Peer Review File
Article information: https://dx.doi.org/10.21037/jtd-23-72

Reviewer A

Dr Lu and collaborators revisited the bioinformatics analysis of GEO dataset from calcified aortic valve patients. They completed this analysis by a single cell RNA-seq from RNA extracted from aortic valve of CAVD patients. They further used machine learning to look for biomarkers of the disease. They found out a gene CTHRC1 specifically expressed in a subpopulation of VICs enriched in fibroblastic cells.

The paper is potentially interesting but is difficult to review. The legends of figures are quite short and not very informative. The lettering is very small which does not help in the understanding of the figures.
Major points:
Comment 1: It is unclear what are the control samples used to be compared with the CAVD samples.

Reply 1: Thank you for your comment. We have modified our text as advised (see Page 6-7, line 94-98).

Comment 2: In figure 4 what are the Tsne plots in A left and right? The same question applied to the trajectory inference plot in B right vs left?

Reply 2: T-SNE visualizes high-dimensional similarities of cells in an easily understandable 2D or 3D scatter plot, the so-called t-SNE map. The proximity of cells in the t-SNE map reflects their distances in the high-dimensional space. Cells that are similar in their analyzed protein-expression pattern will be located closely together in the t-SNE map, thus enabling the visualization of different cellular subpopulations. Importantly, t-SNE has been shown to successfully identify small cellular subpopulations, as low as those comprising 0.25%.

Comment 3: The Tsne plots do not separate well the different clusters; A 3D tsne 1, 2 and 3 should be shown. How many PC were used to get the Tsne plots? How were the data normalized?

Reply 3: The plots separate the different clusters, but due to the amount of cell subsets, the colors may not be very distinct from each other. The data normalization and PC have modified as advised (see Page 8, line 124-133).

Comment 4: What are the genes used to identify the cell clusters? More specifically what are the genes used to discriminate between the immune cell subpopulation.

Reply 4: If a gene is specifically expressed in a particular type of cell, it can be used to define that type of cell. The way to discriminate between the immune cell subpopulation have modified as advised (see Page 9, line 153-159).

Comment 5: A Violin plot should be shown to visualize expression of CTHRC1 in the different cell clusters.

Reply 5: The expression of CTHRC1 in the different cell clusters is as follows. There was no significant expression of CTHRC1 in cell clusters not listed in the table.

cluster	p_val_adj	avg_logFC	pct.1	pct.2
4	6.7269E-77	0.48274557	0.542	0.154
6	3.4341E-20	0.64607972	0.383	0.383

Comment 6: CTHRC1 has been reported by Joy Lincol's lab to be involved in the reparation of the endothelial leayer of valve (Nordquist et al ATVB, 2021) . How do the authors conciliate these data with the pro-calcification effect that they report?

Reply 6: We have modified our text as advised (see Page 15-16, line 289-296).

Comment 7: On the other hand the lab of Elena Aikawa reported a role of CTHRC1 in VIC calcification (Decano et al Cell Reports 2022 ;39(2):110685). The authors should quote and discuss this publication.

Reply 7: We have modified our text as advised (see Page 14-15, line 270-274).

Reviewer B

The manuscript by Lu et al., details an interesting in silico analysis of potential biomarkers of aortic valve disease. The authors study three microarray datasets and find CTHRC1 to be a potential biomarker for CAVD. Additionally, single cell analysis shows that this gene is highly expressed in a VIC subset related with the calcification process and that several immune cell types are correlated with the expression of this particular gene.

Manuscript is very interesting; however, it is very complex and difficult to follow, it requires further clarification and description of the variables studied. Most figures are not self-explicative, many graph axes lack of information and figure legends are too succinct. The correlation analysis does not describe what type of variable are the authors comparing when they say that immune subsets were correlated with CTHRC1 gene expression, are they talking about cell counts? Frequencies? In the correlation graphs, what do the X and Y-axes number represent, mRNA level, cells? There is no info regarding this. Also, when they refer to the gene expression, are they referring to the total expression of the gene in the sample? The expression of that gene in a particular cell subset?

In general, I find that the results and materials are difficult to follow because important information is lacking. Another example is the immune cell subsets definitions, when authors refer to activated subsets, this definition is based on which markers? How are each of these immune cell subsets defined?

Some paragraphs are quite difficult to read as the English quality is low and grammar is not correct. Manuscript language should be extensively edited by a professional editing service.

Comment 1: Another example is the immune cell subsets definitions, when authors refer to activated subsets, this definition is based on which markers? How are each of these immune cell subsets defined?

Reply 1: Thank you for your comment. The way to discriminate between the immune cell subpopulation have modified as advised (see Page 9, line 153-159).

Comment 2: The correlation analysis does not describe what type of variable are the authors comparing when they say that immune subsets were correlated with CTHRC1 gene expression, are they talking about cell counts? Frequencies? In the correlation graphs, what do the X and Y-axes number represent, mRNA level, cells? There is no info regarding this.

Reply 2: The method of correlation analysis is to test the correlation between the expression of our genes and the content of immune cells. If the P value of correlation test is less than 0.05, it means that the expression of our target gene is correlated with the content of immune cells. By checking the correlation, we can get graphs like the figures (figure 6). In these figures, the horizontal axis is the name of the gene, and the vertical axis is the name of the immune cell. Each dot represents each sample, and each sample has its gene expression and its immune cell count. And through these points we can simulate a phase where if this line goes up from left to right, there's a positive correlation between the expression of this gene and the amount of this immune cell, and if it goes down from left to right, there's a negative correlation between the amount of this immune cell and the expression of this gene.

Comment 3: Most figures are not self-explicative, many graph axes lack of information and figure legends are too succinct.

Reply 3: The caption of the picture has been added.