Identification the key genes between polycystic ovary syndrome and Non-Alcoholic Fatty Liver Disease

Pu Yifu (puyifu150695@163.com)  
Sichuan University - Wangjiang Campus: Sichuan University  
https://orcid.org/0000-0002-6297-9184

Research Article

**Keywords:** Bioinformatic analysis, Polycystic ovary syndrome, Non-alcoholic fatty liver disease, Gene

**Posted Date:** December 8th, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-1145395/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Purpose Many studies show correlation between polycystic ovary syndrome (PCOS) and non-alcoholic fatty liver disease (NAFLD), but the underlying pathogenic genes are not clear. This study by using the bioinformatics method aims to search the key genes involved in these 2 diseases. Methods The Gene Expression Omnibus (GEO) datasets coming from the GEO database GSE63067 -NAFLD patients and healthy controls, and GSE34526 -PCOS patients and normal controls, are downloaded. Differentially expressed genes (DEGs) of 2 diseases datasets and the common genes are obtained. After GO and KEGG enrichment analyses of common genes are performed. To find the key genes between NAFLD and PCOS, a protein–protein interaction (PPI) network is carried out. In addition, the diagnostic value of key genes in PCOS is analyzed. Results According to NAFLD and PCOS downloaded datasets, 34 common genes, 21 key genes, 15 GO terms and 4 KEGG pathways are obtained. Further, based on the top 6 key genes, the corresponding area under the curve (AUC) by constructing ROC curves in the PCOS is 0.909 (95% CI, 0.775–1.000). Conclusions The study identify some key genes in the occurrence and progression between NAFLD and PCOS. In the future, to verify our results, it need experimental and clinical research.

Introduction

Polycystic ovary syndrome (PCOS) is the most common endocrine disorder in reproductive age women[1, 2]. Both genetic and environmental factors play vital roles in the occurrence and development of PCOS[3]. PCOS patients always present central obesity, dyslipidemia, inflammatory and insulin resistance (IR) which seem to be the pathogenesis[4]. Meanwhile, another common metabolic syndrome-Non-alcoholic fatty liver disease (NAFLD) is multiple definite causes, including environmental and genetic factors-presenting dyslipidemia, dietary rhythm, and insulin resistance, inflammatory factors and so on[5]. With the high prevalence of NAFLD[6, 7] and PCOS[8, 9] and both pathogenesis being similarly, some studies focus on the PCOS and NAFLD patients. A study[10] proves that PCOS is associated with more severe nonalcoholic steatohepatitis (NASH), including advanced fibrosis. Others find the common risk factors in both disorders including central obesity, insulin resistance, chronic inflammation, and hyperandrogenemia[11–16]. Besides, two studies pay adolescents and/or children and consider the relationship between both diseases[17, 18]. Some reviews summarize the pathophysiological mechanisms of PCOS and NAFLD and discuss the diagnostic approaches and management options for two diseases[19–24]. Research related the relationship between 2 diseases has progressed steadily, however, the research related genetics is lacking of exploration. Therefore, to identify new targets of two diseases and explain the relationships between NAFLD and PCOS, the Gene Expression Omnibus (GEO) database gene expression profile of NAFLD and PCOS are downloaded to make bioinformatics analysis to obtained molecular mechanisms.

Materials And Methods

Database

The GSE63067 dataset, coming from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo) - a free open-access database, contains the gene expression profiles of 11 NAFLD patients and 7 non-NAFLD controls. Meanwhile in the GSE34526 dataset, 7 PCOS patients and 3 healthy controls are obtained to analysis. This study do not need Ethics Committee approval. Data is download by R Software (version 3.6.3) R package - GEOquery package [2.54.1 version][25].

Differentially expressed genes (DEGs) in NAFLD and PCOS

DEGs, for NAFLD and normal controls, PCOS patients and normal controls, are identified by using the limma R package[3.42.2 version][26]. P-value <0.05 and |log2FC| >1 in two datasets are set as selected criteria. 2 sets of DEGs are found according to criteria, then these DEGs being from the 2 diseases are analyzed by R (version 3.6.3) R package: ggplot2 package. Common genes obtained are used for GO and KEGG and others.

Geneontology (GO) and pathway enrichment analyses

DEGs functions—biological processes (BP), cellular component (CC) and molecular function (MF) and KEGG pathway enrichment are analyzed by R software (version 3.6.3) ggplot2 package (3.3.3 version) and cluster profiler package (3.14.3 version) [27].

PPI network and key genes

To explore the interaction among the common genes of 2 diseases, STRING 11.0 (http://string-db.org/) and Cytoscape 3.8.2 (https://cytoscape.org) are used for constructing a PPI network[28, 29]. The nodes represent the proteins and the lines show the interactions between proteins. Then we find the key genes from node1 STRING data by excel and represent them according to the number of proteins in the form of histogram.

Diagnostic efficiency of top 6 genes in PCOS

By using R software (version 3.6.3) pROC package, we analysis receiver operating characteristic (ROC) curves and calculate the area under the curve (AUC) of the top 6 genes in the PCOS. These result showed the diagnostic efficiency of genes for PCOS.

Results

Study process is showed in Figure 1.
Differential expression analysis in NAFLD and PCOS

According to \(|\log_{2}(FC)| > 1\) and \(p\text{-value}<0.05\), in dataset GSE63067, 125 DEGs, 112 highly expressed in NAFLD group (11 samples; group2) and 13 highly expressed in normal group (7 samples; group1), are filtered. In datasets GSE34526, 1692 DEGs, 1148 highly expressed in PCOS group (7 samples; group2) and 544 highly expressed in normal group (3 samples; group1), are filtered. Heatmaps and Volcano Plots of PCOS and NAFLD, which shown the gene expression profiles, are presented in Figures 2 and figure 3. For the Venn Diagram, 34 intersecting common genes coming from 2 diseases are obtained and are shown in Figure 4. The Table 1 presents the details of common genes.

GO and KEGG

According to the P-value <0.05, 34 common genes coming from NAFLD and PCOS are identified functional enrichment and KEGG pathway analyses. Inflammatory and immune response (neutrophil mediated immunity, neutrophil activation, neutrophil activation involved in immune response, neutrophil degranulation, leukocyte migration) are found in GO biological processes (BP). Enrichment of cytoplasmic vesicle lumen, secretory granule lumen, external side of plasma membrane, secretory granule membrane and collagen–containing extracellular matrix are found in cellular component (CC). Moreover, receptor activity (C–C chemokine receptor activity, cytokine receptor activity) and binding-related function (organic acid binding, carboxylic acid binding, RAGE receptor binding) are found in the molecular function (MF) section. In particular, changes-leukocyte transendothelial migration, TNF signaling pathway, fluid shear stress and atherosclerosis, staphylococcus aureus infection are mostly enriched in the KEGG pathway. These results are presented in Figure 5.

PPI network and key genes

A PPI network is created to identify the key genes from the common genes, the result is represented in Figure 6. Figure 7 and Table 1 shows these detailed genes. Moreover, the top 6 genes are used for further research.

ROC: diagnostic value of top 6 genes

The top 6 hub genes to validate the diagnostic value for PCOS are obtained from the above analysis. ROC curves and the corresponding area under the curve (AUC) are calculated. The AUC of top 6 genes combination (MMP9, FPR1, S100A9, TREM1, ICAM1 and S100A12) in PCOS patients and normal controls is 0.909 (95% CI, 0.775–1.000). The detailed result of PCOS is represented in Figure 8.

### Table 1

<table>
<thead>
<tr>
<th>Common gene 34 (total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADAMTS1, ANXA1, CSAR1, CCR1, CD44, FCN1, FPR1, HAMP, HMOX1, ICAM1, IL1RN, IRR1, MMP9, MX2, MXD1, NCF2, PLAUR, S100A12, S100A9, TREM1, VNN2 (21 key genes)</td>
</tr>
<tr>
<td>ACKR3, ADM, AMID1, CCDC71L, FABP5, HLA-DRA, KIAA0040, mir-223, *NAMPT, SAMSN1, SGMS2, WNK3, PPP1R18</td>
</tr>
</tbody>
</table>

*mir-223: miRNA

### Discussion

NAFLD and PCOS, pathogenesis being similarly, are highly prevalent diseases in the world. In this study, to explore the molecular mechanisms in these 2 diseases and discover the early targets to prevent disease development, through searching the datasets of NAFLD and PCOS from GEO, we find 34 common genes including 1mirRNA-mir223. After we perform GO enrichment and KEGG pathway enrichment analyses and construct a PPI network, top 6 among 21 key genes found by above analysis represent a good diagnostic value in PCOS patients.

Matrix metalloproteinase-9 (MMP9) may play an essential role in PCOS and NAFLD. A study [30] proves that PCOS patients have elevated serum concentrations of MMP-2 and MMP-9. Another [31] proves that MMP9 may be involved in the pathogenesis of PCOS and matrix metalloproteinases (MMPs) play vital roles in follicular development. A recent study [32] makes a rat model of PCOS and proves that metformin treatment by decreasing the expression of MMP-2 and MMP-9 alleviates PCOS. Another [33] proves that MMP-9 levels are increased in obese PCOS women. Moreover, study [34] made a rodent model of non-alcoholic steatohepatitis present that MMP-9 activity is negatively correlates with adiponectin—which can protect from inflammation and fibrosis in metabolic liver disease. Besides, study [35] proves that MMP-9 deficiency enhances regeneration of steatotic livers. Study [36] proves that MMP9 levels can act as predictive factors for poor prognosis of nonalcoholic fatty liver patients.

Formyl peptide receptor 1 (FPR1) is powerful neutrophil chemotactic factors. Binding receptor stimulates intracellular calcium mobilization and superoxide anion release. A review— decoding cell death signals in liver inflammation [37], points out that FPR1 by mitochondrial damage, potent neutrophil activators and plasma membrane permeabilization causes liver inflammation. However, the study of the relationship between FPR1 and PCOS has not been found.
S100A9 plays a vital role in the regulation of inflammatory processes and immune response. It can induce neutrophil chemotaxis, adhesion and increase the bactericidal activity of neutrophils by promoting phagocytosis. A study [38] observes that PCOS follicular fluid exosomes contain S100-A9 protein which can enhance inflammation and disrupt steroidogenesis via activation of nuclear factor kappa B signalling pathway. The study of the relationship between S100A9 and NAFLD has not been found.

Triggering receptor expressed on myeloid cells 1 (TREM1) stimulates neutrophil, monocyte-mediated inflammatory responses and triggers release of pro-inflammatory chemokines and cytokines, as well as increase surface expression of cell activation markers. Study [39] proves that TREM-1 is upregulation at messenger RNA and protein levels in NAFLD model mice. The study of the relationship between TREM1 and PCOS has not been found.

Intercellular adhesion molecule 1 (ICAM1) is ligands for the leukocyte adhesion protein integrin alpha-L/beta-2. During leukocyte trans-endothelial migration, ICAM1 engagement promotes the assembly of endothelial apical cups. Study [40] proves that the patients with PCOS and NAFLD are heavier cardiovascular risk profile compared with patients with PCOS involving serum markers (ICAM-1). ICAM1 in NAFLD and PCOS need further exploration.

S100A12, which is similar to S100A9, may play a prominent role in inflammatory processes and immune response. It can act as an alarmin or a danger associated molecular pattern molecule and stimulate innate immune cells via binding to receptor for advanced glycation endproducts (AGER), which can activate the MAP-kinase and NF-kappa-B signaling pathways. It's worth noting that it needs further research to confirm in PCOS and NAFLD.

For mir-223, some researches focus on relationship between PCOS and mir-223 [41–47]. Moreover, others draw on relationship between NAFLD and mir-223 [48–56]. In this study, we find the mir-223 as the common gene between PCOS and NAFLD. Our research also indicated that mir-223 might become a potential target in PCOS and NAFLD treatment.

Besides, we performed gene biological functions and pathways, such as inflammatory reactions, NF-kappa B signaling pathway, which are closely involved in 2 diseases. Meanwhile there coexist advantages and disadvantages in this study. For disadvantages, the included samples are relatively small, key genes predicted need larger-sample, experimental and clinical research to verify. For advantages, we explore and analyze the top 6 key genes and 1 common gene. The AUC calculated by the top 6 key genes in the PCOS is 0.909 (95% CI, 0.775–1.000), showing a good diagnostic efficiency. Importantly, we point out the direction of future research for PCOS and NAFLD from molecular level.

Conclusions
In this study by bioinformatic analysis, 34 intersection genes and 21 key genes are obtained from 2 diseases. The top 6 key genes (MMP9, FPR1, S100A9, TREM1, ICAM1, S100A12) are selected and their diagnostic values are validated by R software. The result may provide potential targets for the prevention and treatment of NAFLD and PCOS.

Declarations
Acknowledgement
None.
Declaration of Competing Interest
None.
Ethics approval
The dataset, coming from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo) - a free open-access database, so ethical approval is not required.
Consent to participate
It is not required.
Availability of data and material
The dataset, coming from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo) - a free open-access database
Code availability
R software.
Authors’ contributions
PuYifu designed the study. PuYifu and HuKaiFeng finished the calculations and the writing.
Funding
None.

There are no conflicts of interest to declare.

This manuscript has not been published or presented elsewhere in part or in entirety and is not under consideration by another journal. We have read and understood your journal’s policies, and we believe that neither the manuscript nor the study violates any of these.

References


45. Zhao D, Huang X, Chang X, Yao Q, He Q (2019) MMP-9 activity is increased by adiponectin in primary human hepatocytes but even more so correlates with serum adiponectin in a rodent model of non-alcoholic steatohepatitis. Exp Mol Pathol 91(2):603.


47. Zhao D, Huang X, Chang X, Yao Q, He Q (2019) MMP-9 activity is increased by adiponectin in primary human hepatocytes but even more so correlates with serum adiponectin in a rodent model of non-alcoholic steatohepatitis. Exp Mol Pathol 91(2):603.


49. Zhao D, Huang X, Chang X, Yao Q, He Q (2019) MMP-9 activity is increased by adiponectin in primary human hepatocytes but even more so correlates with serum adiponectin in a rodent model of non-alcoholic steatohepatitis. Exp Mol Pathol 91(2):603.


Figures

![Figure 1](image)

**Figure 1**
study process
Figure 5

Legend not included with this version

Figure 6

PPI network
Figure 7

Key genes (21 total)

A

B

Figure 8

Legend not included with this version