Peer Review File

Article information: http://dx.doi.org/10.21037/atm-20-3337

Responses to the comments of Reviewer A

Comment 1: Conclusions from the study are largely overstated and need to more directly reflect the findings of the study.

Reply1: Thank you for your suggestion. We have re-written the conclusions according to your suggestion.

Changes in the text: see Page 7, line 4-8, Page 32, line 8-16, Page 33, line 1-5.

Comment 2: The authors identify the hub genes as important regulators of disease but do not describe how the hub genes are themselves regulated. Hub gene selection is based primarily on intramodular connectivity scores, which does not necessarily indicate that these genes are the primary drivers of disease as they themselves are likely regulated by upstream genes involved in the enriched pathways. Therefore, it would be more accurate to identify these hub genes as relevant to disease. This issue should be discussed by the authors.

Reply2: Thank you for pointing out this issue. Just like what the you said, some “hub genes” may show high intramodular connectivity due to the regulation of their upstream hub genes. And these regulated “hub genes” may be significantly associated with the disease, but they are not the primary drivers of disease. We added this point in revised manuscript and modified the title of the article.

Changes in the text: Page 1, line 2-5, Page 31, line 3-10.

Comment 3: Abstract Background: “Viral myocarditis is a common cardiovascular disease”. Viral myocarditis is not a “common” cardiovascular disease. While it does affect many individuals, the number of affected individuals per year is lower than the criteria for the definition of a “rare” disease.

Reply3: Thank you for your suggestion. I am sorry for this mistake, and we have modified our text.

Changes in the text: see Page 5, line 5, Page8, line 2.

Comment 4: Page 6-First and 2nd paragraph: “However, a subset of patients experience explosive myocarditis.” Please provide what is meant by “explosive myocarditis" as this is not a term typically used to describe myocarditis.

Reply4: Thank you for pointing out this issue. I am sorry for our incorrect description here, and we have modified our text.

Changes in the text: see Page 8, line 7, Page 9, line 2.

Comment 5: Introduction Page 7-Third line: The author states that the “turquoise module significantly correlates with the acute stage of viral myocarditis, while the brown and yellow modules significantly correlate with the chronic stage of viral myocarditis”. Briefly introduce how the different modules are related to the various stages of myocarditis, or how these modules are comprised. It is not clear from this description what genes or networks are being discussed.

Reply5: Thank you for your suggestion. In WGCNA, genes with similar expression patterns were clustered into the same modules. And the relationships between different stages of the myocarditis and modules were calculated to identify highly related modules. Among them, turquoise module significantly correlates with the acute stage of viral myocarditis, while the brown and yellow modules significantly correlate with the chronic stage of viral myocarditis. We have modified our text as advised.

Changes in the text: see Page 9, line 11-16.

Comment 6: Materials and Methods, Data collection: Page 8-first paragraph, third line: PBS does not “infect” mice.

Reply6: Thank you for this advice. I am sorry for our incorrect description, and we have modified our text as advised.

Changes in the text: see Page 11, line 3.

Comment 7: Materials and Methods, Data preprocessing: Page 8-second paragraph, 4th line: Please explain how these missing values came to be, is this because of the source of the data? If so, is it common practice to simply use an algorithm to impute nearest neighbor values in the context of bulk-microArray datasets? How many values were missing from the dataset? Why not simply exclude missing values? Describe these methodological issues in more detail.

Reply7: Thank you for your comments. In our study, we used the impute.knn function in R software to impute the missing values of the expression matrix.

Impute.knn is a function in the impute package version 1.58.0, which uses k-nearest neighbors to impute the missing expression values. The function used the Euclidean metric to find k nearest genes that were similar to the expression profiles of each gene with missing value, and estimated the missing value through the expression values of nearest genes. There are individual missing values in the gene expression matrix obtained from the raw data. If we simply exclude the missing values, it may influence the subsequent analysis. Because the subsequent analysis algorithms require a complete matrix of gene array values as input. Therefore, we used impute.knn function to impute the individual missing values in the gene expression matrix. We have described this method in more detail in the manuscript.

Changes in the text: see Page 11, line 10-15.

Comment 8: Why is selection of significant modules based on the correlation coefficient rather than the p-value assigned to the correlations? Although the criterion of .9 and -.9 as cutoff for significance is fine, it seems somewhat arbitrary considering a positively and negatively correlating module is utilized for the chronic timepoint whereas only a positively correlating module is utilized for the acute timepoint. Is there a reference that would justify the criterion for selection of significant modules in this manner?

Reply8: Thank you for this important concern. It is generally believed that the higher the correlation coefficient between the module and the traits, the more important the modules are to the traits. And the p-value is used to test whether the correlation coefficient is statistically significant. In previous studies, the modules with the highest correlation or the modules with the highest positive and negative correlation have been used for further analysis (Wang Tao,Zheng Xuan,Li Ruidong et al. Integrated bioinformatic analysis reveals YWHAB as a novel diagnostic biomarker for idiopathic pulmonary arterial hypertension.[J] .J. Cell. Physiol., 2019, 234: 6449-6462). In our study, our aim was to find the underlying genes and mechanisms most related to the occurrence and development of viral myocarditis. And we believed that the modules whose correlation coefficient was greater than 0.9 or less than-0.9 showed a strong correlation with the disease Therefore, we selected modules with a correlation coefficient greater than 0.9 or less than-0.9 in both acute and chronic disease stage.

Changes in the text: None.

Comment 9: Figure 3 and Figure 4 do not clearly indicate the time points for each module (acute vs chronic).

Reply9: Thank you for your comment. Because the enrichment results of a module were unchanged whether in the acute disease stage or the chronic disease stage, we did not indicate the time points in the enrichment results. In the weighted gene co-expression network analysis, genes with similar expression patterns are clustered into a module. Genes in the same module are fixed. Due to the different correlation between each module and different stages of the disease, the results of module enrichment have different significance for different time point. For example, the genes in the turquoise module showed a high correlation with the acute disease stage, but not the chronic disease stage, indicating that the results of enrichment play an important role in regulating the acute viral myocarditis, but have no special significance for the chronic viral myocarditis. As for figure 3, it shows the expression of genes in key modules in each sample. Samples A1 to A6 represent acute disease stage. Samples C1 to C6 represent chronic disease stage.

Changes in the text: None.

Comment 10: The authors do not describe the restriction parameters of the gene enrichment analysis performed using GO and KEGG – i.e., Were pathways available for enrichment analysis restricted to gene sets containing larger than #?genes and fewer than #?genes? If not, please justify how this would not lead artifacts in the data results and conclusions derived from the enrichment analyses.

Reply10: Thank you for pointing out this issue. According to your comments, we found that there were deficiencies in our enrichment analysis, which may lead to inaccurate enrichment results. Therefore, we carried out a new enrichment analysis on key modules. The enrichGO and enrichKEGG function from the clusterProfiler package were, respectively, used to perform GO and pathway enrichment analysis. We used the Benjamini-Hochberg (BH) method to correct the p value of the enrichment terms, and set the adjusted p-value cutoff to 0.05, q-value cutoff to 0.2. An adjusted p-value < 0.05 was considered to be significant, and the identified significant analyses were sorted by gene counts. Figures 4 and Figures 5 were merged and modified.

Changes in the text: see Page 6, line 11-12, Page 12, line 14-15, Page 13, line 1-3, Page 18-19, Page 26, table1-4, Figure 4.

Comment 11: The authors state that they found certain pathways to be more and less enriched across the dataset and within certain modules; however, it is not described how the authors determined and ranked enrichment. Were gene sets given enrichment scores by GO and KEGG? Did the authors utilize a t-statistic in combination with the number of terms counted as hits in the gene set enrichments – if so, how were these weighted in a calculation? Please justify and explain how enrichment was determined and the significance of the order in which the pathways are listed in tables 1-3.

Reply11: Thank you for pointing out this issue. The clusterProfiler package offers a gene classification method, to classify genes based on their projection at a specific level of the GO corpus, and provides functions, enrichGO and enrichKEGG, to calculate enrichment test for GO terms and KEGG pathways based on hypergeometric distribution. To prevent high false discovery rate (FDR) in multiple testing, q-values are also estimated for FDR control. In our re-analysis, we used the Benjamini-Hochberg (BH) method to correct the p value of the enrichment terms, and set the adjusted p-value cutoff to 0.05, q-value cutoff to 0.2. An adjusted p-value < 0.05 was considered to be significant, and the identified significant analyses were sorted by gene counts.

Changes in the text: see Page 6, line 11-12, Page 12, line 14-15, Page 13, line 1-3, Page 18-19, Page 26, table1-4, Figure 4.

Comment 12: It is not clear why the authors didn't utilize KEGG in the most significant modules at both the acute and chronic time points.

Reply12: Thank you for your comment. We have utilized KEGG enrichment analysis in the key modules (the results are shown in table 4). Because the enrichment results of a module were unchanged whether in the acute disease stage or the chronic disease stage, we did not indicate the time points in the enrichment results.

Changes in the text: None.

Comment 13: The tables do not clearly indicate the time points (acute vs chronic) for the enrichment analyses.

Reply13: Thank you for your comment. Because the enrichment results of a module were unchanged whether in the acute disease stage or the chronic disease stage, we did not indicate the time points in the enrichment results.

Changes in the text: None.

Comment 14: Discussion P18, halfway through first paragraph: The authors state "Neutralizing antibodies appear around days 4 post infection and play critical roles in limiting further viral replication in the heart," but neutralizing (IgG) antibodies do not occur until about 1 week after infection. Correct this statement.

Reply14: Thank you for your suggestion. I am sorry for our inaccurate description, and we have modified our text as advised.

Changes in the text: see Page 23, line 6.

Comment 15: Discussion P19, beginning at line 3: The authors state that "Previous studies show that Itgb2 may be involved in the regulation of B. burgdorferi induced carditis and autoimmune carditis (28,29). Therefore, Itgb2 may modulate immune responses and inflammatory regulation in the early stage of viral myocarditis." Provide a better explanation for why a gene that protects against a bacterial infection will also protect against a viral infection. More information about the gene in relation to the disease needs to be described.

Reply15: Thank you for pointing out this issue. I am sorry for our inaccurate description. The role of Itgb2 in viral myocarditis needs to be further verified and we have modified our text.

Changes in the text: see Page 24, line 4-5.

Responses to the comments of Reviewer B

Comment 1: The discussion section, although interesting, should be shortened in order to be more easy to follow.

Reply1: Thank you for your suggestion. We have modified our text as advised.

Changes in the text: see Page 22-33.

Comment 2: The conclusion is unbalanced. It should be rewritten and shortened focusing on the clinical implications of the study

Reply2: Thank you for this advice. We have re-written the conclusions according to your advice.

Changes in the text: see Page 7, line 4-8, Page 32, line 8-16, Page 33, line 1-5.