Peer Review File

Article information: https://dx.doi.org/10.21037/tcr-23-286

Reviewer A

Your manuscript, "Abnormal genetic and epigenetic patterns of m6A regulators associated with tumor microenvironment in colorectal cancer", shows a relevant landscape of genes related to the m6A regulation in patients with colorectal cancer (CRC).
You used public repositories to get data about expression, methylation, copy number variations, and genomic mutations for characterizing these genes and finding additional features. Although the potential contribution of this manuscript to its field, I would like to comment on some concerns:

Major comments

Comment 1: In line 92, you described you downloaded raw data from TCGA and cBioPortal repositories. How did you process raw data? Please, add all details as possible. This kind of analysis must be reproducible.

Reply 1: We downloaded level 3 raw data (expression, methylation, copy number variations data) from TCGA repositories, which were previously analyzed by the TCGA teams.

The mRNA expression profiles of CRC tissues and control samples were obtained from TCGA via UCSC Xena (1). The data analysis is based on the UCSC Xena website. Briefly, the gene expression profile was measured experimentally using the Illumina HiSeq 2000 RNA Sequencing platform by the University of North Carolina TCGA genome characterization center. This dataset shows the gene-level transcription estimates, as in log2(x+1) transformed RSEM normalized count. Genes are mapped onto the human genome coordinates using UCSC Xena HUGO probeMap.

DNA methylation profile was measured experimentally using the Illumina Infinium HumanMethylation450 platform. Beta values were derived at the Johns Hopkins University and University of Southern California TCGA genome characterization center. DNA methylation values, described as beta values, are recorded for each array probe in each sample via BeadStudio software (2). DNA methylation beta values are continuous variables between 0 and 1, representing the ratio of the intensity of the methylated bead type to the combined locus intensity. Thus higher beta values represent a higher level of DNA methylation, i.e. hypermethylation and lower beta values represent a lower level of DNA methylation, i.e. hypomethylation. Microarray probes

are mapped onto the human genome coordinates using Xena probeMap derived from the GEO GPL13534 record.

Copy number profile was measured experimentally using a whole genome microarray at a TCGA genome characterization center. Subsequently, the GISTIC2 method was applied using the TCGA FIREHOSE pipeline to produce gene-level copy number estimates (3). GISTIC2 was further thresholded the estimated values to -2,-1,0,1,2, representing homozygous deletion, single copy deletion, diploid normal copy, low-level copy number amplification, or high-level copy number amplification. Genes are mapped onto the human genome coordinates using UCSC Xena HUGO probeMap.

As for ATAC-seq analysis (Assay for Transposase-Accessible Chromatin with high throughput sequencing), normalized count matrix: a prior count of 5 is added to the raw counts, then put into a "counts per million", then log2 transformed, then quantile normalized, the result is TCGA_ATAC_Log2Counts_Matrix.180608.txt.gz. We downloaded TCGA_ATAC_Log2Counts_Matrix.180608.txt.gz from the synapse, averaged the values in the file (i.e. log2((count+5)PM)-qn values) across all technical replicates and all biospecimens belong to the same TCGA sample (4). The result is the genomic matrix TCGA_ATAC_peak_Log2Counts_dedup_sample. Peak location information is from file TCGA_ATAC_Log2Counts_Matrix.180608.txt.gz. Peak to gene mapping is defined as all peaks whose summit locates within 500kb from a gene's TSS sites.

The mutation analysis is directly based on the visualization results of the cBioPortal website, which is hosted by the Center for Molecular Oncology at Memorial Sloan Kettering Cancer Center (https://www.cbioportal.org/).

We added the data process in the corresponding method.

Changes in the text: The mRNA expression profiles of CRC tissues and control samples were obtained from TCGA via UCSC Xena (1). The data analysis is based on UCSC Xena website. Briefly, the gene expression profile was measured experimentally using the Illumina HiSeq 2000 RNA Sequencing platform by the University of North Carolina TCGA genome characterization center. Gene expression was estimated as in log2(x+1) transformed RSEM normalized count.

DNA methylation profiles of CRC patients were obtained using the Infinium HumanMethylation450 array from the GDC Data Portal. Beta values were derived at the Johns Hopkins University and University of Southern California TCGA genome characterization center. DNA methylation values, described as beta values, are recorded for each array probe in each sample via BeadStudio software (2).

As for ATAC-seq analysis, normalized count matrix: a prior count of 5 is added to

the raw counts, then put into a "counts per million", then log2 transformed, then quantile normalized. The values were averaged (i.e. log2((count+5)PM)-qn values) across all technical replicates and all biospecimens belong to the same TCGA sample (4).

Comment 2: In line 136, you wrote that you included 51 control samples. Nevertheless, it was not mentioned in the data acquisition. Could you check it, please?

Reply 2: As suggested, we added corresponding descriptions to the control samples in the data acquisition.

Changes in the text: The mRNA expression profiles and DNA methylation data of these tumor tissues and 51 control samples were obtained from TCGA via UCSC Xena.

Comment 3: In addition, you compared these 51 control samples with 380 primary tumors. It is a bit inconsistent with the 534 CRC patients whose information was downloaded in line 93.

Reply 3: After careful checking, we found that among 534 COAD patients, only 380 patients had RNA-seq data, so we used this part of the data for gene expression analysis. Indeed, it is a bit inconsistent with the 534 CRC patients whose information was downloaded in the method. Therefore, we corrected the description in the expression analysis. Also, we describe this more appropriately in the data download section.

Changes in the text: Of these, 380 patients had RNA-seq data and 441 patients had DNA methylation. The mRNA expression profiles and DNA methylation data of these tumor tissues and 51 control samples were obtained from The Cancer Genome Atlas (TCGA) via the University of California Santa Cruz (UCSC) Xena (1).

Comment 4: In Figure S1, you compared the expression levels of m6A-related genes between mutated and WT samples. However, it appears that he included all variants as one mutated group. Shouldn't you differentiate pathogenic variants from silent ones? Subsequently, there is no clear rationale for only selecting RBM15 and ZC3H13 for this classification.

Reply 4: As shown in Figure 2A, we defined synonymous mutations as mutated groups, which include inframe mutation, missense mutation, truncating mutation, and fusion. According to the American College of Medical Genetics and Genomics (ACMG) guidelines, mutations are classified as pathogenic, likely pathogenic, uncertain significance, likely benign, and benign depending on the supporting evidence (5). Indeed, a detailed breakdown of mutation might be more convincing. However, the classification of mutations in this study is according to the cBioPortal for Cancer Genomics website. Therefore, we did not have detailed information on the mutation to look for the corresponding evidence to support the classification. Furthermore, Due to the small number of patients, the further classification of mutation may result in too

few samples to be analyzed.

Although we did not differentiate the pathogenic variants from silent ones, the expression of ZC3H13 was significantly downregulated in the mutation group, while RBM15 expression was upregulated in the mutation group (Figure S1), suggesting that these mutations might affect their expression. Of course, the specific types of mutations that altered the expression of ZC3H13 and RBM15 need further research. This part's results are not perfect, but they at least provide a small hint.

Changes in the text: none.

Comment 5: For a leading discussion, would it be interesting to compare the performance of an m6A gene-based panel? Could it approximate your findings to further research and potential clinical applications?

Reply 5: In CRC, Chen et al. provided a comprehensive analysis of four RNA modifications and revealed the potential function of these writers in TME, transcriptional and post-transcriptional events, and identified their therapeutic liability in targeted therapy and immunotherapy (6). Zhang et al. found that nine candidate m6A-related mRNA biomarkers (LRRC17, NFKB1, NOS2, PCDHB2, RAB7A, RPS6KA1, RRNAD1, TLE6, and UBE2H) closely related to the clinicopathology and prognosis of CRC (7). Furthermore, Yue et al. developed an m6Ascore by PCA to quality m6A methylation patterns at an individual level, which predicts performance for overall survival and clinical efficacy of immunotherapy in patients with CRC (8). These studies mainly focused on the influence of m6A regulators themselves on CRC. In our study, we pay more attention to the upstream regulation mechanism of the expression of m6A regulators. Indeed, it would be interesting to compare the performance of an m6A gene-based panel in the discussion. Therefore, we added that to the discussion.

Changes in the text: In CRC, Other researchers provided a comprehensive analysis of four RNA modifications, identified m6A-related mRNA biomarkers related to the clinicopathology and prognosis, and even developed an m6Ascore to predict performance for overall survival and clinical efficacy of immunotherapy (6-8). These studies mainly focused on the influence of m6A regulators themselves on CRC. In our study, we pay more attention to the upstream regulation mechanism of the expression of m6A regulators.

Minor comments

Comment 6: Please, use conventional nomenclature guidelines for human genes. Write all gene names in italics.

Reply 6: As requested, we have written all the gene names in italics.
Changes in the text: All gene names are in italics.

Comment 7: Describe all abbreviations in their first mention. For example, "ATAC-Seq"
Reply 7: As requested, we have described all abbreviations in their first mention.
Changes in the text: We have described all abbreviations in their first mention, such as assay for transposase-accessible chromatin with high throughput sequencing (ATAC-Seq), University of California Santa Cruz (UCSC), etc.

Reviewer B

Although the amount of data analysis presented in the manuscript is considerable, the organization and aim of the work is unfocused.

Comment 1: There is a general lack of consistence in the presented data.
The percentage of mutated CRC samples, for instance, is generally low (a maximum of 7% mutated samples in the m6A regulators). Furthermore, no molecular or pathological correlation is given. Thus, the message given in the text is not sustained by these low percentages.
Reply 1: In this study, 24 m6A regulators had somatic mutations with frequencies ranging from 0.7% to 7% (Figure 2A), which overall is low. In spite of this, the expression of ZC3H13 was significantly downregulated in the mutation group compared with the wild-type group (Figure S1). Moreover, RBM15 expression was upregulated in the mutation group. This suggests that the mutations in ZC3H13 and RBM15, to some extent, might affect their expression. Indeed, it might be more convincing to give a molecular or pathological correlation. However, the analysis was limited by the small sample size of patients with mutations. We changed the subheading from “m6A regulators was mutated in CRC” to “The mutation characteristic of m6A regulators in CRC”, which is perhaps more appropriate.
Changes in the text: The mutation characteristic of m6A regulators in CRC.

Comment 2: the heterogeneity of the methylation patterns is, apparently, not relevant to the gene expression of the m6A regulators. This might be due to a variety of regulatory mechanisms. Thus the epigenetic significance of these observations can not represent a significant message in the manuscript.
Reply 2: Indeed, the heterogeneity of the methylation patterns is not relevant to the gene expression of the m6A regulators in this study. Although it is a negative result, this

further illustrates that DNA methylation does not play a significant role in the regulation of m6A regulators' gene expression, which is not quite the same as the traditional idea that methylation in promoter regions inhibits gene expression. This might be due to a variety of regulatory mechanisms. Thus, we naturally led to an analysis of the transcriptional regulation of m6A regulators. It is not reasonable that the epigenetic significance of DNA methylation represents a significant message in the manuscript. Therefore, we changed the subheading from "m6A regulators showed abnormal methylation patterns in CRC" to "DNA methylation did not associate with m6A regulators' expression".

Changes in the text: DNA methylation did not associate with m6A regulators' expression.

Comment 3: The chromatin accessibility is poorly described and the conclusions are unclear, overall. 7 regulators out of 19 is not a high percentage. The contribution of chromatin accessibility is highly speculative.

Reply 3: Chromatin accessibility is an important factor affecting gene expression. In this study, we firstly analysis the regulating mechanism of m6A regulators' expression from the perspective of epigenetic transcriptional regulation. Although the proportion controlled by chromatin accessibility is not a high percentage, at least seven genes were affected by this pattern. Indeed, the contribution of chromatin accessibility is highly speculative in this present state, which needs further study. About this part, we added the corresponding description to the discussion.

Changes in the text: However, the contribution of chromatin accessibility is highly speculative in this present state, which needs to be further explored.

Comment 4: The number of m6A regulators analyzed in the manuscript, together to the highly descriptive observations, lead to an inconclusive message.

Reply 4: In this study, we explored the molecular characterization and clinical significance of m6A modifications in CRC and revealed the genetic and epigenetic patterns of m6A regulators, which include expression, somatic mutation, CNV, DNA methylation, and chromatin accessibility. It is well known that different genes have different patterns of regulating gene expression. When analyzing the pattern of multi-gene expression, it is bound to appear that different genes are regulated in different ways. Indeed, it is true that we can't get around these inconsistencies very well. Therefore, we changed the conclusion from "m6A regulators were frequently dysregulated in CRC as a result of genetics (mutations and CNV) and epigenetics (chromatin accessibility)" to "m6A regulators were frequently dysregulated in CRC

partly due to genetics (CNV) and epigenetics (chromatin accessibility)".

Changes in the text: m6A regulators were frequently dysregulated in CRC partly due to genetics (CNV) and epigenetics (chromatin accessibility).

Comment 5: My suggestion is to revisit the results and the analysis by focusing on selected regulators, divided by category. This should lead to a better view of the disregulated ones, if that is the case.

Reply 5: Thank you very much for the reviewer's valuable suggestions. According to the suggestion, we have revised the subheading of the results and the description of the conclusion.

Changes in the text: The mutation characteristic of m6A regulators in CRC.

DNA methylation did not associate with m6A regulators' expression.

In summary, m6A regulators were frequently dysregulated in CRC partly due to genetics (CNV) and epigenetics (chromatin accessibility).