Author's response to reviews

Title: Small RNAs in metastatic and non-metastatic oral squamous cell carcinoma

Authors:

Patricia Severino (patricia.severino@einstein.br)
Liliane S Oliveira (liliane.oliveira@einstein.br)
Flavia M Andreghetto (flavia.andreghetto@einstein.br)
Natalia Torres (natalia.torres@einstein.br)
Otavio Curioni (curioni@usp.br)
Patricia M Cury (pmcury@hotmail.com)
Tatiana N Toporcov (toporcov@usp.br)
Alexandre R Paschoal (paschoal@utfpr.edu.br)
Alan M Durham (aland@usp.br)

Version: 3 Date: 6 May 2015

Author's response to reviews: see over
Dear Editor,

We are submitting a revised version of the manuscript and a point-by-point description of the changes made according to the suggestions and doubts of the referees. We thank the referees for their time and useful comments. The manuscript was definitely improved with the changes they suggested.

Referee 1
We would like to thank Dr. Shi for his time and for the useful comments and suggestions. We accepted most of the suggestions and we hope to have made clear all the points addressed by Dr. Shi.

Major Compulsory Revisions:

1. Potential sample selection bias was introduced by the use of a small group of T1 and T2 tumors with LN metastasis vs. another small group of T3 and T4 tumors. What criteria have been used to minimize the bias? Such possibility should be thoroughly discussed at least.

We agree with the reviewer that choosing T1 and T2 tumors with LN metastasis vs. a group of T3 and T4 tumors could introduce a bias. MiRNAs associated with tumor size and associated molecular characteristics (such as hypoxia levels, for instance) could have been selected. The reasoning behind dividing samples in these two groups was to minimize biological variation within the groups and, possibly, to select stronger markers linked to the metastatic phenotype due to their earlier presence in the tumor progression (i.e. T1/T2 tumors already presenting lymph node metastasis) or maybe a protective role due to their presence in non-metastatic larger tumors. We included this explanation in the first paragraph of the Results section. Once we selected what we considered to be the stronger candidates, we aimed at validating this result in additional samples from plasma and tissue, this time without
this stringent TNM classification. In this version of the manuscript we included additional results validating expression levels found by sequencing using real-time PCR in an additional set of 15 patients belonging to distinct pathological stages. We included the following sentence in the introduction: “Selected markers were validated in plasma collected from HNSCC patients before surgery and in additional tumor samples at various pathological stages”. Results show that some of the findings could be validated between N+ and N0 samples regardless of the tumor size (T1, T2, T3, T4).

Minor Essential Revisions:

1. Which edition of the TNM staging criteria was used? The 7th edition? Cite the original source so that it becomes obvious to readers.
   Yes, we used the 7th edition. We would like to thank the reviewer for pointing out that this detail was missing. We have included this information in the manuscript (Methods, section Patients and Samples).

2. It seems that primary tumor tissues were used in the study, not the metastatic lesions from positive lymph nodes. This should be made clear.
   Yes, primary tumor tissues were used in the study. We have made this information clearer in Methods, section Patients and Samples

3. On page 13, in the section for “Small RNA library construction and sequencing”, 18 libraries were made from 8 samples with LN metastasis and 8 without metastasis. What are the remaining 2 samples? Both 8s should be 9.
   The correct sentence should read: “10 samples presented lymph node metastasis at the time of diagnosis and 8 samples
did not present metastasis.” We corrected the sentence in the manuscript.

Discretionary Revisions:

1. Verifying with qPCR the differential expression of putative miRNA, candidate 12375, in the tissues would be interesting to have in this manuscript.
   It is our intention to keep on studying this possible new small RNA molecule. We will investigate its expression in additional HNSCC samples, as well as the presence of a related pre-miRNA in tissue. This should be included in a work carrying out more in silico and molecular analysis and could not be included in this manuscript at this point since it is under development.

Referee 2

We would like to thank Dr. Cascione for the careful reading of the manuscript and for his suggestions. We provide point-by-point responses to the requests and we hope we were able to adequately clarify the doubts raised by Dr. Cascione.

Major:

- The last sentence of the abstract is confusing "...suggest the evaluation of miRNA detection for the evaluation of oral squamous cell carcinoma metastatic potential.". Please, clarify this sentence.
  The sentence was modified to “... suggest the detection of microRNAs as a tool that may assist in the evaluation of oral squamous cell carcinoma metastatic potential.” We hope the meaning of this sentence is clearer now.

- Please use the new microRNA nomenclature. The two arms of a
microRNA were previously annotated as mir/mir*, with the general idea that mir*(mir-star) were degraded. But as we started sequencing at higher coverage, previously thought mir* were also found to be expressed at reasonable depth (specifically in developmental stages/tissues). miR-Base changed the nomenclature denoting the two arms as mir-5p/mir-3p.

We have included this information in Table 2, where differentially expressed miRNAs are presented in the manuscript text, and we have also changed additional files 2 and 3 to include results for each of the arms. We agree that since this information is available due to the sequencing technique and that it could be of interest to the community, we should include it in the manuscript.

- They sequenced 18 OSCC samples, the author wrote

  "Eighteen small RNA libraries were constructed, one for each OSCC sample: 8 samples presented lymph node metastasis at the time of diagnosis and 8 samples did not present metastasis.". How do they come from N0=8+N1=8 to a total of 18 samples

The correct sentence should read: “10 samples presented lymph node metastasis at the time of diagnosis and 8 samples did not present metastasis.” We corrected the sentence in the manuscript.

- Please, specify in Table 1, (add a column) which sample presented lymph-node metastasis and which were metastasis free.

This information can be found in Table 1 under “Pathological Stage”. N0 = no lymph node metastasis, N>0 = presence of lymph node metastasis. Since this is a common notation we didn't think we should include any explanations. Due to the doubt raised here, and in order to satisfy readers who are not familiar with the TNM staging, we have included this
Please upload your data to SRA (http://www.ncbi.nlm.nih.gov/sra) or specify where your sequencing data is available. We have submitted the data to SRA but during this process (which is long and we encountered a few difficulties) we ended up talking to a curator who pointed out the following issue: “Human derived sequence datasets have the potential to contain sufficient genotype information for the identification of a research subject. Therefore the consent status of the human subjects in your study must be established prior to data transfer. If the subjects have consented to the public display of genetic information, your data can be archived in the open SRA.” This kind of consent was not in the scope of our project, and we are definitely not allowed to share information that could lead to patient identification sometime in the future. Unfortunately, we have not found another solution for this situation so far. I am discussing with dbGaP, who provide restricted access to datasets, but it hasn’t been decided yet. For the meantime we added the following sentence in the manuscript “Due to concerns regarding public sharing of patient sequence dataset, raw sequencing results are available upon request but, depending on the scope of the study, it will have to be submitted to the Ethics Committee approval.” I hope you understand and, in case you have another solution for this situation we would be glad to hear it.

The authors wrote "Six miRNAs were found to be upregulated in metastatic OSCC with statistic significance.". They did not mention the p-value cutoff used to consider a miRNA statistically differentially expressed. We included the cut-off for statistical significance in the Table 2 legend as well as in the text. Additionally, we have reviewed
this information in Methods, section Differential Gene Expression Analysis, in order to make it clearer.

- The p-values reported in table 2, are they adjusted for multiple testing?
  Yes, they were. This information can be found in the legend for Table 2.

- The authors evaluated the expression of the differentially expressed miRNAs in N0 vs N1 patients in plasma of 30 patients by real-time. Did they perform any statistical test on the real-time data? Do they have a p-value of each miRNA? If so, did they adjust the p-values for multiple testing?
  The aim of this analysis was mostly qualitative. We wanted to see if miRNAs found in tissues could be found in the plasma of patients since this kind of detection would have a much more interesting clinical application. Considering miRNAs previously found to be circulating in human plasma by large scale screening studies, most plasma samples evaluated presented miRNA expression levels in agreement with the sequencing results and this is what we intended to show. We are currently working on increasing the number of patients evaluated in order to propose the use of these specific molecules as plasma biomarkers, but this was not the aim of the result presented in this work.

Minor:

- Why do you only focus on miRNAanda predictions in your miRNA target analysis?
  Many different algorithms for target scanning are available but the results usually include different target sets (Brief Bioinform. 2014 Jan;15(1):1-19. doi: 10.1093/bib/bbs075. Epub 2012 Nov 22). However, a 2010 review showed higher
specificity of miRanda in a comparison against RNAHybrid and TargetScan (J Integr Bioinform. 2010 Mar 25;7(3) and it is also the algorithm used by mirBase, the source on which we based the first part of our analysis. Taking these pieces of information together we decided to use only miRanda for the target prediction. This, in our view, would produce more consistent results with the rest of the paper. We made a small change on the article to emphasize the reason for our choice (Methods, section Identification of Novel Differentially Expressed Small RNA-like Molecules)

- Page 8, "...comprehension of it mechanistic role " should be "...comprehension of its mechanistic role"
  It has been corrected.

- Page 16 "Characterizing this putative precurso miRNA..." should be "Characterizing this putative precursor miRNA..."
  It has been corrected.

- The author should cite and use miRandola, the Extracellular/Circulating microRNAs database. miRandola: Extracellular Circulating microRNAs Database PLoS ONE 7(10): e47786 doi:10.1371/journal.pone.0047786
  Thank you for this suggestion. We included this reference in page 10, since it was used to double check which of the differentially expressed miRNAs identified when comparing tissue derived from N+ and N0 samples had been previously detected in plasma.