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: 1 Date: 18 Jun 2019

Author’s response to reviews:

Dear Editors,

Thank you for your letter and the comments concerning our manuscript entitled “Genome-wide analysis reveals the associations between alternative splicing and methylation across human solid tumors” (Article number: MGNM-D-19-00004). We found reviewers’ comments were very useful. Our manuscript was strictly revised according to the comments.

Attached please find 3 files.

1. Point-by-point response to reviewers.
2. Revised manuscript text-highlighted in red.
3. Revised manuscript text un-tracked.

The results have not been published in any part elsewhere and we conform to the submission requirements of BMC Medical Genomics. The corresponding author, on behalf of all coauthors, has declared no conflicts of interest

Yours sincerely,
Yimin Zhu

Professor of Epidemiology
Dept. of Epidemiology and Health Statistics
Zhejiang University School of Public Health
Point-by-point response to reviewers
Reviewer 1:
The authors use data from TCGASpliceSeq and TCGA DNA methylation to investigate the connection between alternative splicing and DNA methylation in a pan-cancer setting. The authors 1) characterize alternative splicing events in 10 solid tumor types; 2) identify differential methylation in splice sites between tumor and normal; 3) perform correlation analysis between splicing events and DNA methylation; and 4) perform survival analysis. The methodology is sound and the results are interesting, but several major issues must be addressed.

Major:

1. It is rather difficult to follow exactly what is happening during the analyses, e.g. how many CpGs are included in the different analyses, how many samples are used to in each analysis. It is also unclear what platforms are used for DNA methylation (only 450K or also 27K). Two things should be done to improve this: 1) make a flow chart (perhaps as Figure 1 in the results) where all analyses are mentioned, including how many samples are used, how many CpGs and splice events are going in, and how many significant associations are identified. 2) Go through Materials and Methods and add more details such as sample numbers, how the statistical tests are carried out etc.
Response: Thanks a lot and we have added a workflow chart and more details in Materials and Methods section according to your suggestion.

2. The introduction lacks some key references, first of all Kahles et al., Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients, Cancer Cell 2018. Crucially, the authors must evaluate how much of the content of their Figure 1 and chapter "Abnormal alternative splicing patterns in different types of cancers" is already covered by Kahles et al. and adjust this chapter and figure accordingly.
Response: Thanks for your suggestion. We have added Kahles’s paper and other key references to the Introduction section (page 4, paragraph 1). In addition, we have revised the Figure 1 and the corresponding chapter (page 10, paragraph 1).

3. Also for the introduction, Kahles et al. cite other key papers that the present authors should also mention in the introduction. "More recently, the analysis of alternative splicing has also been shown to be of prognostic value for multiple cancer types, including non-small cell lung cancer (Li et al., 2017), ovarian cancer (Zhu et al., 2017), breast cancer (Bjorklund et al., 2017), uveal melanoma (Robertson et al., 2017), and glioblastoma (Marcelino Meliso et al., 2017)."
Response: Thanks and revised as suggestion (page 4, paragraph 1).

4. For the Results section, the authors should try to describe their results better in the text. It should be possible to read the text without looking at the figures and vice versa.
Response: Thank you for the valuable comments. Revised accordingly.
5. The correlation analysis between splice events and DNA methylation is the core of the paper, but the authors describe both the methods and the results rather poorly. Be more specific about what CpGs and splice events are included in the analysis and what the significant findings are. I don't think Figure 3A should have two panels, rather make one panel with the distributions of the CpG-PSI associations. The importance of the findings should also be much more specifically discussed in the discussion.
Response: Thank you very much. Revised accordingly.

6. Another point regarding the correlation analysis: does it really make sense to only look at differentially methylated CpGs from tumor-normal? What if there is both hyper- and hypomethylation in tumors that is associated to both down- or upregulation of PSIs? This would be lost. At least, this limitation should be discussed.
Response: Thank you for your valuable comments. In the present study, we mainly focused on the regulation of DNA methylation on cancer-specific alternative splicing. Thus, we supposed that the differential methylation between tumor and normal tissue is the potential cause for alternative splicing. As you mentioned, we might omit the hyper- and hypo-methylation in tumors that is associated to either down- or up-regulation of PSIs, and we discussed this as a limitation in our revision.

7. Several points may be raised concerning the survival analysis. 1) Why are the Kidney cancers overrepresented? The authors don't discuss this. 2) Are the observed associations to survival associated to known subtypes of the studied cancer? For instance in breast cancer, it is known that splicing events is associated to ER status (Bjørklund et al. 2017) and therefore also associated to survival. How is this for kidney cancers? If the authors don't want to go into details about cancer subtypes, the importance of the survival analysis must be down-played in the text.
Response: Thank you for your comments. 1) We added several implications and predictions for the overrepresentation of kidney carcinoma in the Discussion section (page 16, paragraph 2) .2) our study mainly paid attention to the methylated-splicing feature across multiple cancer types instead of the difference within a cancer. While the alternative splicing events are associated with different subtypes of cancers, we supposed that AS might be one feature of cancer subtype but not the consequence. As Kahles et.al found, different cancer subtypes could be distinguished based on exon skip splicing features [1]. In addition, the renal cell carcinoma mainly included clear cell (RCC), papillary, and chromophobe, which represent on the order of 65%, 20%, and 5% of all RCC cases, respectively[2]. In present study, the RCC have already been classified as kidney renal clear cell carcinoma (KIRC) and kidney renal papillary carcinoma (KIRP) to analyze. Therefore we did not further go into details about the molecular cancer subtypes in our analysis.


8. The Discussion should be improved. Several paragraphs read as introduction, and should be moved there. Other paragraphs are merely repetition of the results, and this should be removed. What should be added is more insightful evaluation of how your results are important, and how they fit and add to existing knowledge.
Response: Thanks. Revised accordingly.
9. The language needs to be improved overall.
Response: Thanks. Revised accordingly.

Minor:

- "10 solid tumors" should probably be "10 solid tumor types"

- On several occasions (e.g. Figure 3 legend), it should probably say "Significant correlation", not "Positive correlation".

- Why not use boxplots in Figure 4A?

- Add p-values to Figure 4

- Add all results as supplementary data. This is useful for other researchers

- Explain "cassette exon"
Response: Thanks so much for your suggestion and we have revised them in our revision. For the “cassette exon”, the explanation of it is a splicing event in which an intervening exon between tow other exons form the mature mRNA sequence can be either included or skipped in order to generate two distinct protein isoforms. We should apologize that we have misused this in our paper and we have revised it.

Reviewer 2:

This study used the publicly available TCGA data in 10 solid tumor types to examine the association between methylation and splicing. They observed 442 exons with differential level of both expression and methylation in these tumor genomes. They also found 30 exons with both expression and methylation of the boundary CpGs. In principle, this study reported some features that might be interesting for understanding tumorigenesis and biomarkers. However, the results are likely false positives because there are many data points in the analyses and such an association might be artifacts without robust computational approaches and/or validation.

1. The most important concern is that whether the features you reported are real regarding the biological or clinical significance or by randomness. You need to perform additional computational work (e.g. randomization or permutation tests) to evaluate how likely such enriched features are random or unique from other sites.
Response: Thanks and we performed permutation tests in revised manuscript, and we found similar results with the previous one.

2. One specific question as above, in page 13 line 34, you reported "There were 30 exons showed both expression and CpG methylation levels significantly associated with overall survival (Table 2)". You should check whether those genes without your approaches are already significant in survival test. If so, your finding may be not the driving factor.
Response: Thank you for your valuable comments. We further performed survival analysis between gene expression and overall survival (in Table 2). We found that the most splicing events were not biased by their gene themselves.
3. Page 10, the first paragraph, the overview of the splicing results, are those similarly reported in original TCGA analysis papers? It is good to justify what are novel work from previous work.
Response: Thanks. The original TCGA analysis paper mainly reported their web-based resource TCGA SpliceSeq. The features of alternative splicing across different cancers were not showed. In present study, we wanted to explore the relationship between DNA methylation and cancer-specific alternative splicing (AS), which provided new perspective of regulatory mechanisms of cancer-specific AS.

4. Page 8, "200 nucleotides in size for intronic regions and 39 nucleotides for exonic regions 22." How these parameters were selected?
Response: Thank you for your valuable comments. The parameter selected in present study was based on the reference “Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. Nature Genetic.2008”. In their study, the authors examined the exon neighborhoods in the size of 200 nucleotides for intronic regions and 39 nucleotides for exonic regions to identify splicing cis regulatory elements in sequences in and adjacent to the tissue-regulated cassette exons. This definition was also defined based on previous reports.[1-3].


5. Page 8 (and also page 9), pathway enrichment analysis: why did you not use the adjusted p values for significance? Using p <0.05 for pathway enrichment cutoff is too generous and are likely false results.
Response: Thanks. Revised accordingly.

6. English writing needs substantial improvement. For example, in abstract "However, in human tumors, there is limited researches connecting dots between them."
Response: Thanks for your valuable comments. Revised accordingly.

7. Figure S1: it is better to swap X and Y axis. The numbers are easy to see and interpret on Y axis. This change will also be consistent with Figure S2.
Response: Thanks. Revised accordingly.

8. Page 12, line 20, what does this phrase mean "1.2:1.3 exons (chr19: 52390166-52391148: -)"?
Response: Thank you very much for pointing out our mistakes. We wanted to indicate exon 1.2 and exon 1.3 which located in Chr19: 52390166-52391148. We have revised this in our revision (page 12, paragraph 2)

9. Gene symbols should be italic.
Response: Thanks. Revised accordingly.