Peer Review File

Article information: https://dx.doi.org/10.21037/atm-22-3761

Reviewer A

The paper titled "Identification of candidate biomarkers and mechanisms in foam cell formation from heterogeneous cellular origins via integrated transcriptome analysis" is interesting. The results provide a comprehensive landscape of the transcriptional regulations in macrophages and VSMCs under the ox-LDL treatment from a bioinformatics respective, which may contribute to a better understanding of the pathophysiological mechanisms of foam cell formation. However, there are several minor issues that if addressed would significantly improve the manuscript.

1) What are the different VSMC plaque phenotypes? What are their presumed cellular and paracrine functions? What are the regulatory mechanisms controlling VSMC plasticity? What are their effects on atherosclerosis and plaque stability? It is recommended to add relevant contents.

Reply 1: Thank you for your suggestions. VSMCs are the main cell type in early arterial intimal thickenings and a major component of most stages of human atherosclerosis. VSMCs can express scavenger receptors and become foam cells on exposure to lipoproteins. We have added relevant content in the Introduction part, and we also cite some review articles with detailed descriptions of underlying biological mechanisms (ref no.10 and 11). Additionally, foam cell formation depends on the cell's ability to release excess cholesterol, the rate-limiting component of which is mediated by the membrane lipid transporter ATP-binding cassette transporter A1 (ABCA1), which was also reflected in our results and described expandedly in the Discussion part (paragraph 4).

Changes in the text: Introduction part, paragraph 2.

2) All figures are not clear enough. It is recommended to provide clearer figures again.

Reply 2: We have re-uploaded the high-resolution figures, and we can provide PDF versions if required.

Changes in the text: none.

3) This study is based on bioinformatics analysis. It is recommended to increase in vivo and in vitro experimental studies, which may be more meaningful.

Reply 3: Thank you for your comments. We agree with you on the experimental validation will bring more confidence to the findings. However, with the current resources and data available, we are unable to carry out the required work, so we fully explained the shortcomings of the current study in the discussion section of the article. Nevertheless, we included an external cohort from GSE9874 to further validate the reliability of selected DEGs, and we also combined the current published literature to discuss the possible mechanisms in the discussion section. Our study has the merit of screening a number of statistically significant biomarkers through a series of bioinformatic analyses. Of course, our results are preliminary and it is hoped that in the future other researchers

will be interested in conducting corresponding molecular biology studies and validation of large cohorts based on our results.

Changes in the text: Discussion part, paragraph 6.

4) How to guide future experimental research and clinical transformation based on the results of this study? It is recommended to add relevant contents.

Reply 4: Thank you for your comments. In the present study, we comprehensively performed a series of bioinformatics analysis to investigate and compare the DEGs and their related biological pathways in both human macrophages and VSMCs under the treatment of ox-LDL, as well as applied machine learning to evaluate and select the candidate biomarkers statistically and further validated in an external cohort, which may provide relevant transcriptional signatures and potential therapeutic targets in atherosclerosis development. Particularly, we highlighted the ABCA1-ANXA1 interactions, ferroptosis and *BTG2*-related pathways, and *TNFSF4*-related pathways in the discussion part with published literature to suggest their feasibility in future experimental research and clinical transformation.

Changes in the text: Discussion part, paragraph 3-5.

5) This study is only a preliminary screening of key regulatory genes. It is more meaningful to increase the function of key regulatory genes.

Reply 5: Thank you for your comments. In addition to the discovery of DEGs that are closely associated with foam cell formation, we also applied gene ontology and pathway enrichment via three databases to investigate predominant changes in functional pathways of DEGs. Additionally, to comprehensively analyze the molecular mechanisms and functional versatility of the common DEGs in foam cell formation, we constructed the PPI network based on the topological features acquired from STRING database, suggesting several multi-cellular biomarkers and candidate pathways that might participate in the initial atherogenic events, which holds considerable promise to emerge capability to ascertain corresponding mechanisms and offer projections into developing new approaches for potential therapeutic modulation and cardiological practice.

Changes in the text: None.

6) Formation of foam cell macrophages, which sequester extracellular modified lipids, is a key event in atherosclerosis. How lipid loading affects macrophage phenotype? It is recommended to add relevant contents.

Reply 6: Thank you for your comments. Similar to the VSMCs mentioned above, ox-LDL accumulation is one of the essential processes responsible for atherogenesis and foam cell formation within the lipid-rich subendothelial space of the affected artery. This theory is wildly accepted and briefly introduced in the beginning of the manuscript. The aim of this study is to investigate the potential mechanisms of foam cell formation during the initial stage of atherogenesis, as well as identify the relative contributions of heterogeneous foam cell populations with shared functions,

which may hold significant feasibility of foam cell specific-treatments in the future. The detailed information about ox-LDL treatment was added in the Materials and methods section.
Changes in the text: Materials and methods section, microarray datasets collection.

7) What is the value of integrated transcriptome analysis in exploring pathological pathways? What is the biggest challenge facing? It is recommended to add relevant content.

Reply 7: Thank you for your comments. Transcriptome analysis has become a compelling approach to unravel the dynamic expression of genes and capturing cell physiology and molecular mechanisms holistically since its rapid development and adoption. In the present study, we comprehensively performed a series of bioinformatics analysis to investigate and compare the DEGs and their related biological pathways in both human macrophages and VSMCs under the treatment of ox-LDL, which allowed the integration of available data from relevant studies and may become potential targets for interventions in the initial phase of atherosclerotic process through both macrophages and VSMCs simultaneously. The biggest challenge lies in interexperiment variabilities such as different experimental procedures and different array platforms from different laboratories. Rather than simply merging the datasets, we analyzed them separately and independently after rigorous data testing before taking the intersection and performing external validation to minimize the impact of these confounding factors on the results.
Changes in the text: Introduction part, paragraph 2; Discussion part, paragraph 6.

8) What is the role of low-density lipoprotein in smooth muscle cell mediated foam cell formation? It is recommended to add relevant contents.

Reply 8: Thank you for your comments. Many research have supported the concept that oxidized-LDL (ox-LDL) accumulation is one of the essential processes responsible for atherogenesis and foam cell formation, and the datasets we included in this study were all high-throughput gene expression under ox-LDL treantment. We have added relevant contents in the Introduction part, and we also cite some review articles with detailed descriptions of underlying biological mechanisms (ref no.10 and 11).
Changes in the text: Introduction part, paragraph 2.

Thank you very much for your constructive comment. With your suggestions, we hope that the revised version will make the structure of our manuscript more complete and at the same time increase its novelty.

Reviewer B

The reviewed manuscript entitled ‘Identification of candidate biomarkers and mechanisms in foam cell formation from heterogeneous cellular origins via integrated transcriptome analysis’ written by Jing Xu and Yue-jin Yang presents interesting results concerning the transcriptomic differences

between the formation of foam cells from macrophages and vascular smooth muscle cells. The authors identified and investigated differentially expressed genes that could contribute to atherogenesis and serve as biomarkers. The article is well structured, scientifically sound, and contains informative visualisations. I have only minor comments on this manuscript.

1. In the Abstract section, a morphological description of foam cells and their role in the process of atherosclerotic plaque formation could be extended to make the study background more complete.

Reply 1: Thank you for your constructive comment. Taking into account the word limit of the abstract, we have modified it with your suggestions.

Changes in the text: Abstract.

2. If there are differences between ox-LDL treatment procedures used in the GSE54666 and GSE68021 studies, they should be mentioned.

Reply 2: Thank you for your suggestion. We have added the detailed treatment procedures in the Materials and methods section. Additionally, more information could be acquired from ref no.17 and no.18.

Changes in the text: Materials and methods section, microarray datasets collection.

3.Radar plots in Figure 2 are very informative, but genes regulated in the up and down order are mixed. It could be more convenient for readers to analyse these plots if groups of up- and down-regulated genes would be gathered together.

Reply 3: Thank you for your suggestion. The radar plots were produced via Omicshare website tools (https://www.omicshare.com/tools/Home/). We tried to modify the image layout as you suggested, but the current version does not support the operation. Since the radar image shows only a small part of the DEGs, we think the layout of the images does not affect the display of relevant information. In addition, detailed information of all the DEGs from GSE54666 and GSE68021 is available in Table S1 and S2.

Changes in the text: None.

I believe that my suggestions will be helpful to the authors in increasing the quality of the reviewed manuscript.

Reviewer C

This is an interesting review about the main pathways involved in foam cell formation using transcriptome information.

The main concerns are the following:

The revision has been concentrated in oxidized LDL as the major driver atherogenic lipoproteins involved in foam cell formation. However, there are other atherogenic lipoproteins mainly present

in the human arterial intima that have not been considered in this revision. This makes this review lose originality and timeliness.

Thank you for your comments. Based on the available datasets and the predominant role of ox-LDL in atherosclerosis, we aimed to investigate and compare the DEGs and their related biological pathways in both human macrophages and VSMCs under the treatment of ox-LDL, which may become potential targets for interventions in the initial phase of atherosclerotic process through both macrophages and VSMCs simultaneously. Particularly, we highlighted the ABCA1-ANXA1 interactions, ferroptosis and *BTG2*-related pathways, and *TNFSF4*-related pathways in the discussion part with published literature to suggest their feasibility in future experimental research and clinical transformation. Our study has the merit of screening a number of statistically significant biomarkers through a series of bioinformatic analyses.

Other methodological concerns include:

1)the utilization of an unusual cutoff value for logFC. The more suitable cutoff value for logFC is between 1.5 and 2, and the p value < 0.05.

Reply 1: Thank you for your comments. In the first step of DEGs analysis, we used fold change more than 1 as cut-off value in order to filter out as many DEGs as possible, so that more convergent DEGs could be available in the next step. Additionally, even small differences in expression may have physiological significance, and according to some published literature, this cut-off is considered acceptable (PMID: 35071467, 34925639, 30421875, 24103890, 34617063, 33069778, 32866857, etc.).

Changes in the text: None.

2)Why the authors use for the analysis only the top 5 under and over expressed instead of the overlap

Reply 2: Thank you for your comments. Of all the DEGs, there were 34 up-regulated (Figure 3A) and 26 down-regulated (Figure 3B) DEGs overlapped in the two datasets (Table S5). Next, we investigated the functional enrichment of these 60 overlapped DEGs, and despite the differences in the most significant pathways in heterogeneous cellular origins, as shown in Figure 3C, there were still several representative and common functional pathways. To further narrow down the target DEGs, the top 5 up-regulated and down-regulated DEGs were selected as candidate biomarkers based on their FC in both datasets, which may be potential sensitive biomarkers for detecting disease progression.

Changes in the text: None.

3)Why then the authors take the top 20? And then to build the PPI network all??

Reply 3: Thank you for your comments. The top 20 DEGs you mentioned are the top DEGs in each dataset respectively. In order to capture the variations in cellular transcript levels in response to ox-LDL treatment from a holistic perspective, we firstly analyzed the DEGs in VSMCs-derived and macrophages-derived foam cells separately. For macrophages, the predominant variations in

functional pathways were mainly enriched in terms related to lipid metabolism. In contrast, the annotated biological process based on the DEGs in VSMCs showed tight regulation of immunoinflammatory responses. The analyzing procedures are supported by the official guidelines of DAVID database (ref no.21). The PPI network was further built based on the overlapped DEGs in both datasets to comprehensively analyze the molecular mechanisms and functional versatility.

Changes in the text: None.

4)Why the authors use a z score in figure 2, and not directly the fold change?

Reply 4: Thank you for your comments. As mentioned in the manuscript, the samples included in this study are from different datasets, which may induce interexperiment variabilities such as different experimental procedures and different array platforms from different laboratories. Rather than simply merging the datasets, we analyzed them separately and independently after rigorous data testing before taking the intersection and performing external validation to minimize the impact of these confounding factors on the results. Therefore, the Euclidean genetic distance combining complete distance hierarchical cluster method was used for cluster analysis based on Z score transformation.

Changes in the text: None.