Integrated Bioinformatics Analysis Reveals Potential Mechanisms Associated with Intestinal Flora Intervention in Non-alcoholic Fatty Liver Disease

Yingying Liu  
Beijing University of Chinese Medicine

Xinkui Liu  
Beijing University of Chinese Medicine

Wei Zhou  
Beijing University of Chinese Medicine

Jingyuan Zhang  
Beijing University of Chinese Medicine

Siyu Guo  
Beijing University of Chinese Medicine

Shanshan Jia  
Beijing University of Chinese Medicine

Jiarui Wu (✉️ exogamy@163.com)  
Beijing College of Traditional Chinese Medicine: Beijing University of Chinese Medicine  
https://orcid.org/0000-0002-1617-6110

Haojia Wang  
Beijing University of Chinese Medicine

Jialin Li  
Beijing University of Chinese Medicine

Yingying Tan  
Beijing University of Chinese Medicine

Research

Keywords: non-alcoholic fatty liver disease, intestinal flora disorder, target prediction, signaling pathway, bioinformatics

DOI: https://doi.org/10.21203/rs.3.rs-112110/v1
Abstract

Background

Non-alcoholic fatty liver disease (NAFLD) is a common chronic liver disease that imposes a huge economic burden on global public health. And the gut-liver axis theory supports the therapeutic role of intestinal flora in the development and progression of NAFLD. To this end, we designed bioinformatics study on the relationship between intestinal flora disorder and NAFLD, so as to explore the molecular mechanism of intestinal flora interfering with NAFLD.

Methods

Differentially expressed genes for NAFLD were obtained from GEO database. And the disease genes for NAFLD and intestinal flora disorder were obtained from the disease databases. Using string 11.0 database to establish protein-protein interaction network relationship and cytoscape 3.7.2 software visualization. Cytoscape plug-in MCODE and cytoHubba were used to screen the potential genes of intestinal flora disorder and NAFLD, so as to obtain potential targets for intestinal flora to interfere in the occurrence and process of NAFLD. Enrichment analysis of potential targets was carried out using R 4.0.2 software.

Results

The results showed that PTGS2, SPINK1 and C5AR1 may be the key genes for intestinal flora to interfere with NAFLD. CCL2, IL6, IL1B and FOS may be key genes for the development and progression of NAFLD. The gene function is mainly reflected in basic biological processes, including the regulation of metabolic process, epithelial development and immune influence. The pathway is mainly related to signal transduction, immune regulation and physiological metabolism. The TNF signaling pathway, AGE-RAGE signaling pathway in the diabetic activity, and NF-Kappa B signaling pathways are important pathways for intestinal flora to interfere with NAFLD.

Conclusion

According to the analysis results, there is a certain correlation between intestinal flora disorder and NAFLD. It is speculated that the mechanism by which intestinal flora may interfere with the occurrence and development of NAFLD is mainly related to inflammatory response and insulin resistance. Nevertheless, further research is needed to explore the specific molecular mechanisms.

Background
Non-alcoholic fatty liver disease (NAFLD) is becoming increasingly prevalent and rapidly becoming the leading cause of chronic liver disease worldwide. Around the world, 25.2% of the population suffers from NAFLD, and with the increase in the prevalence rate, NAFLD remains the most common liver disease in the world [1, 2]. NAFLD mainly includes simple steatosis (SS), non-alcoholic steatohepatitis (NASH), and its associated cirrhosis (NASH cirrhosis). NAFLD is characterized by liver fat accumulation of more than 5% in the absence of any secondary cause [2, 3]. NASH is a kind of liver cell damage caused by necrotizing inflammation under the premise of liver steatosis, which can easily develop into cirrhosis and liver cancer. NAFLD is a major cause of chronic liver disease. Under the influence of diet, lifestyle and other factors, the incidence rate increases year by year, and the probability of liver cell progression from simple steatosis to hepatitis, cirrhosis and hepatocellular carcinoma increases [4, 5].

Intestinal flora is the general name of the intestinal microbial community, and the total number of genes encoded by this complex microbial community is 150 times that of the number of genes encoded by human [6, 7]. The function of intestinal flora is not only related to nutrition digestion, but also closely related to immunity and metabolic diseases with the in-depth study [8, 9]. In 1998, Marshall first proposed the theory of "gut-liver axis" [10], that is, after the intestinal barrier is damaged, bacterial translocation endotoxin enters the portal vein system, activates liver Kupffer cells, etc., and then releases a series of inflammatory factors, further aggravating liver damage and disease progression. Many studies [11, 12] have shown that intestinal flora disorder is associated with NAFLD. The composition of intestinal flora varies from SS to NASH, fibrosis, and cirrhosis. Therefore, intestinal flora may be a predictor of NAFLD progression and severity. The involvement of microorganisms in changing the balance between pro-inflammatory or anti-inflammatory signals may lead to the progression of NASH inflammation [13, 14]. The pathogenesis of NAFLD remains unclear, and it is thought to involve a complex interaction between genetic susceptibility variation, environmental factors, insulin resistance, and changes in intestinal flora [15]. There is an urgent need to understand the relationship between NAFLD and intestinal flora disorder, which may serve as predicators of NAFLD progression and severity, and may contribute to improved diagnosis, patient stratification, and identification of new therapeutic targets. To this end, we investigated the relationship between intestinal flora disorder and non-alcoholic fatty liver disease based on bioinformatics. This paper mainly explores the relationship between intestinal flora disorder and SS, NASH, and NASH cirrhosis. The specific workflow is displayed in Fig. 1.

Results

Intestinal flora disorder PPI network

Introduce 80 disease targets related to intestinal flora disorder obtained from the database into String 11.0 database, and set high confidence >0.7. In addition, "the 1st shell" and "the 2nd shell" were set to "no more than 20 interactors" in this study. The protein interaction data was then visualized by Cytoscape 3.7.2 software to obtain the PPI network of intestinal flora disorder (as shown in Figure 2). The network had 86 nodes, which interacted with 930 edges. The size of nodes is positively correlated with the degree value of nodes. Module analysis was conducted through the MCODE plugin. Select the genes for module
1 (41 genes) and module 2 (12 genes) and intersect the first 10 hub genes obtained from cytoHubba plugin to obtain the key disease targets related to intestinal flora disorder. According to the above method, a total of 10 key targets related to intestinal flora disorder were produced (as shown in Supplementary Table 1).

**DEGs screening of NAFLD**

The gene chips obtained from GEO database were analyzed by R 4.0.2 (as shown in Figure 3). In the volcano plots, the red nodes represent upregulated genes and green nodes represent downregulated genes. In the heat maps, red areas represent upregulated genes and blue or green represents downregulated genes. According to the adjusted criteria of \( P \leq 0.05 \) and \(|\log_2 (FC) | \geq 1.5\), 93 DEGs were selected from GSE89632 chip (67 down-regulated genes, 26 up-regulated genes). 50 integrated DEGs were screened from GSE17470, GSE24807, GSE33814, GSE89632 and GSE48452 (24 down-regulated genes, 26 up-regulated genes). And 53 DEGs were screened out in GSE58979 (51 down-regulated genes, 2 up-regulated genes). Information on differentially expressed genes is shown in supplementary table 2.

**NAFLD PPI network**

718 SS disease targets, 379 NASH related targets and 171 NASH cirrhosis relates targets were input into the String 11.0 database, respectively, with a high confidence greater than 0.7. And the obtained protein interaction data were imported into Cytoscape 3.7.2 software to build SS PPI network (Figure 4-a), NASH PPI network (Figure 4-b) and NASH cirrhosis PPI network (Figure 4-c). The module analysis of the MCODE plugin and the hub genes of cytoHubba plugin were used to screen the key disease targets of these three PPI networks. The key targets of the disease are obtained by intersecting the genes of the first two modules and the top 10% of the hub genes.

The PPI network of SS has 591 nodes and 5295 edges. Module analysis was conducted through the MCODE plugin. Select the genes for module 1 (25 genes) and module 2 (45 genes) and intersect the first 72 hub genes obtained from cytoHubba plugin to obtain the key disease targets related to SS. According to the above method, a total of 57 key targets related to SS were produced (as shown in Supplementary Table 1). The PPI network of NASH has 300 nodes and 1746 edges. Module analysis was conducted through the MCODE plugin. Select the genes for module 1 (18 genes) and module 2 (21 genes) and intersect the first 38 hub genes obtained from cytoHubba plugin to obtain the key disease targets related to NASH. According to the above method, a total of 36 key targets related to NASH were produced (as shown in Supplementary Table 1). The PPI network of NASH cirrhosis has 128 nodes and 630 edges. Module analysis was conducted through the MCODE plugin. Select the genes for module 1 (17 genes) and module 2 (9 genes) and intersect the first 18 hub genes obtained from cytoHubba plugin to obtain the key disease targets related to NASH. According to the above method, a total of 17 key targets related to NASH were produced (as shown in Supplementary Table 1).
Merge of intestinal flora disorder PPI network and NAFLD PPI network

The PPI network of intestinal flora disorder and the PPI network of SS, NASH and NASH cirrhosis were merged through the merge function of Cytoscape 3.7.2 software, respectively. And the possible targets of intestinal flora in the treatment of NAFLD were found. We have obtained 20 possible targets for treating SS, 7 for treating NASH, and 7 for treating NASH cirrhosis. These are potential targets for intestinal flora to intervene in different stages of NAFLD. The information of merge genes is shown in Table 1.

Merge network of NAFLD progress

The SS PPI network, NASH PPI network, and NASH cirrhosis PPI network are merged by the merge function of Cytoscape software respectively, in order to find key targets for NAFLD development. The module analysis of the MCODE plugin and the hub genes of cytoHubba plugin were used to screen the key progress targets of NAFLD. The key progress targets of NAFLD are obtained by intersecting the first two modules and the top 10% of the hub genes, in the merge network.

The merge network of SS PPI network and NASH PPI network has 171 nodes and 1044 edges (as shown in Figure 5-a). Module analysis was conducted through the MCODE plugin. Select the genes for module 1 (27 genes) and module 2 (9 genes) and intersect the first 18 hub genes obtained from cytoHubba plugin to obtain the key disease targets related to the progression from SS to NASH. Based on the above approach, a total of 17 key targets related to the progression from SS to NASH were obtained. The merge network of NASH PPI network and NASH cirrhosis PPI network has 50 nodes and 207 edges. (as shown in Figure 5-b) Module analysis was conducted through the MCODE plugin. Select the genes for module 1 (14 genes) and module 2 (5 genes) and intersect the first 11 hub genes obtained from cytoHubba plugin to obtain the key disease targets related to the progression from NASH to NASH cirrhosis. Based on the above approach, a total of 11 key targets related to the progression from NASH to NASH cirrhosis were obtained. The merge network of SS PPI network and NASH cirrhosis PPI network has 108 nodes and 530 edges. (as shown in Figure 5-c) Module analysis was conducted through the MCODE plugin. Select the genes for module 1 (15 genes) and module 2 (6 genes) and intersect the first 5 hub genes obtained from cytoHubba plugin to obtain the key disease targets related to the progression from SS to NASH cirrhosis. Based on the above approach, a total of 5 key targets related to the progression from SS to NASH cirrhosis were obtained. The information of merge genes is shown in Table 1.

The intestinal flora disorder PPI network was merged with the merge network of NAFLD progress through the Merge function of Cytoscape 3.7.2 software, and the potential targets for intestinal flora interfering NAFLD progress were found. We obtained five potential targets (AKT1, F2, ICAM1, PTGS2, CRP) for intestinal flora to intervene the progression of NAFLD from SS to NASH, three potential targets (CRP, ICAM1, F2) for intestinal flora to intervene the progression of NAFLD from NASH to NASH cirrhosis. And seven potential targets (NOS3, IL2RA, F2, CD8A, NOS2, ICAM1, CRP) for intestinal flora to intervene the progression of NAFLD from SS to NASH cirrhosis were obtained. These are considered as potential targets for intestinal flora to intervene in the NAFLD process.
Subsequently, we compared all merge genes with DEGs of NAFLD, and found seven overlapping targets (CCL2, PTGS2, IL6, IL1B, FOS, SPINK1 and C5AR1). In this study, these 7 targets were the core potential targets for intestinal flora to intervene in NAFLD.

**GO function enrichment and KEGG pathway enrichment analysis**

We used R 4.0.2 (https://cran.r-project.org/doc/FAQ/R-FAQ.html#Citing-R) software to perform GO and KEGG enrichment analysis on the protein targets of the merge networks. In the bubble chart, the X-axis represents the number of target genes (Gene Ratio), and the Y-axis represents the KEGG pathway or GO term where the target gene is significantly enriched. The size of the dots intuitively reflects the size of the Gene Ratio, and the color depth of the dots reflects different p-value ranges.

The merge genes enrichment analysis of intestinal flora disorder and NAFLD PPI network is as shown in Figure 6. (1) Analysis of merge of intestinal flora disorder and SS PPI networks: (1) KEGG pathway enrichment analysis (Figure 6-a) found 148 pathways, of which 20 pathways had p-value and q-value less than 0.05. In the GO enrichment analysis (Figure 6-b), a total of 541 terms were found, of which 512 terms were related to biological processes (BP), 11 terms were related to cell composition (CC) and 18 terms were related to molecular function (MF). (2) Analysis of merge of intestinal flora disorder and NASH PPI networks: KEGG pathway enrichment analysis (Figure 6-c) found 121 pathways, of which 22 pathways had p-value and q-value less than 0.05. In the GO enrichment analysis (Figure 6-d), a total of 616 terms were found, of which 590 terms were related to biological processes (BP), 7 terms were related to cell composition (CC) and 19 terms were related to molecular function (MF). (3) Analysis of merge of intestinal flora disorder and NASH cirrhosis PPI networks: KEGG pathway enrichment analysis (Figure 6-e) found 57 pathways, of which 12 pathways had p-value and q-value less than 0.05. In the GO enrichment analysis (Figure 6-f), a total of 347 terms were found, of which 305 terms were related to biological processes (BP), 10 terms were related to cell composition (CC) and 32 terms were related to molecular function (MF).

The merge genes enrichment analysis of merge network of NAFLD progress is as shown in Figure 7. (1) Analysis of merge of intestinal flora disorder, SS, and NASH PPI networks: (1) KEGG pathway enrichment analysis (Figure 7-a) found 115 pathways, of which 15 pathways had p-value and q-value less than 0.05. In the GO enrichment analysis (Figure 7-b), a total of 670 terms were found, of which 636 terms were related to biological processes (BP), 12 terms were related to cell composition (CC) and 22 terms were related to molecular function (MF). (2) Analysis of merge of intestinal flora disorder, NASH and NASH cirrhosis PPI networks: KEGG pathway enrichment analysis (Figure 7-c) found 22 pathways, of which 21 pathways had p-value and q-value less than 0.06. In the GO enrichment analysis (Figure 7-d), a total of 323 terms were found, of which 299 terms were related to biological processes (BP), 5 terms were related to cell composition (CC) and 19 terms were related to molecular function (MF). (3) Analysis of merge of intestinal flora disorder, SS and NASH cirrhosis PPI networks: KEGG pathway enrichment analysis (Figure 7-e) found 57 pathways, of which 12 pathways had p-value and q-value less than 0.05. In the GO enrichment analysis (Figure 7-f), a total of 347 terms were found, of which 305 terms were related to
biological processes (BP), 10 terms were related to cell composition (CC) and 32 terms were related to molecular function (MF).

**Sort-target-term-pathway network**

Figure 8 was obtained by visualizing the merge genes, GO terms and KEGG pathways through Cytoscape 3.7.2 software. The network has 153 nodes (6 nodes for merge sort, 20 nodes for merge genes, 3 nodes for GO sort, 69 nodes for GO terms, 13 nodes for KEGG BRITE, 42 nodes for KEGG pathways) and 480 edges. The targets with degree value greater than 10 were AKT1, ICAM1, NOS3, PTGS2, NOS2, F2, CRP, EDN1, CSF2, CD8A and CDH1. The top 10 GO terms of degree value are mainly BP terms (response to molecule of bacterial origin, response to lipopolysaccharide, reactive oxygen species metabolic process, neurotransmitter metabolic process, nitric oxide biosynthetic process, neurotransmitter biosynthetic process, nitric oxide metabolic process, reactive oxygen species biosynthetic process, positive regulation of reactive oxygen species metabolic process, and reactive nitrogen species metabolic process). The KEGG pathway with higher degree value is mainly related to signal transduction (TNF signaling pathway, HIF-1 signaling pathway, Apelin signaling pathway, JAK-STAT signaling pathway), human disease (AGE-RAGE signaling pathway in diabetic complications, Human T-cell leukemia virus 1 infection, Kaposi sarcoma-associated herpesvirus infection, Fluid shear stress and atherosclerosis) and endocrine system (Estrogen signaling pathway, Relaxin signaling pathway).

Interference of intestinal ora with the pathological process of NAFLD is closely associated with inflammation and insulin resistance. TNF signaling pathway, AGE - RAGE signaling pathway in the diabetic activity and NF-kappa B signaling pathway will promote the up-regulation of CCL2, IL6, IL1B, FOS, SPINK1, C5AR1 and PTGS2 after activation, which will lead to liver inflammation and promote the occurrence and development of NAFLD. Intestinal flora can act on SPINK1, C5AR1, and PTGS2 to improve NAFLD. CCL2, IL6, IL1B, FOS and NF-κB may play an important role in the occurrence and development of NAFLD. KEGG Database (https://www.kegg.jp/kegg/kegg1.html) and software of Pathway Builder Tool 2.0 were used to generate the figure. As seen in Figure 9, the major predictive signaling pathways for intestinal flora interfering with NAFLD were constructed.

**Discussion**

NAFLD is a manifestation of obesity and metabolic syndrome affecting the liver, and its pathogenesis is associated with numerous factors [2]. The exact pathogenesis of NAFLD has not been fully elucidated. The current theories that are mainly concerned are the second hit theory and the gut-liver axis theory. Intestinal flora has a strong effect on the liver, and can participate in the development and progression of NAFLD through the “gut-liver axis” [16–18]. Most scholars believe that factors related to the pathogenesis of NAFLD include immune activation [19], inflammatory response [20], bile acid metabolism interference [21], insulin resistance [22], and fasting-induced adipose factor [23].

CCL2, also known as monocyte chemoattractant protein 1, is mainly secreted by monocytes and macrophages [24]. Mouse experiments [25, 26] showed that when the chemokine CCL2 was
overexpressed, macrophages in adipose tissue were recruited to secrete a large number of inflammatory factors, promoting liver steatosis, inflammation and fibrosis. CCL2 deficiency can resist lipid accumulation and insulin resistance induced by high-fat diet. Many studies [27, 28] have suggested that CCL2 expression is upregulated in the liver of animal models of NAFLD, and its circulating level in NAFLD patients is also increased. NAFLD can gradually develop from SS and NASH to NASH cirrhosis and even liver cancer. CCL2 can affect the development and progression of NAFLD by participating in processes such as steatosis and inflammation. IL-6, which is produced by T cells, B cells, macrophages, and endothelial cells, is a cytokine with a variety of physiological effects and an important inflammatory factor, induces the production of TNF-α, and plays an autocrine role in the production of other pro-inflammatory cytokines [29]. Clinically relevant studies have reported significantly elevated serum IL-6 and TNF-α levels in NAFLD patients, suggesting that high levels of TNF-α and IL-6 may be involved in the occurrence and development of NAFLD [30]. A large number of foreign studies have also shown that there are significant immune disorders in the surrounding tissues and liver of NAFLD patients, accompanied by excessive production of inflammatory factors such as TNF-α and IL-6 [31, 32]. In-depth studies have shown that IL-6 can leads to liver steatosis, insulin resistance and aggravation of inflammation [30, 33]. Peripheral adipocytes and hepatocytes can secrete IL-6, which in turn mediates and promotes macrophage infiltration, thereby participating in the development and progression of NAFLD [30]. The protein encoded by IL1B is a member of the interleukin 1 cytokine family. This cytokine is an important mediator of the inflammatory response, and is involved in a variety of cellular activities, including cell proliferation, differentiation, and apoptosis [34, 35]. Liver Kupffer cells can produce IL-1B, which exacerbates liver inflammation and steatosis [36]. IL-1B promotes the production of nitric oxide and cell apoptosis in islet cells, leading to selective destruction of islet cells, thus further inducing insulin resistance [36]. In addition to playing a role in inflammation, IL1B may contribute to NAFLD pathogenesis by promoting insulin resistance and altering lipid metabolism [37, 38]. FOS is considered as a regulator of cell proliferation, differentiation and transformation, which is closely related to inflammatory response and tumor. FOS can encode leucine zipper proteins that can dimerize with proteins of the JUN family, thereby forming the transcription factor complex AP-1 [39]. Christoph's work revealed significant changes in genes associated with metabolic processes, transport, signal transduction and redox in the mouse NASH model, and AP-1 appears to be the key transcriptional regulator of these changes [40]. FOS expression is related to the occurrence of insulin resistance [41]. Significantly increased DNA binding of NF-κB and AP-1 in the NASH patients play a major role in oxidative stress and insulin resistance of NASH pathophysiological mechanisms [42]. SPINK1, as a secretory trypsin inhibitor, is a potential marker for the diagnosis of liver cancer [43]. SPINK1 expressed in both the pancreas and the gastrointestinal tract, and the trypsin inhibitor encoded by this gene can be secreted from pancreatic acinar cells into pancreatic juice, which is often present at high levels in the pancreas and pancreatic juice [44–46]. In the normal gastrointestinal tract, SPINK1 is thought to play a protective role in both gastric mucosa [47] and colonic mucosa [48]. Evidence shows that SPINK1 is an important growth factor linking chronic inflammation and cancer [49–51], and the progression of NAFLD to liver cancer and cell cancer may be driven by spink-1 gene [46, 52, 53]. Complement receptors CR1 and CR3 are responsible for the phagocytic and adhesive properties of neutrophils, whereas the C5a receptor mediates the pro-inflammatory and chemotactic
actions of the C5aR1 [54]. C5aR1 has been shown to promote primary tumor growth and immunosuppression [55, 56]. C5aR1-mediated tumor promotion and immunosuppression have been attributed to the C5aR1 ligand C5a [57, 58]. Studies have shown that reducing C5 levels can reduce inflammatory responses, thus significantly reducing the degree of liver fibrosis and cirrhosis [59, 60]. Sendler [61] used C5a receptor antagonist to inhibit C5aR1 and found that it could greatly reduce the fibrogenesis after pancreatic necrosis. PTGS2 is the principal isozyme responsible for production of inflammatory prostaglandins and plays an important role in inflammation and proliferation of a variety of cells and tissues. PTGS2 and its products affect the metabolism of fat, carbohydrate and protein, thus interfering with the metabolism of normal substances in liver and affecting the occurrence and development of liver disease [62]. The increased expression of PTGS2 causes triglycerides to accumulate in liver cells [63], which may lead to insulin resistance. Insulin resistance can lead to increased free fatty acids and transported to liver cells for accumulation, increased release of inflammatory factors, and up-regulated expression of PTGS2 [64]. Inhibition of PTGS2 expression may be related to the prevention of liver histological changes caused by intestine endotoxin [65, 66].

Insulin resistance and IL-6 promote and cause each other, which is an important cause of NAFLD [67, 68]. C5AR1 [59, 60], PTGS2 [64] and pro-inflammatory factors such as IL-1B and IL-6 [69, 70] are involved in the sensitivity of liver to insulin through inflammatory signaling pathways and induce the formation of insulin resistance. CCL2 [27, 28], FOS [41], SPINK1 [44–46] and PTGS2 [62, 63] in adipose tissue may increase insulin resistance. Insulin resistance is closely related to the formation of NAFLD, and changes in intestinal flora composition can improve insulin resistance [71]. Therefore, we speculated that L6, IL1B, CCL2 and FOS may be key genes for the development and progression of NAFLD, which may contribute to the relationship between NAFLD and intestinal flora disorder mainly through inflammation and IR. SPINK1, C5AR1 and PTGS2 may be key genes for intestinal flora to interfere with NAFLD, possibly through endotoxin, insulin resistance and inflammation intervention.

In order to find out the mechanism of intestinal flora interfering with NAFLD, this study conducted GO and KEGG enrichment analysis of the merge genes. The results of GO functional enrichment analysis showed that the gene functions of intestinal flora interfering with NAFLD are mainly reflected in BP, including regulating metabolic processes, epithelial development and affecting immunity. For example, positive regulation of reactive oxygen species metabolic process, regulation of reactive oxygen species metabolic process, reactive oxygen species metabolic process, acute inflammatory response and external side of plasma membrane. KEGG enrichment analysis found that the pathway intervention of intestinal flora in NAFLD was mainly closely related to signal transduction, immune regulation and physiological metabolism. For example, Arginine biosynthesis is related to metabolism, Kaposi sarcoma-associated herpesvirus infection is related to inflammatory phenotype, and Neuroactive ligand-receptor interaction is related to signal molecules and their interactions. TNF signaling pathway, AGE-RAGE signaling pathway in diabetic complications and NF-kappa B signaling pathway have been shown several times with good enrichment in all KEGG enrichment results. We believe that these three pathways are the key pathways for intestinal flora to treat NAFLD. AGE/RAGE activation can increase oxidative stress and trigger a series of inflammatory reactions, angiogenesis, fibrosis, thrombosis, cell proliferation and apoptosis [72].
Japanese researchers reported that in the case of diabetes, serum AGE levels is positively correlated with insulin resistance and negatively correlated with adiponectin levels, which is an important biomarker to distinguish NASH from SS [73]. Fehrenbach et al. found that AGE/RAGE activation could increase the synthesis of reactive oxygen species, activate NF-kappa B signaling pathway in HSCs, and make them differentiate into fibroblasts [74]. Activation of the NF-kappa B signaling pathway is an important inflammatory response mechanism, such as inflammatory bowel disease and inflammation during liver injury [75–77], and is also associated with fatty liver [78]. LPS binds to toll-like receptors on the macrophage surface and activates IKK, which phosphorylates or degrades I-κB. Free NF-κB enters the nucleus and promotes the transcription of iNOS and inflammatory cytokines such as TNF-α, IL-1β, and IL-6 [79, 80]. Literature studies have shown that NF-κB and TNF signaling pathways are important inflammatory pathways during cholestatic liver injury [81]. Alexander et al. found that the infiltration of neutrophils and inflammatory macrophages in Nlrp3 mutant mice depends on TNF signaling, which can improve LPS-driven liver injury, prevent the activation of hepatic stellate cells in Nlrp3 mutant mice, and trigger liver inflammation [82]. The pathogenesis of many chronic liver diseases, such as viral hepatitis, alcoholic liver disease, and explosive liver failure, is closely related to the dysfunction of TNF signaling pathway [83]. Activation of the TNF signaling pathway may trigger NASH and liver fibrosis [84]. Therefore, we speculated that TNF signaling pathway, AGE-RAGE signaling pathway in diabetic complications and NF-kappa B signaling pathway may play an important role in the intervention of intestinal flora in the occurrence and development of NAFLD.

In summary, we believe that the pathogenesis of NAFLD is mainly related to insulin resistance and inflammation. The main mechanism by which intestinal flora interferes with the pathogenesis of NAFLD is to inhibit the expression of inflammatory genes in the related pathways and reduce insulin resistance, thereby reducing NAFLD. Although our research basically discussed the mechanism of intestinal flora interferes with NAFLD, there are still some limitations. First, the research data comes from existing databases, so the authenticity and completeness of the results depend on the data. Second, the results do not reflect all the genuine cellular network characteristics in the organism, so further experiments will be needed to confirm the data in the future.

**Conclusion**

In this study, bioinformatics was used to explore the relationship between intestinal flora disorder and NAFLD, and the mechanism of intestinal flora interfering with NAFLD was discussed. Through Merge function intestinal flora disorders PPI network and NAFLD PPI network, we have acquired 20 potential targets for treating SS by intestinal flora, 7 for treating NASH, and 7 for treating NASH cirrhosis. The intestinal flora disorder PPI network was merged with the merge network of NAFLD progress through the Merge function of Cytoscape 3.7.2 software. We obtained 5 potential targets for intestinal flora to intervene the progression of NAFLD from SS to NASH, 3 potential targets for intestinal flora to intervene the progression of NAFLD from NASH to NASH cirrhosis, and 7 potential targets for intestinal flora to intervene the progression of NAFLD from SS to NASH cirrhosis. Finally, the 7 targets (CCL2, IL6, IL1B, FOS, PTGS2, SPINK1 and C5AR1) were selected as the core potential targets for intestinal flora to
intervene NAFLD. CCL2, IL6, IL1B and FOS are mainly related to the occurrence and development mechanism of NAFLD, while PTGS2, SPINK1 and C5AR1 are mainly related to the intervention of intestinal flora in the occurrence and development of NAFLD. GO enrichment analysis showed that the gene functions of intestinal flora interfering with NAFLD are mainly reflected in basic biological processes, including regulating metabolic processes, epithelial development and affecting immunity, such as positive regulation of reactive oxygen species metabolic process, acute inflammatory response and external side of plasma membrane. KEGG enrichment analysis showed that the pathway through which intestinal flora interfered with NAFLD was mainly closely related to signal transduction, immune regulation and physiological metabolism. TNF signaling pathway, AGE-RAGE signaling pathway in diabetic complications and NF-kappa B signaling pathway are the main pathways.

In conclusion, we predicted that the intervention process of intestinal flora in NAFLD was mainly related to inflammatory response and AGE/RAEG signal transduction. Interfering with these mechanisms of intestinal flora may lead to the goal of curing NAFLD. However, as our study is based on data analysis, further experiments are needed to confirm this result. The preliminary results of this study confirmed the intervention role and related mechanisms of intestinal flora in the occurrence and development of NAFLD, laying a good foundation for further exploration of its mechanism of action.

**Materials And Methods**

**Collect disease targets associated with intestinal flora disorder**

Intestinal flora disorder, intestinal dysbiosis, intestinal flora imbalance, intestinal microbiota disorder and dysbiosis of intestinal flora are taken as retrieval words. Proteins related to intestinal flora disorder are collected through the disease databases. Including Therapeutic Targets Database (TTD) (https://db.idrblab.org/ttd/) [85], DigSee (http://210.107.182.61/geneSearch/) [86] and DisGeNET (http://www.disgenet.org/search) [87]. Species were selected as "Homo sapiens", and gene-disease score was set ≥ 0.1 in DisGeNET database. 80 proteins related to intestinal flora disorder were finally obtained.

**Collect disease targets associated with NAFLD**

Screening of differentially expressed genes (DEGs) in NAFLD: search in GEO database (https://www.ncbi.nlm) [88]. And use the following keywords to screen gene expression profile : (1) non-alcoholic fatty liver disease, (2) the tissue source is "human liver tissue", (3) the research type is "Expression profiling by Array", (4) Nonalcoholic fatty liver in the experimental group and healthy liver in the control group. Finally, we have acquired one SS gene chip (GSE89632), five NASH gene chips (GSE17470, GSE24807, GSE33814, GSE89632, GSE48452), and one NASH cirrhosis gene chip (GSE58979). Among the GSE58979 microarray samples, it is considered to be a normal control group for steatosis of less than 5%. And screen NASH cirrhosis DEGs from visceral sample. R 4.0.2 software was used to screen DEGs. During the screening process, the screening condition was log2 (FC) > 1.5, P≤0.05. Finally, we obtained 93 DEGs of SS, 51 DEGs of NASH, and 53 DEGs of NASH cirrhosis.
Collection of disease proteins associated with NAFLD: NAFL, simple fatty liver, steatosis, NASH, steatohepatitis, NASH cirrhosis and NASH-cirrhosis were searched in the disease databases for retrieval words. We collect disease targets related to SS, NASH, and NASH cirrhosis in the Therapeutic Target Database (TTD) (https://db.idrblab.org/ttd/), DigSee (http://210.107.182.61/geneSearch/) and DisGeNET (http://www.disgenet.org/search). Species were selected as "Homo sapiens", and gene-disease score was set ≥ 0.1 in DisGeNET database. After removing repetitive disease targets, 643 SS related proteins, 333 NASH related proteins, and 120 NASH cirrhosis related proteins can be acquired in the end. Combining all NAFLD genes, we finally gained 718 SS related targets, 379 NASH related targets, and 171 NASH cirrhosis relates targets.

**Construction of protein-protein interaction (PPI) network and screening of key genes**

String 11.0 database (https://string-db.org/) is a database that stores known and predicted protein interactions, including direct and indirect interactions of proteins. It scores the information of each protein interaction. The higher the score, the higher the confidence of protein interaction [89]. Protein names related to intestinal flora disorder and NAFLD collected in the disease database were input into the String 11.0 database for retrieval. And protein interaction data with a high confidence interval score greater than 0.7 were selected to ensure the reliability of the data. The obtained protein interaction data were imported into Cytoscape 3.7.2 (http://www.cytoscape.org/) software to construct PPI network related to intestinal flora disorder and NAFLD. Cytoscape is an open source bioinformatics analysis software used to construct molecular interaction networks composed of protein, gene and drug interactions for visual browsing and analysis [90].

The key targets of disease are screened using MCODE plug-ins and cytoHubba plug-ins included in the Cytoscape 3.7.2 software. The MCODE plug-in was used for module analysis to select the genes of the first two modules with high scores. Screen Hub genes using cytoHubba plug-in. The genes screened by the above two methods were intersected to obtain the key targets for the treatment of diseases.

By using the Merge function of Cytoscape 3.7.2 software, the PPI network of intestinal flora disorder is merged with SS, NASH, and NASH cirrhosis PPI network respectively. To select the intersecting parts of the network, proteins that exist in the intersection may be a potential core target for intestinal flora to interfering with NAFLD. These potential core targets were systematically analyzed to explore the relationship between intestinal flora and NAFLD. The PPI network of NAFLD was separately merged to find the connection between different stages of NAFLD diseases at the intersection, so as to help explore the relationship between intestinal flora and NAFLD.

**Gene ontology (GO) and kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment**

To evaluate the role of potential core targets by bioinformatic annotation, the R 4.0.2 (https://cran.r-project.org/doc/FAQ/R-FAQ.html#Citing-R) software with the Bioconductor package was manipulated, including GO knowledgebase (http://geneontology.org/), KEGG pathway enrichment analysis (https://www.genome.jp/kegg/) [91-93].
Abbreviations

NAFLD
Non-alcoholic Fatty Liver Disease
SS
Simple steatosis
NASH
Non-alcoholic steatohepatitis
NASH cirrhosis
Non-alcoholic steatohepatitis associated cirrhosis
PTGS2
prostaglandin-endoperoxide synthase 2
SPINK1
serine peptidase inhibitor Kazal type 1
C5AR1
complement C5a receptor 1
CCL2
C-C motif chemokine ligand 2
IL6
interleukin 6
IL1B
interleukin 1 beta
FOS
Fos proto-oncogene, AP-1 transcription factor subunit

Declarations

Availability of data and materials statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation. All data obtained or analyzed during this study are available from the published article and its supplementary information files. The datasets during the current study are available from the corresponding author upon reasonable request.

Funding

This work was supported by the Young Scientists Training Program of Beijing University of Chinese Medicine, State Key Laboratory of Generic Manufacture Technology of Chinese Traditional Medicine (Grant nos. 2010DQ740377) and the National Nature Science Foundation of China (Grant nos. 81673829).

Acknowledgments
This work was supported by the Young Scientists Training Program of Beijing University of Chinese Medicine, State Key Laboratory of Generic Manufacture Technology of Chinese Traditional Medicine (Grant nos. 2010DQ740377) and the National Nature Science Foundation of China (Grant nos. 81673829).

Authors' information

Affiliations

Department of Clinical Chinese Pharmacy, School of Chinese Materia Medica, Beijing University of Chinese Medicine, Beijing, China: Yingying Liu, Wei Zhou, Jingyuan Zhang, Xinkui Liu, Siyu Guo, Shanshan Jia, Jiarui Wu, Haojia Wang, Jialin Li & Yingying Tan.

Authors' contributions

YL and JW conceived and designed the study. WZ provided useful suggestions on the design. XL and JZ provided guidance on the methodology. SG, SJ, HW and JL collected the information and preprocessed the data of research. YL drafted the initial manuscript. YT reviewed and revised the manuscript. All authors read and approved the final version of the manuscript.

Corresponding author

Correspondence to Jiarui Wu.

Ethics declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References


43. Hass HG, Jobst J, Vogel U, Scheurlen M, Nehls O. Overexpression of tumor-associated trypsin inhibitor (SPINK1/TATI) in hepatitis C-associated hepatocellular carcinoma: potential implications...


Additional Material

Supplementary table 1 Information of disease key targets obtained from the PPI network

Supplementary table 2 Information of differentially expressed genes obtained from the GEO database

Supplementary table 3 List of GO enrichment results to genes in all enrichment analysis
Supplementary table 4 List of pathway enrichment results to genes in all enrichment analysis

Tables

Table 1 Information of merge networks genes
<table>
<thead>
<tr>
<th>Merge target</th>
<th>Official Full Name</th>
<th>Type of merge</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCL2</td>
<td>C-C motif chemokine ligand 2</td>
<td>SS-NASH, NASH-NASH cirrhosis, SS-NASH cirrhosis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNF</td>
<td>tumor necrosis factor</td>
<td>SS-NASH, NASH-NASH cirrhosis, SS-NASH cirrhosis</td>
</tr>
<tr>
<td>TLR4</td>
<td>toll like receptor 4</td>
<td>SS-NASH, NASH-NASH cirrhosis</td>
</tr>
<tr>
<td>CXCL8</td>
<td>C-X-C motif chemokine ligand 8</td>
<td>SS-NASH, NASH-NASH cirrhosis, SS-NASH cirrhosis</td>
</tr>
<tr>
<td>IL4</td>
<td>interleukin 4</td>
<td>SS-NASH, NASH-NASH cirrhosis</td>
</tr>
<tr>
<td>IL1A</td>
<td>interleukin 1 alpha</td>
<td>SS-NASH</td>
</tr>
<tr>
<td>ICAM1</td>
<td>intercellular adhesion molecule 1</td>
<td>SS-NASH, NASH-NASH cirrhosis, intestinal flora disorder -SS, intestinal flora disorder -NASH, intestinal flora disorder -NASH cirrhosis</td>
</tr>
<tr>
<td>IFNG</td>
<td>interferon gamma</td>
<td>SS-NASH, NASH-NASH cirrhosis</td>
</tr>
<tr>
<td>PTGS2</td>
<td>prostaglandin-endoperoxide synthase 2</td>
<td>SS-NASH, intestinal flora disorder -SS, intestinal flora disorder -NASH</td>
</tr>
<tr>
<td>IL13</td>
<td>interleukin 13</td>
<td>SS-NASH</td>
</tr>
<tr>
<td>IL6</td>
<td>interleukin 6</td>
<td>SS-NASH, NASH-NASH cirrhosis</td>
</tr>
<tr>
<td>Protein</td>
<td>Description</td>
<td>SS-NASH</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>IL1B</td>
<td>interleukin 1 beta</td>
<td></td>
</tr>
<tr>
<td>IL17A</td>
<td>interleukin 17A</td>
<td></td>
</tr>
<tr>
<td>IL10</td>
<td>interleukin 10</td>
<td></td>
</tr>
<tr>
<td>FOS</td>
<td>Fos proto-oncogene, AP-1 transcription factor subunit</td>
<td></td>
</tr>
<tr>
<td>JUN</td>
<td>Jun proto-oncogene, AP-1 transcription factor subunit</td>
<td></td>
</tr>
<tr>
<td>ADIPOQ</td>
<td>adiponectin, C1Q and collagen domain containing</td>
<td></td>
</tr>
<tr>
<td>CRP</td>
<td>C-reactive protein</td>
<td></td>
</tr>
<tr>
<td>EDN1</td>
<td>endothelin 1</td>
<td></td>
</tr>
<tr>
<td>AKT1</td>
<td>AKT serine/threonine kinase 1</td>
<td></td>
</tr>
<tr>
<td>NOS3</td>
<td>nitric oxide synthase 3</td>
<td></td>
</tr>
<tr>
<td>CSF2RB</td>
<td>colony stimulating factor 2 receptor subunit beta</td>
<td></td>
</tr>
<tr>
<td>IL2RA</td>
<td>interleukin 2 receptor subunit alpha</td>
<td></td>
</tr>
<tr>
<td>CALCA</td>
<td>calcitonin related polypeptide alpha</td>
<td></td>
</tr>
<tr>
<td>F2</td>
<td>coagulation factor II, thrombin</td>
<td></td>
</tr>
<tr>
<td>CD8A</td>
<td>CD8a molecule</td>
<td></td>
</tr>
<tr>
<td>KRT20</td>
<td>keratin 20</td>
<td></td>
</tr>
<tr>
<td>Gene</td>
<td>Description</td>
<td>Disorder</td>
</tr>
<tr>
<td>--------</td>
<td>--------------------------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>CDH1</td>
<td>cadherin 1</td>
<td>intestinal flora disorder -SS</td>
</tr>
<tr>
<td>PRSS1</td>
<td>serine protease 1</td>
<td>intestinal flora disorder -SS</td>
</tr>
<tr>
<td>SPINK1</td>
<td>serine peptidase inhibitor Kazal type 1</td>
<td>intestinal flora disorder -SS</td>
</tr>
<tr>
<td>NOS2</td>
<td>nitric oxide synthase 2</td>
<td>intestinal flora disorder -SS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>intestinal flora disorder -NASH cirrhosis</td>
</tr>
<tr>
<td>IRF6</td>
<td>interferon regulatory factor 6</td>
<td>intestinal flora disorder -SS</td>
</tr>
<tr>
<td>SP1</td>
<td>Sp1 transcription factor</td>
<td>intestinal flora disorder -SS</td>
</tr>
<tr>
<td>C5AR1</td>
<td>complement C5a receptor 1</td>
<td>intestinal flora disorder -SS</td>
</tr>
<tr>
<td>BCHE</td>
<td>butyrylcholinesterase</td>
<td>intestinal flora disorder -SS</td>
</tr>
<tr>
<td>CSF2</td>
<td>colony stimulating factor 2</td>
<td>intestinal flora disorder -NASH</td>
</tr>
<tr>
<td>GCG</td>
<td>glucagon</td>
<td>intestinal flora disorder -NASH</td>
</tr>
</tbody>
</table>

**Figures**
Figure 1

Workflow of bioinformatics
Figure 1

Workflow of bioinformatics
Figure 2

Intestinal flora disorder PPI network. Yellow nodes represent module 1 genes, pink nodes represent module 2 genes, blue nodes represent hub genes, and red nodes represent intersection genes of modules and hub.
Figure 2

Intestinal flora disorder PPI network. Yellow nodes represent module 1 genes, pink nodes represent module 2 genes, blue nodes represent hub genes, and red nodes represent intersection genes of modules and hub.
Figure 3

Volcano plot and Heat map of NAFLD differentially expressed genes. (a) Volcano plot of SS; (b) Volcano plot of NASH cirrhosis; (c) Heat map of SS; (d) Heat map of NASH; (e) Heat map of NASH cirrhosis.
Figure 3

Volcano plot and Heat map of NAFLD differentially expressed genes. (a) Volcano plot of SS; (b) Volcano plot of NASH cirrhosis; (c) Heat map of SS; (d) Heat map of NASH; (e) Heat map of NASH cirrhosis.
Figure 4

NAFLD PPI network. (a) SS PPI network; (b) NASH PPI network; (c) NASH cirrhosis PPI network. Yellow nodes represent the genes of module 1, pink nodes represent the genes of module 2, blue nodes represent hub genes, and red nodes represent the intersection of module genes and hub genes. The size of nodes is positively correlated with the degree value of nodes.
Figure 4

NAFLD PPI network. (a) SS PPI network; (b) NASH PPI network; (c) NASH cirrhosis PPI network. Yellow nodes represent the genes of module 1, pink nodes represent the genes of module 2, blue nodes represent hub genes, and red nodes represent the intersection of module genes and hub genes. The size of nodes is positively correlated with the degree value of nodes.
Figure 5

Merge network of NAFLD progress. (a) Merge network of SS and NASH PPI networks; (b) Merge network of NASH and NASH cirrhosis PPI networks; (c) Merge network of SS and NASH cirrhosis PPI networks. Blue nodes represent module 1 genes, purple nodes represent module 2 genes, yellow nodes represent hub genes, and red nodes represent the intersection of module genes and hub genes. The size of nodes is positively correlated with the degree value of nodes.
Figure 6

Merge genes enrichment analysis of intestinal flora disorder and NAFLD PPI networks. (a) KEGG pathway analysis for merge of intestinal flora disorder and SS PPI networks; (b) GO enrichment analysis for merge of intestinal flora disorder and SS PPI networks; (c) KEGG pathway analysis for merge of intestinal flora disorder and NASH PPI networks; (d) GO enrichment analysis for merge of intestinal flora disorder and NASH PPI networks; (e) KEGG pathway analysis for merge of intestinal flora disorder and NASH cirrhosis PPI networks; (f) GO enrichment analysis for merge of intestinal flora disorder and NASH cirrhosis PPI networks.
Figure 6

Merge genes enrichment analysis of intestinal flora disorder and NAFLD PPI networks. (a) KEGG pathway analysis for merge of intestinal flora disorder and SS PPI networks; (b) GO enrichment analysis for merge of intestinal flora disorder and SS PPI networks; (c) KEGG pathway analysis for merge of intestinal flora disorder and NASH PPI networks; (d) GO enrichment analysis for merge of intestinal flora disorder and NASH PPI networks; (e) KEGG pathway analysis for merge of intestinal flora disorder and NASH cirrhosis PPI networks; (f) GO enrichment analysis for merge of intestinal flora disorder and NASH cirrhosis PPI networks.
Figure 7

Merge genes enrichment analysis of intestinal flora disorder and NAFLD progress PPI networks. (a) KEGG pathway analysis for merge of intestinal flora disorder, SS and NASH PPI networks; (b) GO enrichment analysis for merge of intestinal flora disorder, SS and NASH PPI networks; (c) KEGG enrichment analysis for merge of intestinal flora disorder, NASH and NASH cirrhosis PPI networks; (d) KEGG pathway analysis for merge of intestinal flora disorder, NASH and NASH cirrhosis PPI networks; (e) KEGG pathway analysis for merge of intestinal flora disorder, SS and NASH cirrhosis PPI networks; (f) GO enrichment analysis for merge of intestinal flora disorder, SS and NASH cirrhosis PPI networks.
Figure 7

Merge genes enrichment analysis of intestinal flora disorder and NAFLD progress PPI networks. (a) KEGG pathway analysis for merge of intestinal flora disorder, SS and NASH PPI networks; (b) GO enrichment analysis for merge of intestinal flora disorder, SS and NASH PPI networks; (c) KEGG pathway analysis for merge of intestinal flora disorder, NASH and NASH cirrhosis PPI networks; (d) GO enrichment analysis for merge of intestinal flora disorder, NASH and NASH cirrhosis PPI networks; (e) KEGG pathway analysis for merge of intestinal flora disorder, SS and NASH cirrhosis PPI networks; (f) GO enrichment analysis for merge of intestinal flora disorder, SS and NASH cirrhosis PPI networks.
Figure 8

Sort-target-term-pathway network. Red round nodes represent type of merge. Yellow inverted triangle nodes represent merge genes. Green triangle nodes represent type of GO terms. Blue diamond nodes represent GO terms. Pink square nodes represent the BRITE of KEGG. Purple hexagon nodes represent KEGG pathways. The size of nodes is positively correlated with the degree value of nodes.
Figure 8

Sort-target-term-pathway network. Red round nodes represent type of merge. Yellow inverted triangle nodes represent merge genes. Green triangle nodes represent type of GO terms. Blue diamond nodes represent GO terms. Pink square nodes represent the BRITE of KEGG. Purple hexagon nodes represent KEGG pathways. The size of nodes is positively correlated with the degree value of nodes.
Figure 9

Illustration of crucial putative biological progress for intestinal flora intervening NAFLD.
Illustration of crucial putative biological progress for intestinal flora intervening NAFLD.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile.docx
- Additionalfile.docx