RESPONSE TO REVIEWERS

REVIEWER #3

The submitted manuscript presents the results of a study investigating differentially expressed microRNAs (miRs) in two groups of patients: those with early-onset colorectal cancer (CRC) (below 50 years old) and those with late-onset CRC. The authors aimed to identify miRs, predict their putative targets in CRC, and analyze enriched gene sets associated with early-onset CRC using two patient cohorts from Indian hospitals and CRC patients from the TCGA-COAD dataset. The study provides interesting results that are properly interpreted, well-discussed, and supportive of the final conclusions. However, the manuscript has several limitations that must be addressed before the study can be considered for acceptance.

Point 1: The discussion, particularly its initial paragraphs, should be more concise and specific, avoiding a recapitulation of the results and repetition of facts already presented in the Introduction.

Response: We thank the reviewer for the comments and the extensive review of our work. We have tried to make the discussion more concise and specific avoiding recapitulation of results and previously stated facts (**Section 3, "Discussion", lines 552-665**). We have touched upon the draw of the study only to highlight the results that have broader implications and need to be discussed in more detail. All newly added material in the rewritten discussion of the revised manuscript has been highlighted in yellow for easy reference.

Point 2: The authors should explain why they used only the COAD cohort to validate their CRC results, despite the TCGA also providing the READ cohort. Merging the READ and COAD cohorts might provide a better validation cohort, as the patients from the Indian cohorts included both colon and rectal lesions. If tumor localization in the rectum was an exclusion criterion, this should be clearly stated in the Methods section, specifically in the patient recruitment paragraph.

Response: We appreciate the reviewer's concern and have added TCGA-READ cohort data to the TCGA validations done in **Figure 3** and **Figure 4**. For **Figure 3** we have added the TCGA-READ cohort data alongside the TCGA-COAD data for each miRNA **("Results" Section 2.4, line no. 280-290; Fig 3 A-J, right panels and Figure 3 legend line no. 299-306)**.

For age-specific TCGA analysis (**Figure 4**), we integrated the **COAD** and **READ** cohort data to determine the Average CPM (AveCPM) for each subset. miRNA levels between Normal-Young, Normal-Old, Tumour-Young and Tumour-Old subsets were visualized by plotting the Log_2 AveCPM corresponding to each subset **("Results" Section 2.4, line no. 308-332; Fig 4 A-J and Figure 4 legend line no. 335-347)**. Details regarding the methodology of data acquisition and processing from TCGA-COAD and READ database have been included in the Materials and Methods section **(Section 4.13, line no. 836-850, page no. 52-53)**. Included areas have been highlighted in yellow for easy reference.

Point 3: The enrichment analysis presented in Figure 8 should be repeated using less general levels of gene ontology. Including more detailed Biological Processes (BPs), Molecular Functions (MFs), and Cellular Components (CCs) could provide much more informative insights.

Response: We thank the reviewer for the suggestion. In the previous version of the manuscript we were comparing experimentally validated targets of hsa-miR-148a-3p, hsa-miR-1247-3p, and hsa-miR-326 with differentially expressed (DE) genes in the TCGA datasets of colon adenocarcinoma (COAD) to identify differentially expressed miRNA targets (DE- miRNA targets). We realized that for more detailed enrichment analyses, we needed to increase our DE-miRNA targets. To that purpose, we included those miRNA targets that are differentially expressed in the TCGA dataset of rectal adenocarcinoma (READ) along with our previous COAD DE-miRNA targets for the enrichment analyses **("Results" Section 2.7, line no. 457-472; Fig 7 B, C, D and Figure 7 legend line no. 482-491)**.

Incorporating READ DE target genes lead to a detailed enrichment analysis which has been depicted in Figure 8 **("Results" Section 2.8, line no. 506-531; Fig 8 A-H and Figure 8 legend line no. 535-549)**. All inserted areas have been highlighted in yellow for easy reference.

Point 4: The manuscript contains numerous typographical errors, inconsistencies in spacing, and issues with the use of brackets, units, and formatting. These should be thoroughly corrected in the revised version of the manuscript to ensure clarity and consistency.

Response: We appreciate the reviewer's concern and have tried to correct the typographical errors and other issues that the reviewer has pointed out to the best of our ability.

Point 5: The descriptions of experiments, particularly the IHC and PCR methods, require significant improvements. More detailed information should be provided to ensure that the methods can be accurately reproduced if necessary.

Response: We have added more information related to the Immunohistochemistry (IHC) and Polymerase Chain Reaction (PCR) methods as pointed out by the reviewer. Overall, we have tried to give more detailed information in the Materials and Methods section so that the reader will be able to reproduce the results. Under "Patient recruitment" we have added details regarding cohort size and source of the discovery and validation cohorts used for the collection of biospecimens (**"Materials and Methods", Section 4.1, line 668-680**). For the IHC we have described in detail the steps followed for IHC staining along with details (vendor details, antibody dilutions) of the HRP/ DAB detection system and the primary antibodies used (**"Materials and Methods", Section 4.3, 'Immunohistochemistry', line 700-711**). For PCR we have added the annealing temperature for each pair of primers used (**"Materials and Methods", Section 4.6, 'PCR for APC gene', 'List of PCR Primers', column 4**). We have described in detail the method used to isolate RNA from the colorectal tissues (**"Materials and Methods", Section 4.4, 'RNA isolation', line 724-734**). For miRNA validation by qRT-PCR we have added details related to the Reverse Transcriptase (RT) and RT-qPCR reaction conditions and the vendor details of enzymes used (**"Materials and Methods", Section 4.16, 'miRNA validation by qRT-PCR'**). Sequence details of all primers used have been given. All newly inserted areas have been highlighted in yellow for easy reference.

Point 6: Vendor details should include detailed information, such as the city, state/province (if applicable), and country, when mentioned for the first time. Afterward, only the vendor's name should be used.

Response: We have added all vendor details as instructed. Included details have been highlighted in yellow for easy reference.

Point 7: The inline citations of webpages and documents should be revised to follow the editor's guidelines.

Response: Inline citations of webpages and documents have been revised as per editor's guidelines. Changed areas in the manuscript have been highlighted in yellow for easy reference.

Point 8: line 54-56 - difficult to understand, please revise line 43 - "all" or "males only"?

Response: We have revised the indicated areas as per the reviewer's suggestion.

Line 54-56: Reference: '**Age represents the primary risk factor for CRC, although the cumulative risk for early-onset CRC (0-49 years old) in India has increased (for males) from 0.06 (males) and 0.03 (females) in 1986 to 0.13 (males) and 0.09 (females) in 2012.**' Line 54-56 in the original manuscript corresponds to line 65-67 in the submitted revised manuscript. We have changed this line to '**Age represents the primary risk factor for CRC, although the cumulative risk for early-onset CRC (0-49 years old) in India has increased by 116% in males and 200% in females in the time span of 26 years (1986 to 2012)**'.

Line 43: '**males, all ages**' has been replaced by '**males only, all ages**'.

All changed areas in the revised manuscript has been highlighted in yellow for easy reference.