Author’s response to reviews

Title: Genome-wide analysis reveals the association between alternative splicing and DNA methylation across human solid tumors

Authors:
Xiaohui Sun (11418217@zju.edu.cn)
Yiping Tian (sdtyip@163.com)
Jianming Wang (jw1102@gsbs.rutgers.edu)
Yimin Zhu (zhuym@zju.edu.cn)

Version: 3 Date: 02 Oct 2019

Author’s response to reviews:

Dear Editors,

Thanks a lot for having reviewed our revised manuscript (Manuscript ID: MGNM-D-19-00004R2, Title: "Genome-wide analysis reveals the association between alternative splicing and DNA methylation across human solid tumors"). We would like to express our sincere thanks to the Editorial Office and the reviewers for the constructive and positive instructions and comments.

We are delighted to resubmit a recently revised version of the manuscript which was strictly revised according to the comments. Attached please find the following files.

1. Point-by-point response.
2. Revised manuscript text-highlighted in green.

Thank you for your further consideration of this manuscript. We look forward to your response.

Yours sincerely,

Yimin Zhu
Professor of Epidemiology
Dept. of Epidemiology and Health Statistics
Zhejiang University School of Public Health
E-mail address: zhuym@zju.edu.cn

Point-by-point response

GENERAL COMMENTS: In general, the authors attempted to respond my comments and made some changes in the manuscript, but they did not respond well. The work has potential interest to the readers, but it seems the work was mainly done by a new student or junior postdoc without mentor's careful supervision - by saying that, the authors need to carefully check such data, consider confounding factors, and assess on how the conclusion is reliable. If there were some factors as limitations, please
discuss on it. And figures are not in good resolution (perhaps due to journal submission system, but I have never seen figures in such low resolution during review).

REQUESTED REVISIONS:
Please see my comments below. I think the authors made effort on the changes, but they seemed to be not familiar with some of the analyses they did. The permutation test is very important, as well as the covariate adjusted analysis. The authors have not done well.

My point 1: Your description of permutation test (now it is a permutation test, not multiple permutation tests you stated in previous version) is still vague. I wondered the authors really understood what specific permutation test it should run. The numbers of CpGs and AS events after permutation test are more than those without permutation test. The authors may need to discuss why. Your response is copied below:
"After permutation test, 869 CpGs were found to be highly correlated with 465 cancer-specific AS events (FDR \textless 0.05) (Additional file 1: Table S4). Before constructing the permutation test, we found 710 CpG sites and 442 cancer-specific AS events showed significant associations."

Thank you for checking the survival analysis of using splicing or methylation only and comparing with the joint analysis. Your description of such analysis and comparison is not clear. For example, you found 16 CpGs had statistically significant differences. This is out of what number and what is the percentage? And when I compared your new Table 2 with that in previous version, it seems it changed dramatically. You added more genes but also removed other genes. For example, first row KIRC LPXN in your previous version, it is not included in your new version. If you have major changes, you need to explain for reviewer's justification. And importantly, you need to discuss such information in the manuscript (e.g. survival results by using only splicing or CpGs/methylation).

Response: Thank you for your comments, and we should apologize for our mistakes. We actually used a permutation test followed the procedure for the estimation of the experimental critical values that were proposed by Churchill and Doerge[1]. The description of the permutation test was “In order to infer their statistical significance, we further carried out permutation test. The number of permutations was 1,000 and we considered the permutation P\textlt;0.05 was statistically significant. In the present study, the aovperm() function in R package “permuco” and spearman_test() function in package “coin” were used for permutation test .” After rechecking, we found our mistakes. In this revised version, we considered the permutation P\textlt;0.05 was statistically significant and revised our results thoroughly and highlighted in green. The numbers of CpGs and AS events after the permutation test showed more than those without the permutation test. There are two interpretations of this. Firstly, the FDR approach of Benjamini and Hochberg yielded adjusted P-values that were initially more conservative than the permutation-based counterparts [2]. The FDR approach controls only the expected number rather than the actual number or proportion of false discoveries and under the naive assumption that variables independently. Secondly, the data might distribute non-normally, or there were discrete variables. Therefore, the results obtained by the two methods were inconsistent.

Thank you for your comments. We have depicted the survival analysis more specifically in our revision as “Cox regression analysis was performed adjusting for age, gender, TNM stage and adjuvant therapy, and results showed that 10.3\% (38/369) AS events could significantly distinguish patients with longer versus shorter survival prognoses (Table 2). Because splicing of introns could be regulated under a co-transcriptional mechanism, an alteration in gene expression may affect AS of corresponding genes. Thus, we examined mRNA expression of 35 genes containing cancer-specific AS events and found that the majority of them (85.71\%, 30/35) were not significantly associated with survival, suggesting that
these AS events are partially influenced by their gene expression (Table 2). Additionally, we evaluated DNA methylation of CpGs at the survival-related AS exons boundaries. 14 (19.7%) CpGs in 8 (21.1%) AS exons were found significantly associated with the overall survival of CRC patients. (Table 2).” (Page 15, paragraph 1). Additionally, the major changes in survival analysis were due to adding an adjustment for the “adjuvant therapy (including chemotherapy and radiotherapy)”. After adding “adjuvant therapy” in the Cox regression analysis, some genes were added and some were removed. The added genes seem to be independent prognostic markers for cancers. We also added some description and discussion of this in the section of Discussion as “Additionally, 14 CpGs at the exon boundaries of survival-related AS also showed significant association with overall survival, even after adjusting for age, gender, TNM stage, and adjuvant therapy. These results led us to speculate that the identified CpGs might be the specific cancer drivers, and serve as prognostic biomarkers for cancers.” (Page 18, paragraph 1)


My point 3: You need describe briefly such information in the manuscript so readers can understand the different scope and studies. This is important since you used their analyzed data. And you did it for my point 4.
Response: Thanks for your valuable comments. We have added such a summary in the Discussion section as “With the advantage of high-throughput data, the TCGA data portal provides opportunities for the integration analyses of multi-omics data. TCGA Splice Seq, web-based bioinformatics, providing a clear view of the mRNA splicing patterns of 33 tumor types, across a dataset of more than 10,000 TCGA samples. Ryan et al. identified and calculated each potential splicing event across 33 types of cancer to establish the TCGA Splice Seq database, while did not evaluate the potential mechanism and clinical usage of AS events [27]. In the present study, we integrate AS events from Splice Seq and TCGA data together to comprehensively explore potential regulatory mechanisms of DNA methylation for cancer-specific AS.” (Page 15, Paragraph 3).

Point 6: I asked you to specify how frequently as one example. You answered "the hypomethylation was more frequently observed in tumor tissues (51.3%)" (Page 3, line 9). In this case, the frequency is slightly more than 50%. Please be careful and do not overstate such features - I found you did not use frequently in the second revised manuscript. Please follow my suggestion to check other places, because I only used one example.
Response: Thank you for your suggestion, and we should apologize for our wrong description. We have checked other places in our manuscript and revised them highlighting in green.

Point 7: Please try your best. For example, in your second revised manuscript, abstract, "15,818 AS events". It should be "15,818 AS events".
Response: We should apologize for our mistakes and thank you for your comments. We have rechecked and revised our manuscript
Point 9: You did not respond my comment. I asked covariate "drug history". If no such data, you can explain it in your response/manuscript.
Response: We are so sorry that we did not respond to your comment. There was no information about “drug history” in the TCGA dataset. Thus, we depicted this as a limitation in the Discussion section as” Finally, our research was based on the public data which lacked several important clinical features, such as a history of drug use, and has not been studied in survival analysis.”

Point 10: Thank you for removing Figure 5 to suppl file. However, other figures are barely readable.
Response: Thank you for your comments and we were sorry for the low resolution of our figures. We have revised this in our new revision.