**Reviewer’s report**

**Title:** Tumor classification and biomarker discovery based on the 5'isomiR expression level

**Version:** 1  **Date:** 01 Jan 2019

**Reviewer:** Russell S. Schwartz

**Reviewer’s report:**

The authors have been fairly responsive to the critiques and the paper has been improved in the latest version. They provide a reasonable justification for focusing on the task of classifying cancer types, given the data available to them. They likewise provide a reasonable defense based on prior literature for the use of the chosen wrapper method for feature selection, although I do maintain that a comparison to alternatives on this dataset would be beneficial. They provide updated results from random forests in comparison to support vector machines. They provide some response to the question of how isomiR classification results compare to those from other data such as mRNA expression, although I again think the paper would be improved by actually applying the same classification algorithm to the corresponding TCGA mRNA expression data and comparing results. They clarify a number of prior points of confusion, showing that they have reasonably handled some methodological issues for which that was previously unclear, and have added a requested flowchart. The clarification suggests that they handled the term enrichment analysis reasonably, even if the end results are not greatly enlightening. Given all of this, I believe the paper is clear and technically sound even if the analysis could be taken further in some respects.

I have no new substantive points to raise, although there are a number of minor text errors in need of correction. The paper would benefit from a thorough proofreading, but I noted the following specific errors:

p. 2: ``and the cell-cell adhesion'' should be ``and cell-cell adhesion''

p. 3: ``integrate the comprehensive'' should be ``integrate comprehensive''

p. 3: ``good at deal with data'' should be ``good at dealing with data''

p. 3: ``which contained'' should be ``which contain''

p. 4: ``by combined the machine learning algorithm'' should be perhaps ``the combined machine learning algorithm''

p. 4: ``Besides the machine learning algorithms, such as SVM, SR, SRC, RF, KNN, has'' should perhaps be ``Beside the machine learning algorithms --- such as SVM, SR, SRC, RF, and KNN -- have''. Also, each acronym should be spelled out the first time it is used (e.g., support vector machine (SVM))
p. 4: ``GA is based on Charles Darwin's theory of natural evolution'' would read better as ``GA is based on Darwin's theory of natural selection''

p. 4: "evolution, the GA has proven" should be "evolution. GA has proven"

p. 4: "GA are frequently" should be either "GA is frequently" or "GAs are frequently"

p. 4: "Furthermore, the 5' isomiR may" should be "Furthermore, 5' isomiRs may"

p. 4: "with their canonical miRNA" should be "with their canonical miRNAs"

p. 5: "will be left only one" should be "will be left with only one"

p. 5: "300 generations GA" should be "300 generations of GA"

p. 5: "each of which comprised of 50" should be "each of which is comprised of 50"

p. 5: "the average sensitivity were" should be "the average sensitivities were"

p. 8: "samples from some cancer that" should be "samples from some cancers that"

p. 8: "It is note that" should be "It is noted that"

p. 8: "whether the detected isomiRs are the tissue-specific miRNA" should be "whether the detected isomiRs correspond to the tissue-specific miRNA"

p. 9: "are also list in" should be "are also listed in"

p. 11: "300 "generation" of GA/RF" should be "300 "generations" of GA/RF"

Figure 1: I am unclear what is meant by "digible"

Figure 1: "Combing" should be "Combining"

Figure 1: "Chromosome :50" should be "Chromosome: 50" and would look better in the same font size as the text below it

Figure 1: "Functional enrichment Analysis" should be "Functional enrichment analysis" to be consistent with other capitalization in the figure

**Are the methods appropriate and well described?**
If not, please specify what is required in your comments to the authors.

Yes

**Does the work include the necessary controls?**
If not, please specify which controls are required in your comments to the authors.

Yes

**Are the conclusions drawn adequately supported by the data shown?**
If not, please explain in your comments to the authors.

Yes

**Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?**
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

**Quality of written English**
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published

**Declaration of competing interests**
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests
I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal.