Author’s response to reviews

Title: Identification of biomarkers for amyotrophic lateral sclerosis by comprehensive analysis of exosomal mRNAs in human cerebrospinal fluid.

Authors:

Kentaro Otake (kentarou.ootake@takeda.com)
Hidenori Kamiguchi (hidenori.kamiguchi@takeda.com)
Yoshihiko Hirozane (yoshihiko.hirozane@takeda.com)

Version: 2 Date: 18 Nov 2018

Author’s response to reviews:

Dear Editors;

Thank you for giving us a chance to re-submit a revised version of our manuscript entitled “Identification of biomarkers for amyotrophic lateral sclerosis by comprehensive analysis of exosomal mRNAs in human cerebrospinal fluid.” to BMC Medical Genomics for further consideration. We also appreciate the time and effort you and each of the reviewers have dedicated to providing insightful feedback on ways to empower our paper. We have incorporated changes or explanations in the manuscript to reflect the detailed suggestions you have kindly provided. We also hope that our edits and the responses we provide below satisfactorily address all the issues and concerns you and the reviewers have noted. To facilitate your review of our revisions, the followings are point-by-point responses to the questions and comments.

**Answer to the comments from Reviewer 1**

1. Excellent responses to reviewer comments, with great implementation of new results and revised structure of the paper's message.
   > We sincerely appreciate the encouraging comments.

2. Some remaining grammatical issues (usually the absence of 'the' or lack of pluralised words).
   > They were corrected throughout the manuscript.
3. The opening paragraph to the discussion is still a little understated. A more direct sentence on achieving their aims would be more impactful; for example: "we successful implemented an exoRNA-seq method that identified mRNA changes between ALS cases and NH controls". But this is a stylistic suggestion and up to the authors.

> An achievement of exoRNA-seq development was described in the opening paragraph to the discussion and conclusion as suggested (line 233-235 and 292).

**Answer to the comments from Reviewer 2**

1. The authors made good response and added additional materials (some more details of the data, deposit of the data to GEO, RT-PCR, among others). however, the main concern of large variation of gene expression among four samples remains. As you can see Figure 3D, the variation is quite strong. That says, it is always easy to find some genes showing strong difference among the disease versus control samples. This is a big scientific issue - reproducibility of the results. The statistical power is not sufficient in this study.

> We truly appreciate letting us know the concerns on data interpretation that the reviewer has. The causion of low statistical power derived from small sample size was described as limitation (line 271-274). However, the reproducibility of our method was assured as shown in Figure 3A. It led us to a conclusion that the large variation among four NH samples in Figure 3D was not caused by technical instability but by another factors such as biological variation or artificial characteristics of specimens (i.e. Preservation period of the commercially available CSF after collection was different from each other.) (line 274-281). As this manuscript mainly focused on the technology development of exosomal RNA-sequencing, we want to ask for the reviewer's understanding that we could prove the validity of our technique.

2. The authors replied "5,006 genes passed threshold for calculation (line 180-182, 592-593)." This itself has a big concern. What is the hypothesis here and statistical test? If this observation is reliable, then the whole transcriptome has been changed in the ALS disease form.

> We put all 30,373 genes (not common 4,580 genes in NH specimens) annotated by human genome reference B37.3 into DESeq2 analysis. In DESeq2, adjusted p-value for FDR correction (alpha = 0.05) was not calculated for the genes with (i) no detection in any samples and (ii) ourlier samples identified by cook's distance. As a result, 5,006 genes could be calculated for adjusted p-value up to 1. Of these genes, 543 genes were detected as DEGs with adjusted p-value less than 0.05. These explanations were added to the manuscript (line 184-186, 273-274).
3. Line 180-182, it is not clear how the FDR was calculated. Given that so many genes were compared and sample size was very small, it would be important to disclose with more details, instead of stating "(data not shown, see Additional file 3)."

> FDR was controlled by ordinary Benjamini-Hochberg method with alpha level of 0.05 (line 369-370). Although this is calculated in Array Studio software in this manuscript, we also confirmed same results in BH method in R (CRAN). As above, FDR correction was not applicable in genes with (i) no detection in any samples or (ii) outlier samples identified by cook's distance (line 184-186, 273-274).

4. The authors stated "However, we think it should be noted that representative DEGs were also confirmed by qRT-PCR." This may help, but again the authors did not even much understand the scientific design or result interpretation. It only confirms the expression, but it is risky to conclude this gene is statistically significant. Would the expression represent the same trend if many more samples were selected? This is a fundamental statistical issue - the authors need to consult with biostatistician.

> As the reviewer pointed out, statistical significance of the results in qRT-PCR validation is important although it was not described in the first revision of the manuscript. In fact, p-value of Ct comparison in CUEDC2 between NH and ALS groups was 0.0569 by Welch's t-test after confirmation of the Gaussian distribution by Kolmogorov-Smirnov test (line 215-219, 284, 370-373). At the same time, we have realized statistical significance in this validation step should be examined and discussed in another cohort with large sample size to ensure statistical power (line 286-287) because this manuscript emphasized technology development of exoRNA-seq.

5. They added QC and make the data available to the readers via GEO GSE121519. This dataset is not currently searchable in GEO - this is likely the authors have embargo data and it is understandable. The authors must assure to have it available as academic standard.

> We are sorry that we missed submitting the raw data of QC to GEO. This dataset is now available in GSE121519.
1. As above, while this reviewer still has main concern, the authors need to at least state more clearly this paper is technique based. "

> We deeply appreciate this helpful comment. An argument in technology development of exoRNA-seq was emphasized in the section of abstract, discussion, and conclusion the manuscript. (line 22-25, 40-42, 233-235, and 292)

Again, thank you for giving us the opportunity to strengthen our manuscript with your valuable comments and queries. We have worked hard to incorporate your feedback and hope that these revisions persuade you to accept our submission.

Sincerely,

Kentaro Otake