-regulated and 20 down-regulated targets classified in Table 2 into analysis. The STRING, is the Search Tool for the Retrieval of Interacting Proteins. The Node means the protein, and edge means the interaction between two proteins. Here different color means different type of proteins like colored nodes indicates query proteins and first shell of interactors while white nodes indicate second shell of interactors.

7. What about the survival analysis of other top proteins that are differentially expressed?
Reply: Thank you very much for your comments. Here we focused on the top proteins with our interests to do survival analysis. So, in the manuscript, we just mentioned ALB and HIST2H2AA3 with their potential functions in colon cancer to be elucidated in the future.

8. Role of protein modification is not considered.
Reply: Thank you very much for your comments. Here we only detect peptides from the CRC and the adjacent epithelial tissue (control groups) by LC-MS/MS. So we did not mention the role of protein modification in this manuscript. We plan to do this analysis in later study. Thank you very much.

9. The data from LCMS/MS should be submitted to a database for peer review and public review.
Reply: This mass spectrometry proteomics data has been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD021746.

Recommendation
1. Use an English editing scientific/academic writer's help or English Editing Service.
Reply: Thank you very much. We have updated this revision with the help of an English native speaker.

2. Rewrite the manuscript with proper workflow and with each step in a logical flow.
Reply: Thank you very much. We have updated this revision with a proper workflow in a logical way.

3. Correction of general and critical observation.
Reply: Thank you very much. We have updated this revision with the help of an English native speaker.

4. Add more data only relevant to the topic in consideration.
Reply: Thank you very much. We have updated this revision.