Peer Review File

Article Information: http://dx.doi.org/10.21037/tlcr-20-959

Reviewer A

Article entitled: 'MicroRNA expression profiling and biomarker validation in treatment-naïve and drug resistant non-small cell lung cancer' is interesting. The manuscript is noteworthy as it describes potential markers of resistance to chemotherapy in NSCLC patients. Much research is currently focused on the prognostic and prediction markers of TKI and ICI, and chemotherapy is still used in patients who are not eligible for targeted therapies.

Researchers selected only 5 cisplatin associated molecules (4 in ADC and 5 in SqSC) from a very large group of microRNAs, which indicates the amount of work that needs to be done to select microRNA-prognostic-predictive markers. Below are some notes regarding the manuscript:

We thank Reviewer A for their kind comments and excellent review of this manuscript. We have responded to each of the queries/points below, and feel that these have greatly enhanced the current revision. We hope that Reviewer A will find these revisions satisfactory and will now find the manuscript suitable for acceptance/publication in TLCR.

Comment 1: The number of NSCLC patients is not exposed. It is true that the authors indicate 10 corresponding samples of serum and NSCLC tissue, which is an extremely small group. This should be mentioned in the discussion because in a larger group of patients, the results of such studies could be quite different. I consider this to be the weakness of this article. It is true that the authors mentioned a small group in the paragraph on miRNA-4286 and point out that: 'These data are inconclusive due to the small sample numbers included in this analysis and warrants further validation in a larger cohort of patients.' however, this applies to all the results obtained based on 10 samples of tissue and serum.

Reply 1: We thank Reviewer A for this comment. The authors agree in that the small sample size used in the study and validation of miRNAs using serum and tissue samples from lung cancer patients is a significant limitation in this study. To highlight this study limitation, the discussion section has

been updated to include how the findings arising from biomarker studies using small sample cohorts, can differ to validation outcomes and predictive value when tested in a larger cohort of patients. In addition, the majority of these miRNAs were validated in TCGA datasets (please see Comment 2 below), and feel that whilst the initial cohort of samples studied in this manuscript was indeed small, it was large enough to correctly identify significantly altered miRNAs in these larger datasets, and hope that our discussion of this initial limitation is sufficient for Reviewer A.

Changes in the text: We have modified the discussion to include more information relating to this limitation in sample size and subsequent validations. This has been added to page 31, lines 659-669.

Comment 2: The authors in Supplemental Table 1 point to some results that raise the discussion; however in the materials and methods they do not mention that they performed such an analysis. This needs to be adjusted, because in the description of the table (which should be above the table and not below it), the authors indicate the number of TCGA ADC and SqCC samples.

Reply 2: We thank Reviewer A for this constructive comment, and apologize for failing to include this in our original manuscript. The analysis using in silico TCGA lung cancer datasets was not included in the Methods section of the original manuscript and we agree that this should be corrected. We have now included a new section, “In silico analysis” in the Methods section of the revised manuscript and have adjusted the title of Supplementary Table 1 such that it is placed above the table.

Changes in the text: We have added a new section on “In silico analysis” to the Methods section of the revised manuscript. This addition has been made on page 12, lines 223-241. In addition, the title corresponding to Supplementary Table 1 has been added above the table (see revised Supplementary Table 1) and subsequently adjusted for all Supplementary Tables included in the revised manuscript.

Comment 3: Ethics paragraph should follow the Statistical analysis section. It should also include the statement 'Patient samples used in this study were ethically approved (…) (Ref. No. 041018/8804), which is currently in Patients section.

Reply 3: We thank Reviewer A for this comment. We have now amended this by removing this sentence from the “Patients” section and placed under that of “Ethics”.

Changes in the text: This text has been added to the "Ethics" section of the revised manuscript, page 13, lines 259-260.

Comment 4: The article is quite long.

Reply 4: We appreciate that the manuscript as a whole is a quite long. Based on the peer-review process from previous submissions of this work to other journals, it was necessary to supplement various sections of the manuscript with additional information. This resulted in the need to expand the Methods, Results and Discussion sections of the manuscript to enhance the overall quality of the manuscript. As such, it includes a wider discussion of our findings and how these reflect what is currently known in the literature and that it would be understood by a less knowledgeable reader. This is further supported by Reviewer B who requested that "the results part must be improved with more explanations and discussions". As such, we do apologize but feel that by removing any further text from the current manuscript would not adequately justify the research findings presented. We hope that Reviewer A will understand this and accept that we do not wish to shorten the manuscript as we have to respond to Reviewer B in this regard.

Changes in the text: No changes have therefore been made in relation to the above. Please see comments for Reviewer B who requested additional discussions.

Comment 5: The figures are small and the inscriptions on them are barely visible.

Reply 5: We thank Reviewer A for this comment, and apologize if the figures cannot be read clearly. We have now adjusted the size and resolution of all figures and trust that these will now be more clear and readable.

Changes in the text: New revised figures have been uploaded as part of this revision where the resolution has been increased to 300 dpi and a Bit depth of 24, thereby significantly improving their clarity and readability.

Reviewer B

The manuscript "MicroRNA expression profiling and biomarker validation in treatment-naïve and drug resistant non-small cell lung cancer" by MacDonagh et al. identified a panel of miRNAs which may have diagnostic and prognostic potential as novel biomarkers in lung cancer. However, the result part must be improved with more explanations and discussions.

We thank the Reviewer B for a constructive review of our manuscript. We have tried to incorporate/respond to all of the raised comments, and believe that the manuscript has now been greatly improved as a result.

Comment 1: In Fig. 4, although the authors show that miR-4286 is increased in sera of SqCC patients, the authors should show correlation between miR-4286 levels in sera and cisplatin resistance in NSCLC patients.

Reply 1: We agree with Reviewer B that this is an issue that needs to be resolved. However, these miRNAs were examined in tissues and serum from chemo-naive NSCLC where miR-4286 levels may be affected in the sera as a result of expression from other cell types. As
such this miRNA may not be a suitable candidate to predict/monitor the development of resistance moving forwards. However, our data do show that this miRNA (miR-4286) has prognostic value in NSCLC, and as such, may translate to having a more predictive role pending larger validation studies in the future. We hope that this explanation will prove satisfactory and have added text to the discussion and the Figure legends which we hope will alleviate any concerns of Reviewer B.

Changes in the text: This has been highlighted in the discussion section on page 31, lines 670-674. Furthermore, we have amended the figure legend for Figure 5 to highlight the treatment status of these patients by including “treatment-naïve” to the legend, page 45, line 980.

Comment 2: What kind of cells are secreting miR-4286 in the blood flow? And what is the underlying mechanism (exosomes?). The authors should discuss about this issue.

Reply 2: We thank Reviewer B for this pertinent comment. Of interest in this study, miR-4286 was the only miRNA detectable in the sera of NSCLC patients, where it was significantly increased in SqCC patients relative to ADC patients or healthy controls (non-malignant). As such, this is most relevant

to the query raised regarding the type of cells from which this miRNA may be secreted. While it is difficult to decipher exactly the source of this specific miRNA in our NSCLC patient cohort, we have discussed these possibilities based on current knowledge in the field, and included a discussion of the potential role of exosomes in this process and hope that this will be found to be acceptable by Reviewer B.

Changes in the text: A new section has been added to the discussion to address the possible source of miR-4286, page308, lines 632-658.

Reviewer C

The manuscript "miRNA expression profiling and biomarker validation in treatment-naïve and drug resistance NSCLC" by MacDonagh and colleagues is presented as a study to evaluate the hypothesis that miRNA have diagnostic, prognostic and predictive capacity for anticipating cancer incidence and response to platinum-based therapies. The authors use cell-based models, xenografts, and patient tumor samples to test the hypothesis. Ultimately a 5-gene signature is uncovered that can distinguish cisplatin-resistant cells from normal and adenocarcinoma from squamous cell carcinoma. Expression of this signature was apparent in squamous cell carcinomas with poor outcomes and mir-4286 could be found in serum. Most of these studies seem to follow from earlier work by the senior author. Despite these novel findings, it is unclear whether any of these findings have clinical value.

We thank Reviewer C for these insightful comments following review of our manuscript. We have attempted to address all of the raised concerns and hope that the revised manuscript and responses addressing these are now acceptable by Reviewer C.

Comment 1: The methods are relatively clear, except for the concentrations of cisplatin used in the proliferative, clonogenic, and apoptosis assays. I assume these treatments followed from those outlined in reference 18.

Reply 1: We thank Reviewer C for this comment. The concentrations of cisplatin used in the analyses represented in Figure 2 for proliferation, clonogenic survival and apoptosis are based on similar cisplatin concentrations and dose-response studies previously published by our group (Barr et al., PLoS One, 2013).

Changes in text: Cisplatin concentrations are indicated in the Methods section of the manuscript for proliferation (0-100μM, page 11, line 202), apoptosis (0-100μM, page 11, line 212) and clonogenics (0-10μM, page 11, line 206). The increasing concentrations of cisplatin used in each of the above functional assays are shown on the X-axes of graphs shown in Figure 2.

Comment 2: There are misspellings in lines 117 and 275.

Reply 2: We apologize for these errors in the main body of text and have now amended these.

Changes in the text: The spelling errors have been corrected. The word "hypothesis" has been replaced with "hypothesise" (page 8, line 152) while "artefact's" has been corrected to "artifacts" (page 16, line 326).

Comment 3: The results section outlining the miRNA profiling experiments is quite hard to follow as there are many editorial comments probably best suited for the discussion section. It is not clear why the authors chose separate criteria for supervised v. unsupervised analyses or why the criteria for fold change and cell line expression change were chosen. The mir-1246 probe was one of 22 discarded as being unreliable but ended up as one of the 10 finalists for qPCR validation.

Reply 3: We thank Reviewer C for this comment and apologize for any confusion relating to the analyses of miRNAs carried out in this study, and described in the results section. MiRNA profiling of the paired isogenic cisplatin resistant and parental NSCLC cell lines was carried out commercially by Exiqon MicroRNA array services, where technical data quality assessment and data analysis were performed on these microarrays.

In the analysis of miRNA datasets, multivariate statistical techniques are often used in the analysis of miRNA expression data and their association with particular diseases or conditions. The analysis can be supervised using a discriminant technique to focus on groups of arrays that are of a priori interest. This approach is especially useful as there is no requirement for the filtering of expression data or the generation of miRNA lists or clusters. The method can take an entire microarray dataset and cross reference/integrate it with miRNA prediction databases without the use of user defined thresholds. This can be used in a supervised mode where groups are specified in advance. It can also be used for data exploration in an unsupervised mode. This is used in cases where the samples show great heterogeneity or are poorly characterized, as happens, for example, in many cancer related datasets.

In the unsupervised analysis carried out on the miRNA analyses in question, Principal Component Analysis (PCA) is a method used to reduce the dimensions of large data sets and is a useful way to explore the naturally arising sample classes based on the expression profile. By including the top 50 miRNAs that have the largest variation across all samples, an overview of how the samples cluster based on this variance is obtained. If the biological differences between the samples are pronounced, this will be a primary component of the variation. This leads to separation of samples in different regions of a PCA plot corresponding to their biology. If other factors (such as sample quality) inflict more variation on the samples, the samples will not cluster according to the biology. For the unsupervised analysis used, a small subset of miRNAs were excluded.

Why are some of these miRNAs excluded from the unsupervised analysis (and which ones)?

It is a common phenomenon in such analyses that a small number of probes can sometimes display atypical signal patterns (eg. very high signal levels in different sample types). Therefore, in order to eliminate any possibility that these signals are not representative of biologically relevant miRNA expression and to avoid focusing on potential false-positives, a total of 22 miRNAs were initially excluded from the unsupervised analysis. These included; hsa-miR-1246, hsa-miR-1273g-3p, hsa-miR-1280, hsa-miR1908, hsa-miR-3124-3p, hsa-miR-3686, hsa-miR-3940-5p, hsa-miR-3960, hsa-miR-4279, hsa-miR-4285, hsa-miR-4290, hsa-miR-4443, hsa-miR-4454, hsa-miR-4456, hsa-miR-4467, hsa-miR-4497, hsa-miR-4516, hsa-miR-4639-3p, hsa-miR-4708-3p, hsa-miR-4764-3p, hsamiR-4787-5p,

hsa-miR-4800-3p and hsa-miR-5100. The Heat Map shown in Figure 1B represents the unsupervised hierarchical clustering of the top 50 miRNAs with highest standard deviation and excluding the above 22 miRNAs. These miRNAs were however included in the supervised analysis for overall changes in expression of all miRNAs examined across all samples.

It is recommended that miRNAs are identified that show a sufficient level of regulation across the relevant groups of samples in this study. While it is possible to validate miRNAs that show small regulations, it is also important to consider that smaller fold changes tend to be relatively more affected by technical variance. Such changes are thus associated with increased risk of false-positive signals. Based on this, Exiqon recommended the inclusion of miRNAs showing more than a 2-fold change in expression in subsequent validation studies by qPCR. For this reason, fold-changes in miRNA expression was used in the data presented.

While miR-1246 was one of 22 miRNAs removed from the unsupervised clustering analysis outlined above, these miRNAs were still included in the overall data analyses of those miRNAs that were differentially expressed between PT vs CisR lung cancer cell lines. In doing so, this particular miRNA (miR-1246) was found to be upregulated across all five cisplatin resistant cell lines relative to their parental/sensitive counterparts and as such, warranted validation by qPCR. Under ideal conditions, one would expect to see perfect correlations between miRNA array data and qPCR validations. However, this is very often not the case when using microarrays, mainly due to the probe-based hybridization approach used, while validations use primers and probes that may not be the same as those on the array. Array-based methodologies therefore have a known false-discovery rate, and must always be subsequently validated to confirm the specificity of the result.

Changes in the text: We hope that the information provided has helped to clarify and explain in more detail, the queries highlighted by Reviewer C. In order to further highlight the unsupervised analysis represented in the heat map (Figure 1B), the corresponding legend for this figure has been amended to reflect this and now states, “The heat map represents unsupervised hierarchical clustering”, page 43, lines 930-931. Furthermore, in the Results section, fold-change in expression was added (pages 15-16, lines 304-305) to highlight that the differential changes in miRNAs were represented as fold-changes in expression.

Comment 4: Manipulation of the expression of 3 of the 5 miRNA molecules was largely unremarkable and did not consistently follow previously published observations, albeit in different systems. It is unclear why the authors did not investigate any of the potential gene targets of the miRNAs, especially those of mir-4286, since it was observed that overexpression modulated both apoptosis and clonogenic assays. The authors note in the discussion that "several gene targets implicated in regulation of proliferation and apoptosis" might be affected. Why not have a look?

Reply 4: We thank Reviewer C for raising this very valid point. While miRNA target identification and validation studies were outside the scope of the current manuscript, we have now attempted to address this using in silico analyses. In doing this, TargetScan (http://www.targetscan.org/vert_72/) was used to identify potential candidate mRNAs targeted by miR-4286. The top 10 genes with a known role in cellular proliferation were identified and their expression examined using meta-analysis in SqCC. In addition, their comparative expression was examined in the TCGA-LUSC dataset alone, using Lung Cancer Explorer (https://lce.biohpc.swmed.edu/lungcancer/). Overall, the majority of these top 10 candidate genes showed significantly decreased/downregulated expression in SqCC which, whilst suggestive that these are regulated by this miRNA, have yet to be functionally validated.

Changes in the text: These new analyses have been incorporated into the revised manuscript with additional sections added to the Methods (page 12, lines 224-241), a description of these data in the Results section (page 22, lines 454-462), and an update to the Discussion section (page 29, 624-631). Additional Supplemental Files (Supplementary Table 3, Supplementary Table 4, Supplementary Figure 2) have been provided which incorporates the results of these analyses.

Comment 5: The authors propose that the signature might differentiate histology of lung cancers, but this is a standard pathological analysis tool unlikely to be replaced. mir-4286 can be found in the blood of SqCC disease, and this might be useful as a screening tool.

Reply 5: We agree with this statement by Reviewer C in that, while specific miRNAs may have the ability to distinguish between different NSCLC histologies as demonstrated in our study and other studies, these would not replace current gold standard methodologies for the histopathological diagnosis of lung tumours.

Changes in the text: We have added a line to the relevant section of the discussion to highlight this (page 29, lines 615-618).

Comment 6: I found the analyses in the Supplementary data much more compelling, but they did not necessarily support the data provided from the author's own studies (Supplementary Table1, especially). I think this manuscript could be vastly improved by focusing on the data that have clinical value.

Reply 6: We thank Reviewer C for this comment. The miRNAs identified in this study were validated against TCGA datasets. Whilst the cohort of samples studied in this manuscript was small, it was large enough to correctly identify significantly altered miRNAs in these larger online datasets. In addressing the relevant clinical value of these findings, we have added to the discussion section, to offer an explanation as to the contrasting differences between miRNA expression in our cisplatin resistant cell line model and those in patient samples (tissue vs serum) examined in our study and in the TCGA datasets and their clinical relevance.

Changes in the text: The discussion has been updated to highlight the potential clinical value of these data (page 28, lines 590-606).