Author Response

Thank you and two anonymous reviewers very much for reviewing our manuscript! Both reviewers give valuable and constructive comments for us to improve our manuscript. Based on these comments, we have revised our manuscript.

The novelty of our work contains the following three major points.

• our study detected gene fusion events in IPF tissues, and identified several gene fusion events that occur in IPF tissues with significant-high incidence rates

• our study probes the possible impacts of IL7=AC083837.1 gene fusion on the progression of IPF, and suggests that it may exacerbate IPF symptoms, especially the lung tissue fibrosis, by promoting the signaling pathways of NK cell-mediated cytotoxicity, angiogenesis, and apoptotic process.

• To guarantee the validity and reliability of our research, we study 91 high-quality transcriptomes collected from four IPF projects, with the cross-batch bias being corrected.

The major changes and additions to the revised manuscript are as follows.

• Added a statement about why we choose to study the gene fusion of IL7=AC083837.1

• Replot two figures, the leading-edge heatmap (now figure 4) and the map of subset analysis of all enriched gene sets (now figure5)

• Added 3 supplementary files, including the differential expression gene list, the results of GSEA enrichment analysis, and the results of GSEA co-expression network of IL7.

• Added statements about null hypothesis and type 1 error in legends of two tables

• Added a statement about the significant enrichment of IFN pathways in our GSEA analysis

• Added legends to all supplementary files.

• A colleague who is experienced in academic English writing conducted proof-reading on this revision.

Below are the answers to the detailed questions:

Reviewer #1

Reviewer's #1 comment 1, major revision

1) Observing table 1, there were several gene fusion combinations with a significant incident rate. It is not so evident the reasons why you decide to focus just on the IL7-AC083837.1 and not consider the others (for example, some of them presented better Chi-square p-value). I guess due to IL7-AC083837.1 were the only genes located nearby. Is it correct? A more detailed explanation would be highly appreciated.

A: You are correct that the location of gene loci is one important factor for this decision making. To clarify this confusion, we add a paragraph in the manuscript as follows:

“Based on the study of these gene fusion events in their gene loci, occurrence differences between groups, and known functions in biological processes, we determine to focus our research on the IL7=AC083837.1 fusion, which most possibly involves in the IPF pathology. Firstly, IL7=AC083837.1 and MFAP4=EPN2 are the only two events found exclusively in IPF samples. Secondly, although many of the other fused genes, such as MFAP4=EPN2 and CCDC120=PIM2, also locate in the same chromosomes, IL7=AC083837.1 are the only pair of genes that have common promotor zones. Thirdly, compared with other genes, IL7, as an immune promoter, has a higher possibility to be involved in the imbalanced immune responses in IPF. Noteworthy, the level of IL-7 in serum has been reported to be associated with an increased survival rate in IPF patients.

Reviewer's #1 comment 2, major revision

2) It is not clear the number of genes composing the list you tested in the expression correlation analysis. Probably type 1 errors in null hypothesis testing is needed in the last table (pag. 53-57). Speaking about that table, it is almost impossible to read spreading across 6 pages and lacking of legend. Change the format, please.

A: Thanks for pointing this out and sorry for the confusion. This table shows the correlations between the expression of fused genes (IL7 and AC083837.1) and the 55 known IPF feature genes we collected from public databases. We changed the format of the tables and merged three columns into one to make it clearer. The table now shows the most related results of the Pearson correlation coefficient from our study and the results calculated by the dataset in GeneFriends.org.

About the question on the type I error, the null hypothesis for the test is that the correlation coefficient between the expression of each pair of genes is 0. The type 1 error set for the significance level is 0.05.

We have also added a legend to these tables to explain the results and the test method.

Reviewer's #1 comment 3, major revision

3) Similarly, you declared that your investigation highlighted 1550 fusion events. Thus, type 1 errors in null hypothesis testing value should be reported in the table 1 as well.

A: Thanks a lot for this excellent comment. We have now added the statement on type 1 error in table 1 as suggested.

Reviewer's #1 comment 4, major revision

4) Some potentially interesting results are inaccessible. In particular, there is no trace of the 282 differentially expressed gene list (not even in the supplementary material). In addition, the figure 5 and 6 are too dispersive that resulted as not-informative at all. Did you consider to report the same results in a table format or to "zoom in" in the area of interest of the matrix?

A: Thanks for pointing these problems out. In the previous submission, we decided not to submit the results of DE analysis and other analyses because they would take thousands of pages. We understand that these 282 DE genes are of great importance to this study, so we provided them in a new supplementary file.

For Figures 5 and 6, we wanted to show all the significant gene sets in the previous submission. Thanks for pointing the problem out, we exported the data, extracted the gene sets of interest, and replot them.

Reviewer's #1 comment 5, minor revision

- Approximation of the number at second or third decimal will be appreciated, as well as, the use of exponential number format in case of low values.

A: Thanks for pointing this out. We changed the format of all numbers in our paper according to your suggestion.

Reviewer's #1 comment 6, minor revision

- Supplementary figures lack of relative legend.

A: We added legends to supplementary figures.

Reviewer #2

Reviewer's #2 comment 1, major revision

General concern – Unfortunately, the manuscript is poorly written, making it very difficult to understand the methods, results and discussion. Much of the text has incorrect grammar, contradicting statements and types (even in the title: "involves" should be "is involved"). The manuscript will need major editing and likely English proof-editing if available. Many of the statements made are repetitive and do not make sense. Overall the description of results and discussion lacks coherent speculation on the overall findings (e.g. the association of this fusion event with biological processes indicated by GO terms).

Overall this manuscript will need significant revision before it will be suitable for publication.

A:

- Thanks for pointing out the grammar error in the title. We have corrected it as suggested.

- As suggested, this manuscript was reviewed and revised by a colleague who is experienced in academic English writing to improve readability.

Reviewer's #2 comment 2, major revision

The authors also make statements such as “Our study shows that IL7=AC083837.1 gene fusion was associated significantly enhanced gene expression related to NK cell- mediated cytotoxicity, which, according to the latest studies, will further activate the fibrosis process and exacerbates patients’ respiratory impairment. We have little clue about the mechanism of this impact; thus, more studies in this area is required.”

It would be better to discuss particular genes which are related to NK-cell cytotoxicity and speculate on possible mechanisms which could contribute to the pathogenesis of IPF.

A: Thank you. According to the studies on this topic, the most important gene that is involved in the NK-cell cytotoxicity’s regulation of IPF is IFN-γ, the signaling pathway of which was also enriched in our analysis. We have already summarized all related studies about IFN-γ and IFN family in lines 815-840. To make a stronger point, we added a statement about the significant enrichment of IFN pathways in our GSEA analysis. “In our study, NK cell-mediated cytotoxicity, the signaling pathways and responses of interferon α, β, and γ were significantly enriched in IPF tissues with IL7=AC083837.1 gene fusion, indicating that they might be the pivotal pathways through which this gene fusion accelerates the exacerbation of IPF.”

Reviewer's #2 comment 3, major revision

It would be advantageous to try to demonstrate some functionality of the loss of IL-7 or IL-7R using an in vitro assay. For example, how NK-cell cytotoxicity could induce epithelial or endothelial cell apoptosis, as discussed in lines 489-506.

A: We agree that an in vitro study would strengthen the reliability of our findings. However, because our lab is a dry lab mainly focusing on bioinformatics analysis and data mining, we do not conduct wet-lab experiments to validate our findings. With an attempt to address this comment, we searched the literature again and found two more publications on the knockdown studies of IL7 and IL7R (i.e., References 69 and 70). Thus, we have now cited these two published knockdown studies to strengthen our findings.

Reviewer's #2 comment 4, major revision

In terms of the figures presented, there are a number of concerns, in most of the figures it is unclear what is actually being presented.

Figure 4 may be better suited to be listed as a table.

A: Thanks, this is done.

Reviewer's #2 comment 5, major revision

Figure 5 is difficult to understand what is actually being presented and the “blue” is not discernable.

A: Thanks. We replot the figure focusing on the genes and gene sets of great interest. As we only focus on the enriched pathways in tissues with gene fusion, we removed the blue (inhibited) gene sets from the figure.

Reviewer's #2 comment 6, major revision

From Figure 6 onward, it is unclear what actual data is being presented. Showing GO-term node connections like this does not inform the reader of much. Please consider presenting this data more clearly. Tables could be very helpful.

A: Thanks. Most related results for figure 6,7 and 8 were provided as supplementary tables.

Once again, we thank you for the time you put in reviewing our paper and look forward to meeting your expectations.