Integrated analysis of potential gene crosstalk between non-alcoholic fatty liver disease and diabetic nephropathy

Qianqian Yan\textsuperscript{1,2†}, Zihao Zhao\textsuperscript{1,2†}, Dongwei Liu\textsuperscript{1,2,3,4}, Jia Li\textsuperscript{1,2,3,4}, Shaokang Pan\textsuperscript{1,2,3,4}, Jiayu Duan\textsuperscript{1,2,3,4*}, Jiancheng Dong\textsuperscript{2*} and Zhangsuo Liu\textsuperscript{1,2,3,4*}

\textsuperscript{1}Department of Integrated Traditional and Western Nephrology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, \textsuperscript{2}Research Institute of Nephrology, Zhengzhou University, Zhengzhou, China, \textsuperscript{3}Henan Province Research Center for Kidney Disease, Zhengzhou, China, \textsuperscript{4}Key Laboratory of Precision Diagnosis and Treatment for Chronic Kidney Disease in Henan Province, Zhengzhou, China

Background: Growing evidence indicates that non-alcoholic fatty liver disease (NAFLD) is related to the occurrence and development of diabetic nephropathy (DN). This bioinformatics study aimed to explore optimal crosstalk genes and related pathways between NAFLD and DN.

Methods: Gene expression profiles were downloaded from Gene Expression Omnibus. CIBERSORT algorithm was employed to analyze the similarity of infiltrating immunocytes between the two diseases. Protein–protein interaction (PPI) co-expression network and functional enrichment analysis were conducted based on the identification of common differentially expressed genes (DEGs). Least absolute shrinkage and selection operator (LASSO) regression and Boruta algorithm were implemented to initially screen crosstalk genes. Machine learning models, including support vector machine, random forest model, and generalized linear model, were utilized to further identify the optimal crosstalk genes between DN and NAFLD. An integrated network containing crosstalk genes, transcription factors, and associated pathways was developed.

Results: Four gene expression datasets, including GSE66676 and GSE48452 for NAFLD and GSE30122 and GSE1009 for DN, were involved in this study. There were 80 common DEGs between the two diseases in total. The PPI network built with the 80 common genes included 77 nodes and 83 edges. Ten optimal crosstalk genes were selected by LASSO regression and Boruta algorithm, including CD36, W1P1, CBX7, FCN1, SLC35D2, CP, ZDHHC3, PTPN3, LPL, and SPP1. Among these genes, LPL and SPP1 were the most significant according to NAFLD-transcription factor network. Five hundred twenty-nine nodes and 1,113 edges comprised the PPI network of activated pathway-gene. In addition, 14 common pathways of these two diseases were recognized using...
Gene Ontology (GO) analysis; among them, regulation of the lipid metabolic process is closely related to both two diseases.

Conclusions: This study offers hints that NAFLD and DN have a common pathogenesis, and LPL and SPP1 are the most relevant crosstalk genes. Based on the common pathways and optimal crosstalk genes, our proposal carried out further research to disclose the etiology and pathology between the two diseases.

KEYWORDS
non-alcoholic fatty liver disease, diabetic nephropathy, crosstalk, LPL, SPP1, bioinformatics

Introduction

Diabetic nephropathy (DN) is a rigorous microvascular complication primarily associated with both type 1 and type 2 diabetes mellitus (T2DM) and has been the leading cause of end-stage renal disease (ESRD) worldwide (1–3). Both morbidity and mortality of DN have promptly increased around the world (1, 2, 4). Non-alcoholic fatty liver disease (NAFLD) has become pyramidally ordinary in parallel with the adding popularity of obesity and other components of the metabolic syndrome (5, 6). NAFLD is distinguished as the existence of fat storage ≥5% of liver weight with the absence of excessive alcohol consumption or secondary cause of liver diseases such as autoimmune hepatitis, hemochromatosis, and Wilson’s disease (3, 7–9). Being metabolic diseases, factors that contribute to NAFLD, such as diabetes, chronic inflammation, insulin resistance, and obesity, are also associated with the development of DN.

Several observational studies reported an impressive proportion that there were 70%–86% of patients with NAFLD also suffering from T2DM (6, 8, 10–12). Jia et al. (13) found that the cumulative incidence of DN in patients with NAFLD was much higher than those without it and that the liver fat content was positively correlated with increased occurrence of albuminuria and decreased glomerular filtration rate (GFR). Targher et al. (14) also found that the prevalence of diabetic retinopathy and chronic kidney disease (CKD) was significantly higher in patients with NAFLD. Previous epidemiological studies further suggested several contributors including metabolic syndrome, dysbiosis, unhealthy diets, platelet activation, and processes acting as the linking factors between NAFLD and CKD, which implied the potential correlations involved in the pathogenesis of liver and kidney disease (15). That NAFLD might be a risk factor for DN had been analyzed by some researchers (13). The relationship between NAFLD and DN seems rational and of clinical interest to some extent.

Based on the results of current observational studies, the potential contributions of genetic factors and protein–protein interactions (PPIs) on the correlation of NAFLD and DN should be further analyzed. In this study, bioinformatics analysis was used to disclose the crosstalk mechanisms between NAFLD and DN at the transcriptomic level. The mutual transcription characteristics would offer new insights into the common pathogenesis of NAFLD and DN. The purpose of this study is to recognize optimal crosstalk genes, participant pathways, and transcription factors (TFs). We hypothesize the existence of crosstalk genes between NAFLD and DN, then employed comprehensive bioinformatics and enrichment analyses to identify the common differentially expressed genes (DEGs) and the functional pathways of NAFLD and DN. At last, we identified 10 crosstalk genes, favoring the similarity between these two diseases.

Materials and methods

Study design and data collection

We acquired microarray data from Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/). After earnest review, four gene expression profiles (GSE66676 and GSE48452 were NAFLD datasets, and GSE30122 and GSE1009 were diabetic human kidney disease datasets, with no other complications) were selected. Figure 1 shows the schematic of the research.

Data procession and differentially expressed gene analysis

We combined two datasets for each disease to increase the sample size. R software (version 4.1.1; https://www.r-project.org/) and “Bioconductor” packages were applied to analyze the data. The expression data from different datasets were normalized using the
robust multi-array average and merged together, and the “sva” library was used for combating batch correction to remove batch effects. We then used the Linear Models for microarray data (“limma” package) to identify DEGs by comparing the expression values between NAFLD patients and control cases. Genes with $P < 0.05$ were considered DEGs. The same way was done in the diabetic kidney disease dataset. The “pheatmap” package was used to draw the heatmap of the DEGs in R software.

FIGURE 1
Workflow of this study. We downloaded the gene expression profiles of NAFLD and DN from the GEO database, including two NAFLD datasets (GSE66676 and GSE48452) and two DN datasets (GSE30122 and GSE1009). Datasets were merged, and DEGs were found by R software. The LASSO regression and Boruta algorithm were used to select the optional crosstalk genes. RF model was considered the best model to predict DN using the 10 crosstalk genes. In addition, protein-protein interaction (PPI) and functional enrichment analyses of the DEGs were performed. Graphic created with BioRender.com. NAFLD, nonalcoholic fatty liver disease; DN, diabetic nephropathy; GEO, Gene Expression Omnibus, DEGs, different expression genes.
Gene set enrichment analysis

In order to understand and interpret coordinate pathway-level changes in transcriptomics experiments represented under different conditions, the gene set enrichment analysis (GSEA) was used to determine whether there are statistically significant differences between the two groups as to a defined set of genes (16). R package "clusterProfiler" was used to perform GSEA. Gene sets “c7.all.v7.5.1.entrez” were downloaded from "Downloads (gsea-msigdb.org)” website, and then R software was employed to retrieve systematic functional annotation information. $P < 0.05$ was the cutoff criterion. We examined the pathway-level changes for all DEGs in NAFLD and DN to find out whether there were reductive pathways.

Immune infiltration analysis

The proportions of 22 kinds of immune cells, including naive B cells, memory B cells, plasma cells, CD8 T cells, naive CD4 T cells, CD4 resting memory T cells, activated memory CD4 T cells, helper follicular T cells, regulatory T cells (Tregs), delta gamma T cells, resting natural killer (NK) cells, activated natural killer (NK) cells, monocytes, macrophages M0, macrophages M1, macrophages M2, resting dendritic cells (DCs), active DCs, resting mast cells, activated mast cells, eosinophils, and neutrophils, were obtained, and CIBERSORT algorithm was utilized to analyze the gene expression data between NAFLD and DN.

We brought in all genes that were expressed in both NALFD and DN patients to explore the common ground based on 22 kinds of immune cells between the two diseases. The percentage of each kind of immune cell in the samples was calculated. Single-sample gene set enrichment analysis (ssGSEA) was employed to calculate the degree of penetration of 28 immune cell types on the grounds of the expression levels of genes in 28 published gene sets for immune cells (17).

Identification of potential crosstalk genes and functional enrichment analysis

The potential crosstalk genes were identified as DN-related DEGs overlapped with the NAFLD-related ones. These crosstalk genes could have the potential ability of linking the pathogeneses of NAFLD and DN.

To further determine the biological features of potential crosstalk genes, Gene Ontology (GO) analysis was accomplished by “clusterProfiler” of R Bioconductor packages. A classification method is offered by the “clusterProfiler” packages to classify genes based on their projection at a specific level of the GO corpus and provide functions to calculate enrichment values for GO terms. The enriched function with $P < 0.05$ was considered a significant pathway. Based on this analysis method, we selected the top 20 GO biological processes. To identify the most significant clusters of the crosstalk genes, PPI network of crosstalk genes was constituted by STRING (STRING 11.5; Search Tool for the Retrieval of Interaction Gene; https://string-db.org/). Cytoscape (version 3.8.0) was used to visualize the PPI network.

Identification of optimal diagnostic crosstalk genes

To better screen the risk crosstalk genes between NAFLD and DN, Boruta algorithm and least absolute shrinkage and selection operator (LASSO) regression were performed in R project. The LASSO regression was used to filter the best predictive features while fitting a generalized linear model (GLM) and avoiding overfitting. The Boruta employed a wrapper approach, built around a random forest (RF) classifier. After merging two NAFLD datasets, the expression values of potential crosstalk genes were extracted. The DEGs between NAFLD patients and healthy controls were reserved for feature selection, and the optimal crosstalk genes were initially recognized using the Boruta algorithm and LASSO regression. To narrow it down further, we combined the results of the two algorithms.

On the grounds of the optimal diagnostic crosstalk gene expression on the NAFLD merged dataset, we created the RF model, support vector machine (SVM) model, and GLM to pick out the best model. The response variable was the diagnosis of NAFLD or not, and the DEGs were used as explanatory variables. We then used the explain feature of "DALEX" package in R to find out which was the finest model among these three models aforementioned based on the plot of residual distribution.

Development of the random forest model using optimal diagnostic crosstalk genes

After extracting the gene expression values of the filtered crosstalk genes that constitute the merged gene expression profile, the RF model with the gene expression value and sample type was built (NAFLD and healthy) to further confirm the diagnostic value of these crosstalk genes. The R package "randomForest" was applied to build the RF model. The "ComBat" method of “sva” packages in R project was performed to eliminate the batch effect. It is worth noting that the gene sample expression values were changed after a series of operations that are mentioned above, comparing previous gene expressions. Therefore, the primitive expression profile of the two datasets GSE66676 and GSE48452 was obtained. Afterward,
the optimal crosstalk genes were confirmed by Boruta algorithm and LASSO regression, and then the expression values of the optimal crosstalk genes from the merged data were confirmed. We select the gene expression values of the filtered optimal genes to form the merged gene expression profile, and the RF models with the gene expression profile value and sample type were set up (NAFLD or healthy). The NAFLD merged data were input as training data, and the DN merged data were imported as testing data. The prediction effectiveness was determined by the accuracy rate of the test set.

Transcription factor-adjusted and pathway analysis of the crosstalk genes

We downloaded TFs that regulate the target genes from TRRUST and ChEA3 databases, taking the intersection of TFs from these two databases. Based on the TF–target relationship, the NAFLD-related TF–target pairs were picked out, and the Cytoscape software was used to set up and visualize the TF–target gene interaction network.

In order to pick out activated pathways, the remarkably enriched pathways by the DEGs of NAFLD were screened. We selected the potential crosstalk pathways that may be the bridge of NAFLD and DN and obtained the genes functioning in each pathway. Finally, the Cytoscape software was used to construct the pathway–gene crosstalk network. For the purpose of confirming the functional TFs, which adjusted the crosstalk genes in the activated pathways, we picked out the crosstalk genes in the pathway–gene pairs and identified the NAFLD-related TFs and DN-related TFs. Moreover, 10 crosstalk genes were also included. Finally, the network of these four parts was created.

Results

Identification of differentially expressed genes and functional pathways by gene set enrichment analysis

A total of 215 study subjects were included in the current study. The mean age with standard deviation was 63.29 ± 14.61 years (DN patients) and 52.44 ± 12.90 years (healthy control) for the DN group and 45.92 ± 11.29 years (NAFLD patients) and 33.52 ± 8.82 years (healthy control) for the NAFLD group. The proportion of women was 66.5% and 78.9% for the DN and NAFLD groups, respectively. To identify DEGs between NAFLD and healthy controls, we recruited microarray expression profiles of GSE66676 and GSE48452 from the GEO database. After merging and normalizing the microarray data, 1,265 DEGs between NAFLD and healthy controls were selected by “limma” package ($P < 0.05$). Two DN datasets were also picked out from the GEO website, which was done in the same way. Finally, we got 1,265 DEGs in the NAFLD merged dataset and 1,085 DEGs in the DN merged dataset (heatmap shown in Supplementary Figures S1A, B). GSEA was implemented to reveal the functional similarity between the two diseases. All DEGs of each disease were contained in the GSEA using gene set “c7.all.v7.5.1.entrez.” As a result, three common pathways were identified in NAFLD and DN (Supplementary Figures S2A, B).

Immune infiltration analysis

By employing the CIBERSORT algorithm, we investigated the similarity in immune infiltration between NAFLD patients and DN patients in 22 subpopulations of immune cells. The results acquired from NAFLD patients and DN patients were summarized by R software (Supplementary Figure S3A). The samples were screened according to $P < 0.05$, and the percentage of each kind of immune cell in the samples was calculated. As shown in Supplementary Figure S3B, there are no significant differences between NAFLD and DN tissue in most immune cells, such as macrophage M1, which were considered to be proinflammatory and promote inflammation (18). However, the DN tissue generally included a high ratio of naive CD4 T cells, delta gamma T cells, activated NK cells, and resting mast cells, while resting NK cells had the opposite trend of expression. In the ssGSEA (Supplementary Figure S3C), 17 immune cell subtypes, including activated B cell, NK T cell, immature B cell, effector memory CD8 T cell, and central memory CD4 T cell, demonstrated no significant expression differences between NAFLD and DN. However, Myeloid-derived suppressor cells (MDSC), memory B cells, regulatory T cells, T follicular helper cells, and Type 1 T helper cells showed higher expression in DN patients, while immature DCS were highly expressed in NAFLD patients. The consequences of the CIBERSORT algorithm and ssGSEA manifest that the two diseases are likely to have a similar immune infiltration environment, which laid the theoretical foundation to link them.

Identifying crosstalk genes, Gene Ontology analysis, and construction of the protein–protein interaction network

After overlapping the DEGs of the two datasets, we finally got 80 crosstalk genes. The Venn diagram for the DEGs was given in Figure 2A. The heatmaps of common DEGs between NAFLD and DN were represented in Figures 2B, C. The GO analysis found that common DEGs were most intensively related to neutrophil-related pathways such as neutrophil degranulation, neutrophil activation involved in immune response, neutrophil-mediated immunity, and neutrophil activation. The detailed biological pathways in which DEGs were involved are shown in Figures 3A–D.
Seventy-seven nodes and 283 edges comprised the constructed PPI network of common DEGs (Figure 4A). The most significant module (score = 5.846) was recognized by MCODE, a plug-in of Cytoscape. CytoHubba was used to identify the hub genes (Figure 4B). SPP1 may be the key gene that contacts NAFLD and CKD, since it had the highest score in the biological network.

Prediction of optimal crosstalk genes and building the machine learning model

We then extracted the expression data of the 80 genes from the NAFLD gene expression profile. Gene biomarkers were identified with the LASSO and Boruta algorithms. A total of 15 genes were finally selected (Figures 5A–C; Supplementary Table S2). Furthermore, the optimal crosstalk genes were identified by overlapping biomarkers derived from these two algorithms. We got 10 optimal crosstalk genes in the end.

To select and create the optimal prediction model, three models including SVM, RF, and GLM were created in light of the training NAFLD merged dataset. After that, the “DALEX” package’s explanatory feature in R was utilized to analyze the three aforementioned models. As shown in Figures 6A–C, which revealed the residual distribution, the RF model was confirmed as the best suitable model because it possesses the least sample residual. Ultimately, the expression of the 10 optimal crosstalk genes in NAFLD [CD36, WIPI1, CBX7, FCN1, SLC35D2, CP, ZDHHC3, PTPN3, lipoprotein lipase (LPL), and SPP1] was input to create the RF model. The gene expression profile of the 10 feature genes was also extracted from the DN merged dataset. Treating the NAFLD merged dataset as training data and the DN merged dataset as validation data, the predicted outcome of the RF model was shown in Supplementary Table S1. Supplementary Figure S4 shows the importance of 10 genes in the RF model. The forecast performance of each gene in both NAFLD and DN was shown in Supplementary Figure S5. The area under the curve (AUC) values of LPL and SPP1 in DN were 86% and 80.1%, and the AUC values of LPL and SPP1 in NAFLD were 72.5% and 64.3%, respectively. Supplementary Figure S6 showed the expression of the 10 genes in the two diseases.

Transcription factor–gene regulation network

We got a total of 35 mutual TFs, and the TF–target network was established as shown in Figure 7A. The optimal crosstalk genes with the highest degree were SPP1 and LPL and therefore potentially played a significant role in the TF–target network. Supplementary Figure S7 showed the network between NAFLD TF–target pairs and DN–genes.
Ultimately, 14 crucial pathways, which may play a key role in the progress of NAFLD, were selected. In order to recognize the pathway crosstalk between NAFLD and DN, we established the pathway–gene crosstalk network. Five hundred twenty-nine nodes and 1,113 edges were included in the activated pathway–gene network (Figure 7B). To further explore the relationship between NAFLD and DN, the DN-related TF–target pairs and NAFLD-related TF–target pairs were extracted and the PPI was built. Meanwhile, the PPI between 10 crosstalk genes was obtained, then the activated TF–crosstalk gene network was established (Figure 7C). Consequently, we discovered that crosstalk genes were regulated by many TFs. The closeness of their relationship was indicated by the size of the circle. The highest degree among the 10 crosstalk genes remained to be SPP1 and LPL. DN-related TF–target pairs had a closer relationship with crosstalk genes.
The major outcome of this study was that bioinformatics analysis could expose crosstalk genes between NAFLD and DN. Accordingly, LPL and SPP1 were identified to be the most concerned genes; meanwhile, some participant pathways were identified. According to their expression values in each patient and forecast ability, the latent correlation of these genes was confirmed. The areas under the ROC curve of these two genes are higher than those of most genes (Supplementary Figure S3).

**Discussion**

The major outcome of this study was that bioinformatics analysis could expose crosstalk genes between NAFLD and DN. Accordingly, LPL and SPP1 were identified to be the most concerned genes; meanwhile, some participant pathways were identified. According to their expression values in each patient and forecast ability, the latent correlation of these genes was confirmed. The areas under the ROC curve of these two genes are higher than those of most genes (Supplementary Figure S3).
FIGURE 6
Construction and evaluation of RF model, SVM model, and GLM. (A) Accumulated residual distribution picture of the sample. (B) Boxplot of the residuals of the sample. The root mean square of the residuals was indicated by a red dot. (C) The significance of the variables in the three models.

FIGURE 7
A series of protein–protein interaction (PPI) networks. (A) NAFLD-related TFs selected by TRRUST and ChEA3, indicated by yellow circles. The purple circles represent the top 50 significant differentially expressed genes in the NAFLD merged dataset. The optimal 10 crosstalk genes are indicated by the green circle. (B) Fourteen crucial pathways functioning in both NAFLD and DN, indicated by green squares. The purple-red circles represent the genes from each pathway both NAFLD-related and DN-related. The lilac circles represent the significant differentially expressed genes in the NAFLD merged dataset. The green circle represents the significant differentially expressed genes in the DN merged dataset. (C) DN-related TFs selected by TRRUST and ChEA3, indicated by purple circles. The green circles represent 10 optimal crosstalk genes. The genes from each pathway both NAFLD-related and DN-related were indicated by yellow circles. The pink circles represent NAFLD-related TFs.
furthermore, they have the highest degree in the TF-target network.

According to our results, the aberrant lipid and glucose metabolism plays an important role in the crosstalk between NAFLD and DN. The LPL gene, which belongs to the lipase gene family including hepatic lipase, endothelial lipase, and pancreatic lipase, could combine with lipoproteins and cell surface proteins concurrently, resulting in accumulation and uptake of lipoproteins (19, 20). In 2019, Teratani et al. (21) found that the expression of LPL was changed in hepatic stellate cells in NAFLD patients. Serum obesity-related factors, including interleukin-6, leptin, and free fatty acid (FA), could further affect its circulating level (21). NAFLD would even evolve to hepatic stellate cells and cause excessive collagen deposition, osteopontin secretion in hepatocytes could directly activate fibrosis via a Notch-activated pathway. Notch-mediated osteopontin secretion in hepatocytes could directly activate hepatic stellate cells and cause excessive collagen deposition, despite hepatoacellular injury. Furthermore, by performing chronic γ-secretase inhibitor treatment, liver Notch activity was decreased and hepatic stellate cell activation and liver fibrosis were reduced (28). In the progression of DN, Notch signaling pathway was activated following long-term exposure to hyperglycemia (29). The expression of constitutively active Notch intracellular domain in mature podocytes caused podocyte dedifferentiation, glomerulosclerosis, and apoptosis that substantially caused albuminuria and progressive renal failure (30, 31). By treating with γ-secretase inhibitors, the diabetes-induced glomerulosclerosis and podocyte injury could be prevented, which suggested that inhibiting the overactive Notch pathway in renal cells could be a potential plausible therapeutic approach (32).

Previous studies indicated that hyperglycemia would increase the expression of SPP1, which caused an elevated exposure of cells to proinflammatory cytokines and inflammation indicators including tumor necrosis factor α, transforming growth factor β, and interleukin-1 (33–36). In our current study, the immune filtration analysis also found that there were common expressions of macrophages and DCs between the NAFLD and DN groups, which suggested that inflammatory activations were involved in the crosstalk of the two diseases. It has been proven that in high-glucose conditions, the transcriptional activity of SPP1 was enhanced in proximal tubular epithelial cells (PTECs), which means that when T2DM occurred, the expression of SPP1 will increase and SPP1 promotes the occurrence and development of both diseases (37). Zhang et al. (38) also found that the SPP1 was negatively correlated with GFR in diabetic kidney disease patients. Some researchers believed that SPP1 could be the core target to treat diabetic kidney disease by using traditional Chinese medicine (39). SPP1 is also conjectured to function in the transformation of non-alcoholic steatohepatitis (NASH) to HCC like LPL (40).

In the enrichment analysis part, we found that 14 pathways were involved in the crosstalk between NAFLD and DN, one of which was “regulation of lipid metabolic process.” Several previous studies also demonstrated that the progression of DN was linked to serum lipid abnormalities and renal ectopic lipid accumulation (18, 41, 42). The proportion of kidney-absorbed LDL would be different when the activity of the LDL receptor changed; meanwhile, the expression of LDL receptor would be remarkably suppressed by cholesterol in podocytes (42). Lipid loading facilitates the phenotypic conversion of podocytes, which results in the disappearance of its epithelial features (43–45). Most DN patients performed albuminuria or macroproteinuria during the progression of disease. The albumin also acted as a vehicle for FAs in urine. Consequently, albuminuria may cause extensive accumulation of FAs and accelerate kidney injury in DN patients (46). By analyzing 34 DN patients and 12 healthy controls, Herman-Edelstein et al. (47) found a high degree of correlation between lipid metabolism and GFR. In the situation of continuing hyperglycemia in diabetic patients, TGs and FAs were accumulated (47). Ectopic lipid accumulation in non-adipose tissues, such as liver, kidney, heart, and pancreas, occurs because of raised serum TGs, FFAs, and modified cholesterol (41, 48–51), which appear to play a part in the pathogenesis of DN (52–54). This condition also seems to result in NAFLD. This suggests that diabetes also acts as a link between these two diseases. Consistent with this, obvious neutral lipid accumulation was found both in glomeruli and
tubulointerstitium in diabetic kidneys (47). Two of the features of DN in electron microscope podocyte process effacement, interestingly, lipotoxicity and lipid accumulation, can lead to podocyte malfunction and apoptosis (55). Hence, it is reasonable for DN to see heavy lipid deposition (47).

A previous study found that the total counts of lipid droplets (LDs) decreased when kidney tissue was seriously fibrosed. This process was similar to the progression of NASH (56). Liver is the central organ of lipoprotein metabolism, since it takes part in the production of lipoprotein particles in all categories. It also plays a central role in the metabolism of TGs and cholesterol. Serum TGs and remnant cholesterol would be elevated when liver function is impaired. Then, it comes with altered glucose metabolism and insulin resistance, which are believed to be hallmarks of NAFLD (57). NAFLD occurs at the time of excessive intake of FAs and TGs from the circulation. Unbalanced lipid metabolism is also related to NAFLD advancement from steatosis to NASH; moreover, alterations in liver and serum lipidomic signatures are excellent indicators of NAFLD’s development and progression (58).

Both NAFLD and DN were considered to be affected by chronic inflammation progression, especially in individuals with abnormal serum glucose and lipid concentration (7, 12, 59). In our current study, we also found that there were aberrant proportions and expression levels of immune indicators in both diseases. There were a total of 7 and 17 kinds of immune cells performing higher proportion and expression levels, respectively. Most of them showed no significant differences between NAFLD and DN groups, which indicated the mutual mechanism in the two diseases. Enrichment analysis further demonstrated that several common DEGs were enriched in immune-related functions, including neutrophil activation, neutrophil-mediated immunity, and positive regulation of mitophagy.

Liver-mediated lipid changes are associated with the severity of proteinuria (60). Similarly, the diagnosis of DN refers to the appearance of specific pathologic structural alongside functional changes in the kidney of patients with DM, one of which is proteinuria (61). Renal injury is generally non-reversible on the condition that albuminuria persistently occurs (13). Various mechanisms, such as poor plasma glucose control, activation of sympathetic nervous system, and insulin resistance both in liver and kidney contribute coordinately to the advancement of kidney diseases (62, 63). Metabolic syndrome is a significant contributor to the evolution of chronic kidney dysfunction (64, 65). It has already been confirmed that lipid accumulation is strongly associated with inflammatory stress in the kidney. Through perturbing the LDL receptor pathway and induced phenotypic change and dysfunction in podocytes, inflammation induces lipid accumulation (42). Obesity is closely associated with these two diseases, which have been considered as a risk factor for both NAFLD and DN (9, 66, 67). Obesity, T2DM, and NAFLD can not only facilitate systemic insulin resistance but also boost the accumulation of hepatic fat and impairment of glucose metabolism (3, 66); in the meantime, insulin resistance can stimulate hepatic macrophages. T2DM often causes chronic hyperinsulinemia, which plays a vital role in liver metabolism abnormalities (41). The possibility that kidney dysfunction occurs was remarkably higher in patients with NAFLD based on two cross-sectional studies (14, 68). There is an assumption that along with the advancement of NAFLD into NASH, diabetic kidney diseases would occur (69). With the progression of NASH, metabolic disorders, for example, dyslipidemia, insulin resistance, and glucose intolerance, would collectively advance the renin-angiotensin (RAS) system system and influence nitric oxide formation (70–72), which can facilitate the progression of DN. Simultaneously, a mechanism in NASH adjusted by liver-derived inflammatory mediators and oxidative stress, which boosts the free proinflammatory, procoagulant, pro-oxidant, and profibrogenic factors from the liver, participates in the development of DN (69, 73–75). Renal hemodynamics may be influenced because of activation of the sympathetic nervous system, hence conducing to the onset or deterioration of kidney diseases (3, 63). Therefore, based on these common risk factors, it is reasonable to conjecture the interlink between NAFLD and DN, and early detection and treatment of NAFLD may be of clinical significance for the intervention of DN.

Altogether, this study identified two major crosstalk genes and relevant shared pathways, which emphasize the comparability and underlying relationship between NAFLD and DN. It is reasonable to consider that NAFLD patients are vulnerable to get and develop kidney dysfunction caused by T2DM and insulin insistance. Despite the credibility of such a probability interlink, more research needs to be done to reveal the potential mechanism of these two diseases. Nevertheless, there exist several limitations in our study. Firstly, data on survival times and outcomes are lacking. Therefore, the effect of these crosstalk genes on survival could not be tested. Secondly, the integration of different gene expression datasets might be biased because of discrepancies in the experimental setting for each dataset. Ideally, unnecessary bias should be avoided by making sure all gene expression data have the same experimental settings. In addition, it was a retrospective study. For the purpose of averting analysis bias associated with retrospective studies, a prospective study is recommended to be conducted. Lastly, the current study is entirely ground on computer analyses, thus the validation analysis based on wet-lab will be encouraged to confirm these crosstalk genes we found.

**Conclusion**

To the best of our knowledge, this is the first study of crosstalk mechanisms between NAFLD and DN using bioinformatics analysis, identifying common immune and TF-related mechanisms. LPL and SPP1 are the most relevant crosstalk genes in our study, which suggest that NAFLD and DN may have a common pathogenesis.
Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Author contributions

QY, ZZ, JYD, and DL designed the study, collected the data, analyzed the data, and wrote the manuscript. ZL, JCD, SP, and JL reviewed and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Young Scientists Fund of the National Natural Science Foundation of China (Grant No. 82103916), General Program of the National Science Foundation of China General Project (No. 81970633).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2022.1032814/full#supplementary-material

1. Kato M, Natarajan R. Diabetic nephropathy–emerging epigenetic mechanisms. Nat Rev Nephrol (2014) 10(9):517–30. doi: 10.1038/nrneph.2014.116

2. Xiong Y, Zhou L. The signaling of cellular senescence in diabetic nephropathy. Oxid Med Cell Longev (2019) 2019:7495629. doi: 10.1155/2019/7495629

3. Wang T-Y, Fang R-F, Bu Z-Y, Targher G, Byrne CD, Sun D-Q, et al. Association of metabolic dysfunction-associated fatty liver disease with kidney disease. Nat Rev Nephrol (2022) 18(6):259–68. doi: 10.1038/s41581-021-00519-y

4. Bell S, Fletcher EH, Brady I, Looker HC, Levin D, Joss N, et al. End-stage renal disease and survival in people with diabetes: a national database linkage study. QJM (2015) 108(2):127–34. doi: 10.1093/qjmed/hcu170

5. Neuschwander-Tetri BA. Non-alcoholic fatty liver disease. BMC Med (2017) 15(1):45. doi: 10.1186/s12916-017-0806-8

6. Arslan MS, Turhan S, Dincer I, Mirzak D, Corapcigolu D, Ildiman A. A potential link between endothelial function, cardiovascular risk, and metabolic syndrome in patients with non-alcoholic fatty liver disease. Diabetol Metab Syndr (2014) 14(6):109. doi: 10.1186/1758-5996-6-109

7. Wijarnpreecha K, Thongprayoon C, Boonpheng B, Panjawatanan P, Sharma K, Ungprasert P, et al. Nonalcoholic fatty liver disease and albuminuria: a systematic review and meta-analysis. Eur J Gastroenterol Hepatol (2018) 30(9):986–94. doi: 10.1097/MEG.0000000000001169

8. Heidari Z, Gherebaghi A. Prevalence of non alcoholic fatty liver disease and its association with diabetic nephropathy in patients with type 2 diabetes mellitus. J Clin Diag Res (2017) 11(5):OC04–OC07. doi: 10.7860/JCDR/2017/25931.9823

9. Musso G, Gambino R, Cassader M, Pagano G. Meta-analysis: natural history of non-alcoholic fatty liver disease (NAFLD) and diagnostic accuracy of noninvasive tests for liver disease severity. Ann Med (2011) 43(8):617–49. doi: 10.3109/07853890.2010.518623

10. Targher G, Chonchol M, Bertolini L, Rodella S, Zennari L, Lippi G, et al. Increased risk of CKD among type 2 diabetics with nonalcoholic fatty liver disease. J Am Soc Nephrol (2008) 19(8):1564–70. doi: 10.1681/ASN.2007101155

11. Tolman KG, Fonseca V, Dalpiaz A, Tan MH. Spectrum of liver disease in type 2 diabetes and management of patients with diabetes and liver disease. Diabetes Care (2007) 30(3):734–43. doi: 10.2337/dc06-1539
Angiotensinogen and AT1 antisense inhibition of osteopontin translation in rat diabetic nephropathy 

Yan et al. 10.3389/fendo.2022.1032814

Frontiers in Endocrinology frontiersin.org13

Yan et al. 10.3389/fendo.2022.1032814

Frontiers in Endocrinology frontiersin.org13

Yan et al. 10.3389/fendo.2022.1032814

Frontiers in Endocrinology frontiersin.org13

Yan et al. 10.3389/fendo.2022.1032814
58. Mato JM, Alonso C, Noureddin M, Lu SC. Biomarkers and subtypes of deranged lipid metabolism in non-alcoholic fatty liver disease. *World J Gastroenterol* (2019) 25(24):3009–20. doi: 10.3748/wjg.v25.i24.3009

59. Lv WS, Sun RX, Gao YY, Wen JP, Pan RF, Li L, et al. Nonalcoholic fatty liver disease and microvascular complications in type 2 diabetes. *World J Gastroenterol* (2013) 19(20):3134–42. doi: 10.3748/wjg.v19.i20.3134

60. Vaziri ND. Disorders of lipid metabolism in nephrotic syndrome: mechanisms and consequences. *Kidney Int* (2016) 90(1):41–52. doi: 10.1016/j.kint.2016.02.026

61. Umanath K, Lewis JB. Update on diabetic nephropathy: Core curriculum 2018. *Am J Kidney Dis* (2018) 71(6):884–95. doi: 10.1053/j.ajkd.2017.10.026

62. Targher G, Byrne CD. Non-alcoholic fatty liver disease: an emerging driving force in chronic kidney disease. *Nat Rev Nephrol* (2016) 13(3):297–310. doi: 10.1038/nrneph.2017.16

63. Spoto B, Pisano A, Zoccali C. Insulin resistance in chronic kidney disease: a systematic review. *Am J Physiol Renal Physiol* (2016) 311(6):F1087–F108. doi: 10.1152/ajprenal.00340.2016

64. Chen J, Muntner P, Hamm LL, Fonseca V, Batuman V, Whelton PK, et al. Insulin resistance and risk of chronic kidney disease in nondiabetic US adults. *J Am Soc Nephrol* (2003) 14(2):469–77. doi: 10.1097/01.ASN.0000046029.53933.09

65. Chen J. The metabolic syndrome and chronic kidney disease in U.S. adults. *Ann Intern Med* (2004) 140(3):167–74. doi: 10.7326/0003-4819-140-3-20040203-00007

66. Mantovani A, Petracca G, Beatrice G, Tilg H, Byrne CD, Targher G. Non-alcoholic fatty liver disease and risk of incident diabetes mellitus: an updated meta-analysis of 501,022 adult individuals. *Gut* (2021) 70(5):962–9. doi: 10.1136/gutjnl-2020-322572

67. Wan H, Wang Y, Xiang Q, Fang S, Chen Y, Chen C, et al. Associations between abdominal obesity indices and diabetic complications: Chinese visceral adiposity index and neck circumference. *Cardiovasc Diabetol* (2020) 19. doi: 10.1186/s12933-020-01095-4

68. Targher G, Bertolini L, Chonchol M, Rodella S, Zoppini G, Lippi G, et al. Non-alcoholic fatty liver disease is independently associated with an increased prevalence of chronic kidney disease and retinopathy in type 1 diabetic patients. *Diabetologia* (2010) 53(7):1341–4. doi: 10.1007/s00125-010-1720-1

69. Saito H, Tanabe H, Kudo A, Machii N, Higa M, Yamaguchi S, et al. High FIB4 index is an independent risk factor of diabetic kidney disease in type 2 diabetes. *Sci Rep* (2021) 11(1):11753. doi: 10.1038/s41598-021-88285-6

70. Cohen DE, Fisher EA. Lipoprotein metabolism, dyslipidemia, and nonalcoholic fatty liver disease. *Semin Liver Dis* (2013) 33(4):380–8. doi: 10.1055/s-0033-1358519

71. Goib GB, Pagadala MR, Dasarathy J, Unalp-Arild A, Sargent R, Hawkins C, et al. Renin-angiotensin system and fibrosis in non-alcoholic fatty liver disease. *Liver Int* (2015) 35(1):97–85. doi: 10.1111/liv.12611

72. Targher G. Risk of cardiovascular disease in patients with nonalcoholic fatty liver disease. *Massachusetts Med Soc* (2010) 363(14):1341–50. doi: 10.1056/NEJMra0912063

73. Anders HJ, Huber TR, Isermann B, Schiffer M. CKD in diabetes: diabetic kidney disease versus nondiabetic kidney disease. *Nat Rev Nephrol* (2018) 14(6):361–77. doi: 10.1038/s41581-018-0001-y

74. Musso G, Cassader M, Cohnney S, De Michieli F, Pinach S, Saba F, et al. Fatty liver and chronic kidney disease: Novel mechanistic insights and therapeutic opportunities. *Diabetes Care* (2016) 39(10):1830–45. doi: 10.2337/dc15-1182

75. Fukazawa K, Lee HT. Updates on hepatorenal syndrome. *J Am Soc Nephrol* (2013) 4(9):352. doi: 10.4172/2155-6148.1000352