Aug 09, 2020

Number invitation ID: 03725560
Manuscript NO: 57895
Manuscript Type: Basic Study

Professor Subrata Ghosh and Andrzej S Tarnawski
Editor-in-Chief
World Journal of Gastroenterology

Dear Editor,

We would like to thank the editor and all the reviewers of *World Journal of Gastroenterology*, for taking the time to review our article. We also appreciate the valuable comments from the reviewers. Accordingly, we have made the necessary corrections and clarifications in the manuscript after going over the reviewers' comments. The changes made in response to reviewers' comments are highlighted in red in the manuscript. Below we specify what we have done in response to each of the reviewers' comments. We hope the revised manuscript meets the requirements of your journal for publication. We thank the editor and the referees of *World Journal of Gastroenterology*, once again for the constructive review of our paper.
Sincerely,

Sung Hak Lee, MD, PhD, Associate Professor
Department of Hospital Pathology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea
222 Banpo-Daero, Seocho-Gu, Seoul 06591, Korea
Telephone: +82-2-2258-1617
E-mail: hakjjang@catholic.ac.kr

Reviewer #1:

1. Introduction and Discussion of the manuscript is too long. Need to reduce it. In the last paragraph of the Introduction section authors are describing the methods which is not needed.

We thank the reviewer for the thoughtful comments. As the reviewer recommended Introduction and Discussion sections were revised. Furthermore, methodological description in the Introduction was removed. The word counts for the Introduction section decreased from 626 to 490.

2. Methodology section should be crisp and focus only on how the study and experiments were designed and done.

We thank the reviewer for the thoughtful comments. As the reviewer recommended Methodology section was revised to be more concise. The word counts decreased from 1174 to 990.

3. Results section consists of too much redundancy, and repetition of methodology which is also present in discussion.

We thank the reviewer for the thoughtful comments. As the reviewer recommended Results section was revised to be more concise. The word counts decreased from 1099 to 997.

4. Authors need to specify the various limitation of the study, especially when dealing with two different sample - frozen and Formalin fixed.

We thank the reviewer for the thoughtful comments. As the reviewer recommended the limitation of the study was additionally discussed in the Discussion section. (Related manuscripts are as follows)
(Page 18, Line: 23)
Limitations also exist for the deep learning-based tissue classifiers. One of the limitations is the sensitive nature of deep learning to minute differences in the datasets. Because of the sensitive nature, classifiers applied to very subtly different conditions should be separately built. For example, classifiers for the frozen and FFPE tissues should be separately trained for the same tasks. It requires additional data collection and training overload. In clinical practice, pathologists should take an additional step to determine the kind of classifiers that should be applied for a specific specimen. It is currently inevitable to separately build classifiers to support various real-world tasks in the pathology laboratories. Therefore, manual selection of appropriate classifiers for target tasks is a necessary step that can limit the fully automated adoption of deep learning-based classifiers in the pathology workflows.

5. There is a need of language sanitation of the manuscript by a native speaker.

We thank the reviewer for the thoughtful comments. We've got English editing from native English.

Reviewer #2:

1. The benefits of this methods and its clinical meanings are not well discussed in the present article. The authors mentioned this methods may help to get more accurate molecular tests since this may avoid/balance the tumor heterogeneity. However, highly doubt a whole slide image review will be able to address this issue. Then if we are using representative slides from different location of the tumor, it will add the time and costs for this process.

We thank the reviewer for the thoughtful comments. Basically, the target regions of molecular tests are manually demarcated by pathologists on H&E slides based on tumor cell contents. Then, tissues adjacent to the demarcated regions are scraped for molecular tests. We suggest if the regions with high mutational status can be automatically demarcated by the deep learning-based tissue classifier, the quality of molecular tests can be improved. Furthermore, molecular tests with high spatial specificity can be targeted to regions of different mutational status depending on the purposes of the tests. We discussed the issue and the added discussion is described in the answer to the next question.

2. More importantly, the authors used the mutation identifier for each mutation including APC, KRAS, PIK3CA, SMAD4, and TP53. However, in the real world, these mutations are so commonly seen in colorectal cancer and a lot of patients have a combination of these mutation instead of only one mutation, how to use the current identifier to predict mutation in these patients will be a challenge.

We thank the reviewer for the thoughtful comments. It is possible to identify the mutational status of all five genes in a slide as demonstrated in newly added Supplementary Fig. S3. The representative tissue images showed that a slide can have regions with different mutational status. Our method can help to visually demarcate the heterogeneous regions. Furthermore, overall mutational burden can be recognized with the overlaid probability map of mutation prediction results for all five genes. It may not be easy to obtain this kind of information without the help of deep learning. This is described in the Discussion section.

(Page 18, Line: 14)

For example, Supplementary Figure S3 presents the heatmaps for the mutational status of all five genes in a TCGA frozen tissue slide, demonstrating how different regions of a slide can have different mutational statuses. When an overlaid probability map of mutation was drawn, areas with low and high mutational statuses can be recognized. It may not be easy to obtain this kind of information without the help of deep learning. Hence, molecular tests with high spatial specificity can be targeted to specific regions depending on the purpose of the tests.

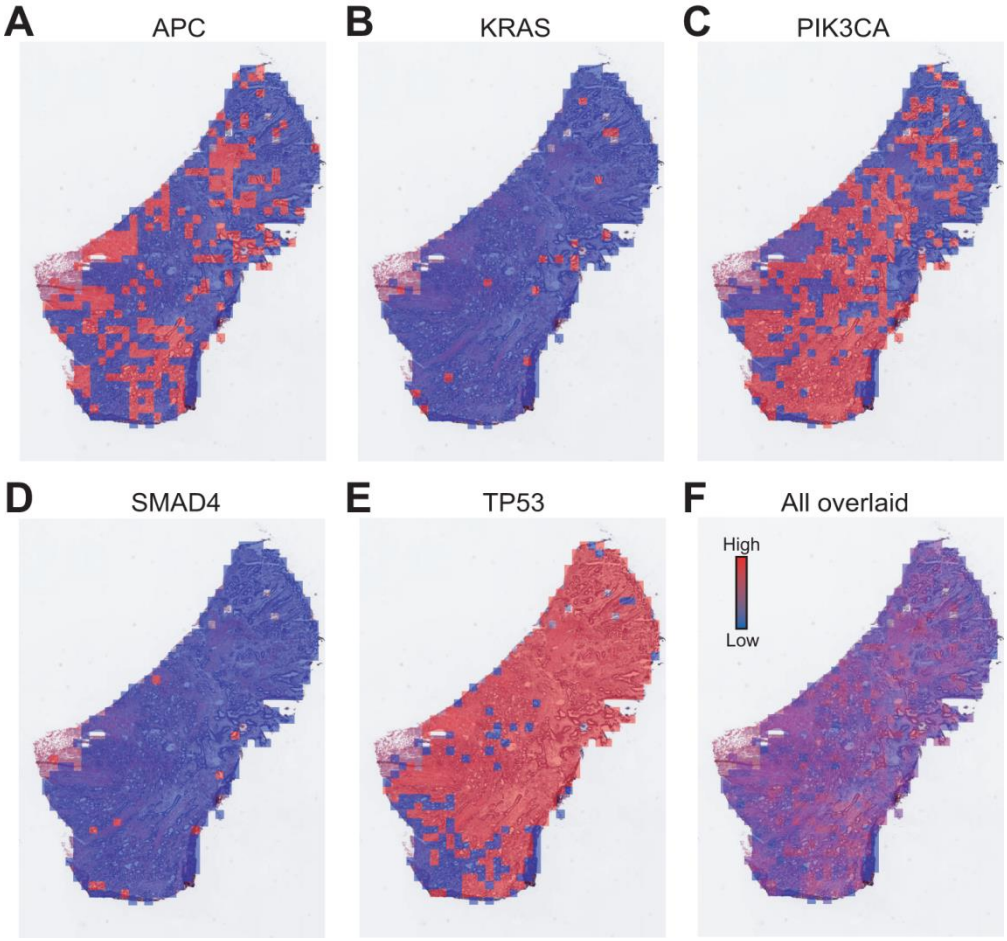

Fig. S3

3. It will be interesting to do a head-to-head comparison between the mutation panel test (molecular test) and this method. It might be reasonable to consider add a table to summarize the pros and cons for each test and may also emphasize the complementary parts of the tests.

We thank the reviewer for the thoughtful comments. We discussed the pros and cons of each method and added a table. (Related manuscripts are as follows)

(Page 19, Line: 6)

In the current study, we used the high-throughput cancer panel to identify mutations in CRC tissues of the SMH dataset. This panel test approach makes it possible to identify diverse clinically actionable

mutations in a single assay. However, it is quite expensive to prepare the equipment necessary to perform the test and to save a large number of data generated. This study demonstrated that a deep learning-based method could be a useful and effective tool for the prediction of actionable mutations from CRC WSIs. However, the interpretation of decision made by the deep learning-based classifier is unclear because of the black box nature of deep learning and should be further studied. Besides this aspect, the advantages and disadvantages between the mutation panel test (molecular test) and deep learning method were described in Table 1.

Table 1

	Mutation panel test	**Deep learning-based method**
Advantages	- High throughput method: multiplex analysis of various genes - Quantitative and sensitive detection of genomic aberrations	- More rapid turnaround time: once trained, the predictions are fast (less than 5 min per gene) and fully automated - Better picture of tumor heterogeneity: heat map analysis provides insights into spatial distribution of mutations - Remote testing: it may be able to detect genetic mutation from pictures taken directly from the microscope at the remote institute
Disadvantages	- Longer turnaround time: run lasts from 1 to 3 days - High complexity of workflow: requires complex sample preparation	- Requires separate classifier for each gene - Requires large training dataset: neural networks work best with more data - Deep learning method is a black box: it is not straightforward to understand how the decision is made