76109_Auto_Edited.docx

WORD COUNT	4852	TIME SUBMITTED	28-MAY-2022 03:34PM
		PAPER ID	86600606

Name of Journal: *World Journal of Hepatology*

Manuscript NO: 76109

Manuscript Type: ORIGINAL ARTICLE

Clinical and Translational Research

Transcriptome Changes in Stages of Non-Alcoholic Fatty Liver Disease

Transcriptome Changes in Stages of Non-Alcoholic Fatty Liver Disease

Jihad Aljabban, Michael Rohr, Saad Syed, Kamal Khorfan, Vincent Borkowski, Hisham Aljabban, Michael Segal, Mohamed Mukhtar, Mohammed Mohammed, Maryam Panahiazar, Dexter Hadley, Ryan Spengler, Erin Spengler

Abstract

BACKGROUND

NAFLD is the most common chronic liver disease in the U.S. and globally. The currently understood model of pathogenesis consists of a 'multiple hit' hypothesis in which environmental and genetic factors contribute to hepatic inflammation and injury.

AIM

We examined the genetic expression of NAFLD and NASH tissue samples to identify common pathways that contribute to NAFLD and NASH pathogenesis.

METHODS

We employed the Search Tag Analyze Resource for GEO platform to search the NCBI Gene Expression Omnibus to elucidate NAFLD and NASH pathology. For NAFLD, we conducted meta-analysis of data from 58 NAFLD liver biopsies and 60 healthy liver biopsies; for NASH, we analyzed 187 NASH liver biopsies and 154 healthy liver biopsies.

RESULTS

Our results from the NAFLD analysis reinforce the role of altered metabolism, inflammation, and cell survival in pathogenesis and support recently described contributors to disease activity, such as altered androgen and lncrna activity. The top upstream regulator was found to be SREBF1, a transcription factor involved in lipid homeostasis. Downstream of SREBF1, we observed upregulation in CXCL10, HMGCR, HMGCS1, FABP5, PEG10, and downregulation of SHBG and IGF1. These molecular changes reflect low-grade inflammation secondary to accumulation of fatty acids in the liver. Our results from the NASH analysis emphasized the role of cholesterol in pathogenesis. Top canonical pathways, disease networks, and disease functions were related to cholesterol synthesis, lipid metabolism, adipogenesis, and metabolic disease. Top upstream regulators included pro-inflammatory cytokines TNF and IL1B, PDGF

BB, and beta-estradiol. Inhibition of beta-estradiol was shown to be related to derangement of several cellular downstream processes including metabolism, extracellular matrix deposition, and tumor suppression. Lastly, we found riciribine (an AKT inhibitor) and ZSTK-474 (a PI3K inhibitor) as potential drugs that targeted the differential gene expression in our dataset.

CONCLUSION

In this study we describe several molecular processes that may correlate with NAFLD disease and progression. We also identified ricirbine and ZSTK-474 as potential therapy.

Key Words: non-alcoholic fatty liver disease; non-alcoholic steatohepatitis; bioinformatics

Aljabban J, Rohr M, Syed S, Khorfan K, Borkowski V, Aljabban H, Segal M, Mukhtar M, Mohammed M, Panahiazar M, Hadley D, Spengler R, Spengler E. Transcriptome Changes in Stages of Non-Alcoholic Fatty Liver Disease. *World J Hepatol* 2022; In press

Core Tip: Our results from the NAFLD analysis reinforce the role of altered metabolism, inflammation, and cell survival in pathogenesis and support recently described contributors to disease activity, such as altered androgen and lncrna activity. The top upstream regulator was found to be SREBF1, a transcription factor involved in lipid homeostasis. Downstream of SREBF1, we observed upregulation in CXCL10, HMGCR, HMGCS1, FABP5, PEG10, and downregulation of SHBG and IGF1. These molecular changes reflect low-grade inflammation secondary to accumulation of fatty acids in the liver. Our results from the NASH analysis emphasized the role of cholesterol in pathogenesis. Top upstream regulators included pro-inflammatory cytokines TNF and IL1B, PDGF BB, and beta-estradiol. Inhibition of beta-estradiol was shown to be related to derangement of several cellular downstream processes including

metabolism, extracellular matrix deposition, and tumor suppression. Lastly, we found riciribine (an AKT inhibitor) and ZSTK-474 (a PI3K inhibitor) as potential drugs that targeted the differential gene expression in our dataset.

INTRODUCTION

Non-alcoholic fatty liver disease (NAFLD) is a chronic liver disease that is characterized by the accumulation of triglycerides within hepatocytes. This process strongly resembles alcohol-induced fatty liver damage but occurs in the absence of excessive alcohol consumption. Akin to obesity, rates of NAFLD are burgeoning and represent a growing health burden; it is estimated that the global disease prevalence is between 20-30%.[1] There is growing evidence that NAFLD is a multisystem disease with both intra- and extra-hepatic manifestations, with strong association between NAFLD and type 2 diabetes mellitus and metabolic syndrome.[2]

NAFLD comprises of a spectrum of disease that includes simple steatosis, non-alcoholic steatohepatitis (NASH), cirrhosis, and hepatocellular carcinoma (HCC). While hepatic steatosis is seen as a generally benign state, NASH is considered a progressive disease state with increased risk of intra- and extra-hepatic disease complications, including cirrhosis.[3] The gold-standard to diagnose NASH is an invasive liver biopsy. As there are no effective non-invasive diagnostic techniques, which makes estimating the true prevalence of NASH difficult; however, it has been estimated that up 25% of patients with NAFLD have concurrent NASH.[4] As rates of NAFLD continue to increase, it is estimated that NAFLD-related cirrhosis will soon surpass chronic hepatitis as the leading indication for liver transplantation.[5]

The increasing prevalence and health burden of NAFLD has made it imperative to understand the pathogenesis of this disease process. The most current, best understood model of NAFLD conceptualizes a 'multiple hit' hypothesis in which interactions between genetics and environmental factors promote inflammation, cellular injury, and liver damage.[6] These 'hits' include lipid accumulation secondary to diet and lifestyle, obesity, and insulin resistance, all of which predispose the liver to inflammation and

fibrosis. However, the mechanisms by which these hits promote disease progression are still poorly understood. In this meta-analysis, we aim to use bioinformatics of publicly available data to elucidate the most common genetic pathways involved in NAFLD and identify potential therapeutic targets for intervention.

MATERIALS AND METHODS

The National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) is one of the largest databases available to researchers. The Search, Tag, Analyze, Resource GEO, or STARGEO, was developed to tag samples from the GEO database and produce robust meta-analyses. The GEO database is a genomics repository comprised of all published samples from omics studies. Briefly, STARGEO uses a standard random model for meta-analysis to generate both meta *p*-values and effects size across studies.[7] Study weight percentages were calculated using the inverse variance method *via* the DerSimonian-Laird estimate.[8] The STARGEO "Tagging" interface was used to gather samples under the "NAFLD," "NASH," and "NASH_NAFLD_Control" tag to conduct two separate meta-analyses: one comparing liver biopsies from NAFLD patients to healthy liver controls and the other comparing liver biopsies from NASH to healthy controls.

Series GSE48452 [9], GSE63067 [10], GSE66676 [11], and GSE107231 [12] were used to gather NAFLD, NASH, and healthy liver samples. Studies were found by searching NAFLD or NASH under human samples on stargeo.org. Studies selected for analysis had to meet the following criteria: Expression analysis was conducted on liver biopsies, the study included contained patients meeting NAFLD or NASH criteria and had matched healthy controls, and biopsies met definitive diagnosis of liver steatosis as below.

In these studies, liver biopsies were performed to diagnose liver disease and healthy liver biopsies were defined as having less than <5% steatosis and patients with evidence of viral hepatitis, alcoholic consumption, and hemochromatosis were excluded. Standard histopathological analysis by blinded pathologists were used to

defined NASH, NAFLD, and healthy liver samples.[13] For example, GSE48452 investigated intra-individual biopsies taken pre and post-bariatric surgery meeting NAFLD, NASH, and healthy liver criteria as above. Only pre-bariatric samples were tagged. For the NAFLD analysis, there was a total of 58 NAFLD liver biopsies and 60 healthy liver biopsies. The NASH analysis featured 187 NASH liver biopsies and 154 healthy liver biopsies.

We were able to extract approximately 20,000 genes for each of the meta-analyses conducted using STARGEO. We analyzed gene signature outputs with Ingenuity Pathway Analysis (IPA) to genes showing statistical significance ($p<0.05$) and an absolute experimental log ratio greater than 0.1 between case and control samples.[14] The genes included in our analysis are further detailed in Supplementary tables S1 and S2. IPA allowed us to define top canonical pathways, disease functions, disease networks, and potential upstream regulators that define NAFLD and NASH pathogenesis. Regulator analysis identifies upstream regulators that best explain the genetic expression in our dataset with *p-values* reflecting the degree of overlap of known effector targets and the gene signature analyzed in IPA. We also used the global molecular network feature of IPA to identify top disease networks. IPA ranks networks from the Global Molecular Network based on the number of focus genes from given networks that match with our analysis. Significance is represented by the p-score, as previously described.[14]

To find potential drug interactions, we used clue.io to analyze our dataset.[15] We inversed the gene expression pattern from the meta-analysis and used the "list-maker" function to identify drugs (see supplementary table 3). We focused on HEPG2 cell lines given they are immortalized HCC cells that relate most closely to the cells studied in our analysis.

All data analyzed were taken from Gene Expression Omnibus. There was no interaction or intervention with human subjects and no involvement with access to identifiable private patient information. As such, no Institutional Review Board (IRB) approval was necessary.

RESULTS

Top Canonical Pathways and Genes of Interest from NAFLD and NASH Analysis

From STARGEO, we were able to extract approximately 20,000 genes from our analysis of NAFLD and NASH liver biopsies compared to normal biopsy controls. Table 1 summarizes top upregulated and downregulated genes from the two analyses. Only genes that demonstrated statistically significant ($p<0.05$) differences in up-and down-regulation and absolute experimental log ratios of 0.1 were analyzed in IPA. Additionally, we used IPA to classify the top canonical pathways for NAFLD and NASH. P-values and experimental log ratios are included in supplemental tables S1 and S2.

For the NAFLD analysis, the genetic changes and top canonical pathways illustrate several disease processes such as dysregulated metabolism, immune cell recruitment, and altered signal transduction. IPA identified liver X receptor (LXR)/retinoid X receptor (RXR) activation (P = 4.35E-05), superpathway of cholesterol biosynthesis (P = 1.08E-04), granulocyte adhesion and diapedesis (2.34E-04), cAMP response element binding protein (CREB) signaling (P = 6.35E-04), and mevalonate pathway (P = 8.96E-04) as top canonical pathways. Among the most upregulated genes are the long non-coding RNAs (lncrna), X-inactive specific transcript (XIST), and LINC00885, with the role of lncrna in liver disease playing an increasing role.[16,17] Additionally, we found upregulation of tumorigenic proteins such as paternally expressed imprinted gene 10 (PEG10) and phosphodiesterase 11A (PDE11A).[18,19] We also noted dysregulated metabolism and increased lipogenesis through upregulation of inositol hexakisphosphate kinase (IP6K3), flavin containing monooxygenase 1 (FMO1), perilipin 1 (PLIN1), 3-hydroxy-3-methylglutaryl coenzyme A synthase (HGMCS1), HMG-CoA reductase (HMCR), fatty acid binding protein 5 (FABP5), and downregulation of insulin-like growth factor binding protein 2 (IGFB2), and insulin-like growth factor 1 (IGF1).[20–22] Upregulation of steroid 5-alpha reductase 2 (S5AR2) and downregulation of sex hormone-binding globulin (SHBG) and nuclear receptor subfamily 0 group B

member 2 (NR0B2) leads to higher androgen activity with implication in liver disease.[23] Interestingly, we found downregulation of the circadian rhythm gene period circadian regulator 3 (PER3).[24] There was also upregulation of several chemoattractants, including CXCL10. Lastly, we noted downregulation of the glycoprotein chitinase 3 Like 1 (CHI3L1), which regulates several cellular processes including proliferation, differentiation, inflammation, and others.[25]

Similarly, the gene expression changes and top canonical pathways from the NASH analysis detailed several pathologic processes. IPA identified cholesterol biosynthesis and insulin-like growth factor 1 (IGF-1) signaling as top canonical pathways. We found upregulation of genes involved in bile acid synthesis and carnitine synthesis including cholesterol 7 alpha hydroxylase (CYP7A1) and gamma-butyrobetaine hydroxylase 1 (BBOX1), respectively.[26–28] Notably, we saw upregulation of the novel myokine fibronectin type 3 (FNDC5), which correlated with NAFLD severity and extracellular matrix deposition.[29] Interestingly, we found upregulation of the lamin-associated gene muscular LMNA interacting protein (MLIP). Lamins and lamin-associated proteins have implications in liver disease.[30] We also found upregulation of several pro-inflammatory genes including the interleukin 13 receptor (IL1RA2) and the immunoglobin (Ig) receptor Fc alpha and mu receptor (FCAMR).[31] Our genetic analysis also highlighted dysregulated apoptosis through the downregulation of pro-apoptotic regulators such as the matricellular protein cysteine-rish angiogenic inducer 61 (CYR61), FOS protein (modulates JUN signaling), and Ras related dexamethasone induced 1 (RASD1) from the RAS family.[32–34] Lastly, we found downregulation of insulin-like growth factor binding protein-2 (IGFBP2), similar for our NAFLD analysis above, and the nicotinamide phosphoribosyltransferase (NAMPT), a rate-limiting enzyme in the NAD+ pathway.[35]

Top Disease Function and Networks

NAFLD and NASH are the result of several complex disease processes in tandem. To define these processes, we used IPA to identify top disease function and networks of

interest. In the NAFLD analysis, disease processes were largely related to lipid regulation, inflammation, and hepatic fibrosis and injury (table 2). Similarly, the disease functions in NASH included processes related to lipid metabolism in addition to other functions such as cancer and cell death and survival. Figure 1 illustrates one of the disease functions, adipogenesis, in the NASH analysis.

Next, we employed the IPA Disease Network feature to further elucidate the pathologic changes in NAFLD and NASH. IPA takes genes from the analyzed dataset and superimposes it onto curated information from the Ingenuity Knowledge base.[14] In table 3, we detail the top disease networks identified for NAFLD and NASH. Figure 2 details the lipid metabolism network from the NAFLD analysis.

Top Upstream Regulators and Causal Analysis

To propose potential drivers of NAFLD and NASH pathogenesis and their downstream effector genes, we used IPA Upstream Regulator analysis.[14] In the NAFLD analysis, beta-estradiol (P = 9.42E-12), cholesterol (1.79E-11), tumor necrosis factor (TNF) (P = 8.73E-10), nuclear receptor coactivator (NCOA2) (P = 1.22E-09), and SREBF1 (P = 12.8E-08) were top upstream regulators. Of these regulators, sterol regulatory element binding transcription factor 1 (SREBF1) demonstrated the highest z-score (2.200), demonstrating how the gene expression signature reflects known downstream SREBF1 gene signaling. Next, we investigated how the genes described above are affected by SREBF family (figure 3). We see activation of SREBF1 and SREBF2 is linked to the changes in expression noted in CXCL10, FABP5, IGF1, HMGCR, HMGCS1, sex hormone binding globulin (SHGB), and PEG10.

In the NASH analysis, TNF (P = 1.22E-19), lipopolysaccharide or LPS (P = 6.27E-16), beta estradiol (P = 1.42E-15, with predicted inhibition), interleukin 1B or IL1B (P = 1.78E-14), and platelet-derived growth factor BB or PDGF BB (P = 1.90E-14) were top upstream regulators. Beta-estradiol demonstrates anti-fibrotic effects in the liver, so we investigated its downstream effects in our dataset (figure 4). Inhibition of beta-estradiol activity is reflected by the changes we noted in the top upregulated genes including crystallin alpha A (CRYAA), BBOX1, CYP7A1, and FNDC5 and top downregulated genes including IGFBP2, nicotinamidenphosphoribosyltransferase pseudogene 1 (NAMP1), and RASD1 genes in our dataset described above. In addition, IPA related inhibition of beta-estradiol to other gene expression changes of interest including upregulation of the PEG10, squalene epoxidase (SQLE), inositol phosphokinase (IP6K3) and downregulation of the tumor suppressor Kruppel-like factor 6 (KLF6).[36,37]

Therapeutic Analysis

To investigate potential drug targets from our dataset, we utilized clue.io. We inputted genes that were both upregulated and downregulated in our NAFD and NASH dataset (see figure 5). We used the query tool from the platform and focused on HEPG2 cell lines, immortalized HCC cells. By looking at compounds that inverse the pathologic expression patterns in our meta-analyses, we identified riciribine (an AKT inhibitor) and ZSTK-474 (a PI3K inhibitor) as potential therapeutic compounds that target the genes in our investigation (see supplementary table 3).

DISCUSSION

NAFLD represents a growing health burden, with an astonishing prevalence of 25% of the global population.[38] A better understanding of pathogenesis is needed to tackle this herculean disease. Here, we use meta-analysis of public data using our STARGEO platform in search of insights to disease and potential therapeutic targets. The gene expression profiles from our analyses can elucidate function and regulatory patterns to disease.[39] Our results from the NAFLD analysis reinforce the role of altered metabolism, inflammation, and cell survival and supports recently described contributors to disease such as altered androgen and lncrna activity.[17,23,40]

Our results demonstrated several changes that are implicated in altered lipid and metabolic homeostasis. It is the accumulation of lipids that lead to several downstream effects that characterized NAFLD development and progression.[20,41] For example, lipid droplets in hepatocytes can lead to hepatic insulin resistance, decreased autophagy, oxidative stress, and interaction with several transcription factors such as SREBF.[20] These lipid droplets can form through the activity of proteins from the perilipin family, such as perilipin 1 (PLN1), which was upregulated in our dataset (table S2).[20] "Superpathway of Cholesterol Biosynthesis" was one of the top canonical pathways, and several top disease functions and networks were related to lipid accumulation (tables 1-3). In addition, cholesterol and SREBF1, a transcription factor involved in lipid homeostasis, were top upstream regulators.[42] SREBF1 stimulates accumulation of lipids in hepatocytes through activation of patatin-like phospholipase domain-containing 3 (PNPAL3).[43] In our results, we illustrate how downstream signaling of SREBF1 and SREBF2 Leads to fatty acid accumulation and other disease functions. Downstream of SREBF1 and SREBF2 signaling, we noted upregulation of CXCL10, HMGCR, HMGCS1, FABP5, and PEG10 in addition to downregulation of SHBG and IGF1. HMGCR catalyzes the first reaction of cholesterol synthesis and HMGCS1 also contributes to hepatic cholesterol synthesis. Increased activity of HMGCSR and HMGCS1 was associated with NAFLD and with fatty acid accumulation.[44] Additionally, FABP5 is a fatty acid binder normally expressed in adipocytes, but

expression in hepatocytes was correlated with fatty acid infiltration in NAFLD.[45] In addition to fatty acid changes, we found downregulation of IGF-1, which leads to hyperglycemia and increases risk for diabetes seen as in NAFLD.[46,47] IGF-1 also has anti-fibrotic effects through attenuation of hepatic stellate cell (HSC) activation in murine models.[48] Furthermore, SHBG was downregulated in analysis, with decreased SHBG levels being associated with inreased insulin resistance in NAFLD patients.[49] Higher levels of SHBG are also associated with lower odds for NAFLD and may have some protective effect.[50] In addition to fatty acid accumulation and glycemia, we related SREBF activity to malignant changes through upregulation of PEG10. PEG10 is a transcription factor that was found to be an oncogene in several solid cancers such as HCC, gastric cancer, and breast carcinoma.[36] PEG10 is upregulated in NASH and NAFLD and may be associated with increased risk for HCC seen in this patient population.[18] Furthermore, our results fortified other changes in NAFLD that are implicated in lipolytic changes that may induce NAFLD. CREB signaling was identified as a top canonical pathway. Awaad, *et al* showed elevated cAMP and CREB levels in a NAFLD murine model and suggest the role of cAMP and CREB as a marker of early NAFLD.[51]

Accumulation of fatty acids in the liver induces chronic, low-grade inflammation, and subsequently, progression of NAFLD to NASH. Our results illustrated the inflammatory changes in NAFLD. The inflammatory response was a top disease function in our analysis (table 3) and the pro-inflammatory cytokine TNF was a top upstream regulator. In murine models, TNF plays an essential role in NAFLD development through upregulation of inflammatory mediators and genes associated with liver fibrosis.[52] TNF also induces hepatic steatosis in murine models through upregulation of SREB proteins.[53] We also noted upregulation of several pro-inflammatory cytokines in our analysis including CXCL10 (table 1, table S1). CXCL10 recruits T cells and macrophages and is an independent risk factor for NASH.[54] Since fatty acids lead to inflammatory changes, it is expected that SREB signaling would lead to downstream pro-inflammatory changes such as upregulation of CXCL10 (figure 3).

Aside from inflammation and metabolic derangements, our results illustrated several other signaling and cellular processes of interest in NAFLD. One such cellular process is protein prenylation. Protein prenylation is a protein post-translational modification where farnesyl (farnesylation) or geranylgeranyl (geranylgeranylation) side chain is added to a C-terminal cysteine residue.[55] The mevalonate pathway, a top canonical pathway in our analysis, affects the ratio of farnesylation and geranylgeranylation. Alteration in this ratio is implicated in NAFLD and NAFLD-associated fibrosis.[56] In addition to post-translational protein modification, our results suggests a role for lncRNAs in NAFLD. LnRNAs are critical mediators of normal liver physiology, with aberrant expression being observed in metabolic, fibrotic, and malignant hepatic changes.[17] We found upregulation of lncRNAs in our analysis, including XIST and LINC0085. XIST is one of the earliest described lnRNAs and assists in the formation of silenced heterochromatin.[57] While not well-described in NAFLD and NASH, XIST has been shown to promote HCC and colorectal cancer.[58,59] Additionally, LINC0085 is a positive cell growth regulator in breast cancer models and may, alongside XIST, cause proliferative and pathologic changes in hepatocytes in NAFLD and NASH.[16] Lastly, recent research has connected the link between circadian rhythm genes with NAFLD.[24] Asynchronization of circadian rhythms, such as from shift work, are correlated with higher prevalence and NAFLD.[60] Per3 is a circadian rhythm gene that regulates adipogenesis, with deletion leading to increased adipogenesis in animal models.[24] Thus, downregulation of Per3 in our results may suggest dysregulation of circadian rhythm and consequent changes in regulation of adipogenesis.

NASH is a subset of NAFLD characterized by steatosis inflammation and fibrosis.[61] It typically takes years for NAFLD to progress for NASH, and while the mechanisms behind this progression are not clear, our current understanding suggests a "multi-hit hypothesis" where multiple modes of fatty acid accumulation and oxidative stress synergistically induce liver inflammation and fibrosis.[61] Aside from lifestyle modifications, obeticholic acid (OCA) is the only FDA-approved treatment of NASH.[62]

The growing burden of NASH necessitates new therapeutics and our analysis of NASH offers insight into potential treatment.

Ingenuity Pathway Analysis of our NASH dataset reinforces the role of cholesterol. Several of our top canonical pathways, disease network, and disease functions were related to cholesterol synthesis, lipid metabolism, adipogenesis, and metabolic disease (figure 1, and tables 1, 3, and 4). The role of lipids in liver injury have been described above.[20,40] Other disease functions and disease networks of note involved cell death and survival, cancer, digestive system disease, and organismal injury (tables 3 and 4). The top upstream regulators in addition to upregulated and downregulated genes reflect activity related to these disease functions. Among our top upstream regulators were pro-inflammatory cytokines TNF and IL1B, PDGF BB, and beta-estradiol (with predicted inhibition).

As already described, inflammation is a major contributor to liver disease. It has been long shown that patients with NASH, and more so those with severe NASH, have elevated levels of TNF.[63] Elevated serum levels of TNF in NASH patients was linked to increased major adverse hepatic events.[64] While TNF inhibition reduces steatosis and fibrosis in murine models, their role in select NAFLD and NASH patient populations has still not been proven effective.[52,65–67] Similarly, IL1B signaling has pro-fibrotic and lipogenic effects in murine models and may have promise as directed therapy in NASH patients.[68][69,70] Lastly, the cytokine PDGF BB exerts its pro-fibrotic effects through activation of hepatic stellate cells and, consequently, is another potential drug target.[71]

Experimental models have shown that estrogen has protective, anti-fibrotic activity through attenuation of HSC activation and generation of reactive oxygen species.[72] Additionally, estrogen receptor agonism in a NASH murine model had therapeutic effects through modulating bile acid receptor signaling and inhibiting fibrosis and adipogenesis.[73] Interestingly, decreased estrogen levels and other hormone changes in menopause may be related to increase risk for NAFLD and NASH.[74] Since beta-estradiol was a top upstream regulator with predicted inhibition in our NASH

analysis, we applied IPA to investigate beta-estradiol signaling and its downstream genetic effects (figure 4).

Our analysis related inhibition of beta-estradiol to derangement of several cellular processes downstream including metabolism, extracellular matrix deposition, and tumor suppression. In regard to metabolism, we related inhibition of estradiol to upregulation of IP6K3, CYP7A1, and SQLE and to downregulation of NAMP1 and IGFBP2. IP6K3 produces inositol pyrophosphates and regulates metabolic control.[75] Deletion in murine models leads to improved glucose tolerance, reduced body weight, and protection from fatty liver disease.[75,76] SQLE is involved in cholesterol synthesis and has been shown in both human and animal studies to promote development of HCC in fatty liver disease.[77] CYP7A1 is a rate-limiting enzyme in the classical pathway of bile acid synthesis with upregulated gene expression in NAFLD and NASH patients alike, but discrepancies exist in post-transcriptional protein levels.[27] The effects of fatty liver disease on CYP7A1 are inconsistent, but bile acid dysregulation is a growing hallmark in this disease.[27] NAMP1 is a critical enzyme in the synthesis of nicotinamide adenine dinucleotide (NAD+). NAD+ functions in mitochondrial oxidative phosphorylation and protection of cells from reactive oxygen species.[78] Depletion of hepatic NAD+ has been shown to be a risk factor for NAFLD in a murine model.[35] There is growing interest in targeting NAD+ in NAFLD.[79] Lastly, IGFBP2 binds to IGF1 and has a positive effects in glucose control.[80] Early epigenetic silencing, *via* methylation, of IGBFP2 predicts development of fatty liver later in mice.[81]

In addition to metabolic changes, our analysis showed pro-oncogenic and fibrotic genetic changes in NASH that may relate to inhibition of beta-estradiol signaling. Through IPA, we correlated inhibition of beta-estradiol signaling to upregulation of PEG10 and FNDC5 and to downregulation of RASD1 and KLF6. Interestingly, we found upregulation of PEG10 in our NAFLD analysis and discussed its pro-oncogenic activity. RASD1 is a member of the Ras superfamily of G proteins that regulate signal transduction through G-protein coupled receptors.[82] RASD1 prevents aberrant cell growth, and its downregulation may lead to increased risk for HCC seen in fatty liver

disease.[34,83] Additionally, KFL6 is a zinc finger transcriptional protein with tumor suppressor function that is inhibited in various cancers, including HCC.[37] Lastly, FNDC5 is a novel myokine that controls extracellular matrix deposition. Higher expression of FNDC5 in HSCs correlated to severity of fibrosis in NAFLD patients.[29] Our results illustrate malignant and fibrotic gene expression changes in both NAFLD and NASH stages of disease and its possible relation with inhibition of beta-estradiol signaling.

Lastly, our analysis suggests potential use of riciribine and ZSTK-474 in the treatment of NAFLD. Dysregulation of the PI3KT/AKT pathway in hepatocytes has been described in NAFLD.[84] Such dysregulation are implicated in hepatic steatosis and fibrosis. While the mechanisms underpinning pathogenesis through the PI3KT/AKT pathway are still under investigation, our results add further evidence of targeting this pathway for therapeutic benefit.

Our meta-analysis approach offers insights into NAFLD and NASH, but this approach is not without limitations. Biological samples in Gene Expression Omnibus have limitations in terms of description of samples. Some details that may present confounding variables are the co-morbidities in patients and differing stages in fatty liver disease, including degree of fibrosis. Other patient characteristics may also influence results such as medications, age, gender, and ethnicity. Samples were also taken under different conditions such as diagnosis of undifferentiated liver disease or in bariatric patients, which may lead to further differences between samples. Though there are set diagnostic criteria for hepatic steatosis on biopsy, the diagnoses were made by separate pathologists across these studies and a meta-analysis approach would not be able to account for these differences. Additionally, while transcriptomic and meta-analysis studies can offer a global view of disease function and regulatory signaling using gene expression patterns, causality necessitates more direct functional experimentation.[39] This approach itself does not offer direct experimental or clinical evidence. Nonetheless, our results offer a foundation to future studies in NAFLD and NASH that warrant further investigation with experimental and human models.

CONCLUSION

We utilized our platform STARGEO to produce genetic signatures from GEO datasets that provide molecular insights to fatty liver disease. We conducted to separate analysis of NAFLD and NASH liver biopsies to investigate genetic changes that define stages of fatty liver disease. Our analyses buttresses how the dysregulation in lipid homeostasis, though such regulators as the transcription factor SREBF1, contribute to steatosis. We also noted upregulation of genes implicated in oncogenesis, such as PEG10, that may partly explain the increased risk of HCC in these patients. We also describe the potential contribution on long noncoding RNAs in NAFLD pathogenesis. From our NASH analysis, we explored how beta-estradiol dysregulation may mechanistically contribute to steatosis and its several consequences such as fibrosis and oncogenesis. Lastly, we used out dataset and clue.io to identify genes that target pathologic genetic changes and signaling, such as PI3KT/AKT signaling, and found ricirbine and ZKST-474 as possible therapeutic targets. Overall, our analysis illustrates several changes that may explain progression of NAFLD pathogenesis and promising directions that warrant further investigation.

ARTICLE HIGHLIGHTS

Research background

NAFLD pathogenesis is poorly understood but may result from a mix of exogenous and genetic factors that lead to fatty infiltration and inflammation.

Research motivation

NAFLD is a growing cause for liver transplant with limited therapeutic options.

Research objectives

Define genetic changes that underlie NAFLD and progression to NASH in pursuit of identifying promising therapeutic targets.

Research methods

We employed our STARGEO platform to conduct meta-analyses of publicly available liver biopsies from NAFLD and NASH patients.

Research results

We identified various genes implicated in inflammation and fatty infiltration, as well as signaling processes that lead to these changes. We also identified riciribine and ZSTK-474 as potential drugs.

Research conclusions

NAFLD and its progression to NASH is likely led by several genetic changes detailed in our manuscript. The genetic changes in our dataset are targeted by ricirbine and ZSTK-474 and warrants further study.

Research perspectives

As NAFLD becomes an increasing clinical burden, a bioinformatics approach is valuable in understanding causes and elucidating treatment avenues.

76109_Auto_Edited.docx

ORIGINALITY REPORT

3%
SIMILARITY INDEX

PRIMARY SOURCES

1	Kamal Khorfan, Jihad Aljabban, Imad Aljabban, Laraib Z. Safeer, Maryam Panahiazar, Dexter Hadley. "Su1494 - Meta-Analysis Reveals Multiple Disease Processes Underlying Non-Alcoholic Fatty Liver Disease", Gastroenterology, 2018 Crossref	79 words — 2%
2	aasldpubs.onlinelibrary.wiley.com Internet	45 words — 1%
3	www.science.gov Internet	29 words — 1%
4	DalSpace.library.dal.ca Internet	13 words — < 1%
5	www.reportworld.co.kr Internet	12 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	< 12 WORDS
EXCLUDE BIBLIOGRAPHY	OFF	EXCLUDE MATCHES	< 12 WORDS