

RESEARCH

Open Access



Identification of key modules and driving genes in nonalcoholic fatty liver disease by weighted gene co-expression network analysis

Zhengmao Song^{1†}, Yun Wang^{1†}, Pingli Lin^{1†}, Kaichun Yang^{1†}, Xilin Jiang^{2,3†}, Junchen Dong³, Shangjin Xie⁴, Rong Rao^{1*}, Lishan Cui^{1*}, Feng Liu^{1,4*} and Xuefeng Huang^{2*}

Abstract

Background Nonalcoholic fatty liver disease (NAFLD) is characterized by excessive liver fat deposition, and progresses to liver cirrhosis, and even hepatocellular carcinoma. However, the invasive diagnosis of NAFLD with histopathological evaluation remains risky. This study investigated potential genes correlated with NAFLD, which may serve as diagnostic biomarkers and even potential treatment targets.

Methods The weighted gene co-expression network analysis (WGCNA) was constructed based on dataset E-MEXP-3291. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were performed to evaluate the function of genes.

Results Blue module was positively correlated, and turquoise module negatively correlated with the severity of NAFLD. Furthermore, 8 driving genes (ANXA9, FBXO2, ORAI3, NAGS, C/EBP α , CRYAA, GOLM1, TRIM14) were identified from the overlap of genes in blue module and GSE89632. And another 8 driving genes were identified from the overlap of turquoise module and GSE89632. Among these driving genes, C/EBP α (CCAAT/enhancer binding protein α) was the most notable. By validating the expression of C/EBP α in the liver of NAFLD mice using immunohistochemistry, we discovered a significant upregulation of C/EBP α protein in NAFLD.

Conclusion we identified two modules and 16 driving genes associated with the progression of NAFLD, and confirmed the protein expression of C/EBP α , which had been paid little attention to in the context of NAFLD, in the

[†]Zhengmao Song, Yun Wang, Pingli Lin, Kaichun Yang and Xilin Jiang are co-first authors and contributed equally to this work.

*Correspondence:

Rong Rao
rongrao317@126.com
Lishan Cui
cuilishan3116@126.com
Feng Liu
liufeng@xmsdwyy.cn
Xuefeng Huang
hxfzy2001@139.com

Full list of author information is available at the end of the article



present study. Our study will advance the understanding of NAFLD. Moreover, these driving genes may serve as biomarkers and therapeutic targets of NAFLD.

Keywords C/EBP α , WGCNA, NAFLD, Bioinformatics analysis

Background

Nonalcoholic fatty liver disease (NAFLD), characterized by excessive liver fat deposition, is a continuous disease spectrum, including simple steatosis, nonalcoholic steatohepatitis (NASH), relevant liver cirrhosis, and even hepatocellular carcinoma in severe cases [1, 2]. NAFLD accounts for 75% of chronic liver disease cases, and is also a common cause of liver transplantation [3–5]. With changes in modern lifestyles, such as high energy intake and sedentary activities, the incidence of NAFLD is rapidly increasing [6]. In addition, NAFLD increases susceptibility to chronic kidney disease, sarcopenia, hyperuricemia, type 2 diabetes and other metabolic diseases and malignancies [7, 8]. Hence, NAFLD is paid much attention and many efforts have been made for its diagnosis and treatment [9].

The current gold standard for diagnosing NAFLD is histopathological evaluation of liver tissue biopsy, which is invasive and risky and with sampling errors [10]. It is challenging and yet tempting to seek non-invasive diagnostic biomarkers with easy detection and high accuracy for diagnosis and even potential treatment of NAFLD [11].

Thanks to the great strides made in bioinformatics in recent decades, we can analyze large and complex gene sequencing data, which has been accepted as an important method in life science research [12–15]. Weighted gene coexpression network analysis (WGCNA) is a novel bioinformatics method that explores the correlations between or within different genomes, as well as the correlations between genes and clinical features, by establishing co-expression modules or gene networks [16–18]. The modules are established based on differences in expression profiles and driving genes that are critical in triggering key cell signaling pathways in important types of cells [19]. It recognizes highly correlated modules, characteristics of gene modules and driving genes. WGCNA contributes to establishing correlations between gene modules and samples and to calculating module membership [20]. At present, WGCNA has been successfully applied in analyses of cancers (e.g., breast cancer, glioblastoma and prostate cancer) [21–23]. By investigating the correlations between tissue microarray data and clinical features, WGCNA predicts the survival outcomes of cancer patients and identifies candidate biomarkers or therapeutic targets of cancers [24, 25].

In the present study, we analyzed the E-MEXP-3291 dataset using WGCNA. After establishing the correlations between gene modules and clinical data of NAFLD,

it was found that the blue module was positively correlated with the severity of NAFLD. Subsequently, the overlap of genes in blue module and upregulated genes in GSE89632 was searched, and 8 driving genes, such as CCAAT/enhancer binding protein (C/EBP α), etc., was identified. After establishing a NAFLD model in mice, immunohistochemical data validated significantly upregulated C/EBP α in the liver tissues of NAFLD mice. We also identified turquoise module and 8 driving genes to be negatively associated with the severity of NAFLD, indicating the regression of the disease. Taken together, our findings provide novel directions and therapeutic targets of NAFLD.

Methods

Dataset acquisition and data preprocessing

The RNA microarray dataset GSE89632 [25,581,263] and the E-MEXP-3291 [21,737,566] profile were downloaded from the GEO (Gene Expression Omnibus) database (<https://www.ncbi.nlm.nih.gov/geo>) and Array-Express (<https://www.ebi.ac.uk/arrayexpress/>). The gene expression level of 24 healthy controls, 20 cases with simple steatosis and 19 cases with NASH were included in GSE89632, besides the steatosis percentage, fibrosis stage, lobular inflammation severity, ballooning intensity, NAFLD activity score, age, sex, liver arachidonic acid level, liver eicosapentaenoic acid level, and liver docosahexaenoic acid level. The gene expression level of 19 normal liver, 10 simple hepatic steatosis sample, 9 NASH with fatty liver and 7 NASH without fatty liver, containing gender and age, were included in the E-MEXP-3291. The summary of the datasets was provided in supplementary Table 1. In the present study, 19 normal liver, 10 simple hepatic steatosis sample and 9 NASH with fatty liver expression profile was used to construct the WGCNA network, and the obtained results were further validated using the GSE89632 dataset.

Establishment of the NAFLD mouse model

Animal procedures were strictly performed based on the ethics review of experimental animals and were approved by the Ethics Committee of Xiamen University. All methods were carried out in accordance with relevant guidelines and regulations of the Ethics Committee, and were reported in accordance with ARRIVE guidelines (Animal Research: Reporting of In Vivo Experiments, <https://arriveguidelines.org>) for the reporting of animal experiments.

Male adult C57BL/6J mice weighing 24 ± 2 g was habituated to a standard environment [temperature: 21 ± 2 °C, humidity: $60 \pm 10\%$, light/dark cycle: 12 h/d (8:00–20:00)]. Mice were given free access to water and diet. They were randomly assigned into two groups and fed either a normal diet or a high-fat diet.

Immunohistopathological characterization of liver.

Livers were sectioned from mice, washed with PBS, fixed with 4% paraformaldehyde/PBS and paraffin embedded. For C/EBP α immunofluorescences, cross-sections were treated for antigen retrieval and incubated with primary antibodies (1:100) followed by fluorescent secondary antibody. Peroxidase activity was revealed by 3-30-diamino- benzidinetetrahydrochloride (DAB, Dako). Images were captured using an Upright Metallurgical Microscope (Leica DM4B, Germany). Negative controls were carried out by omitting the primary antibody.

Construction of WGCNA

The WGCNA R software package was constructed. In brief, genes with expression values >10 in 43 samples were utilized to draw a hierarchical clustering tree (dendrogram) using the *fashClust* function. The soft-thresholding power selected by the *pickSoft* Threshold function was a standard value in the scale-free topology network to make the established network a power-law distribution. It reduced errors and made the results more characteristic of biological data by strengthening strong correlations and weakening weak correlations in a scale-free network feature. The scale-free topology fit index presented an exponential change. Therefore, a good correlation ($R^2=1$) indicated that the data network was in a scale-free topological distribution.

Clinically significant modules

Key modules were screened out by calculating the correlations between module eigengenes and clinical traits. In the linear regression between gene expression and clinical information, \log_{10} transformation of the *P*-value ($GS = \lg P$) was considered gene significance (GS). The average GS of all genes in one module was considered module significance (MS). The module with the highest MS among all modules was believed to be the one that has a most significant correlation with clinical traits.

Function enrichment analysis

To obtain further insight into the function of genes in key modules, Gene Ontology (GO) enrichment analysis was performed for modules with the KOBAS tool (<http://kobas.cbi.pku.edu.cn/kobas3>). The gene lists of modules were uploaded, and we obtained the results of BP and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses. An adjusted *p*-value <0.05 was regarded as significant.

Validation of driving genes

The “limma” R package was used to screen the differentially expressed genes (DEGs) between healthy control samples and NASH samples in dataset GSE89632 for validation. The cutoff value was $\log_2^{|\text{FC}|} > 1$, with an adjusted *P*-value <0.05 . The construction of the volcano plot and hierarchical clustering analysis were carried out using the R packages *ggplot2* and *pheatmap*, respectively. A Venn diagram was created using the online tool *jvenn* (<http://jvenn.toulouse.inra.fr/app/example.html>) to overlap the genes in key modules and differentially-expressed genes (DEGs).

Statistical analysis

The results were expressed as the mean \pm S.E.M. Statistical analysis was performed using unpaired T test for comparisons between two groups followed by the Student-Newman-Keuls test (Prism 5 for Windows, GraphPad Software Inc., USA). *P* values <0.05 were considered statistically significant.

Results

Expression value analysis of microarray data

The E-MEXP-3291 profile containing 43 cases was downloaded from ArrayExpress, including 20 healthy controls, 15 cases with steatosis and 8 cases with NASH. Using the R package, raw data of the E-MEXP-3291 profile were processed for background correction and normalization. Probes and gene symbols were matched using R package annotation. For the multi-matched genes, the median level was regarded as the final expression value. A total of 23,486 genes were identified, and those with an average expression level >5 were selected for the following analysis. Ultimately, 6,731 eligible genes were included for cluster analysis. As shown in Fig. 1A, three clusters of 43 samples were classified.

Construction of WGCNA and identification of key modules

An appropriate soft threshold value was screened out to make the established network a scale-free distribution. The network topology analysis was conducted on the top 20 thresholding powers, aiming to identify the relatively balanced scale independence and mean connectivity of the WGCNA. The power value (β) was confirmed to be 9 (Fig. 1B and C) to produce a hierarchical clustering tree of 6731 genes.

The obtained adjacent and topological modules were subjected to a gene clustering function using dissimilarity. Subsequently, modules were cut by the dynamic-prune algorithm for the establishment of the WGCNA network. Similar modules were merged as the *MEDisThres* set for 0.25, and 7 modules were generated (Fig. 2A and B). Notably, the gray module represented genes that were unable to be allocated to modules.

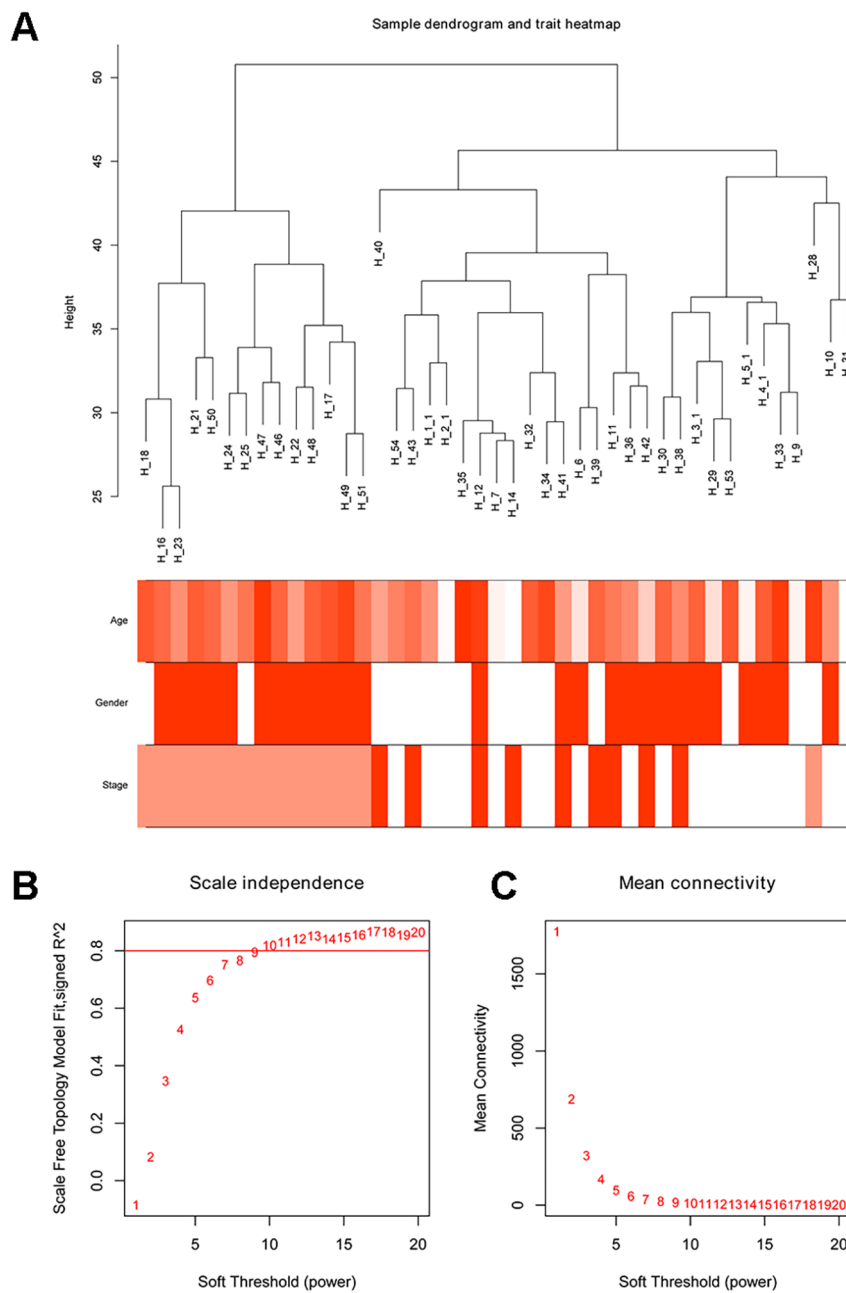


Fig. 1 Sample clustering and soft-thresholding power determination. **A** Clustering was based on the expression data of E-MEXP-3291, and the color intensity was proportional to disease status (healthy controls, simple steatosis and NASH), sex and age. **B** Analysis of the scale-free fit index for various soft-thresholding powers (β). **C** Analysis of the mean connectivity for various soft-thresholding powers. In all, 9 was the most fit power value

According to hierarchical clustering, different colors represent different modules, and those on the top are initially obtained modules through the dynamic-prune algorithm, while those on the bottom are the final merged modules. In detail, there were 609, 1154, 653, 1701, 1234, 527 and 853 genes in the black, blue, brown, green, grey, red and turquoise modules, respectively.

Correlation between modules and key module identification

Interaction association was analyzed among the seven modules, and a network heatmap was depicted (Fig. 3A). It is concluded that every module was validated independently to each other, indicating a relative independence of different genes in different modules. Subsequently, co-expression similarity in modules was investigated

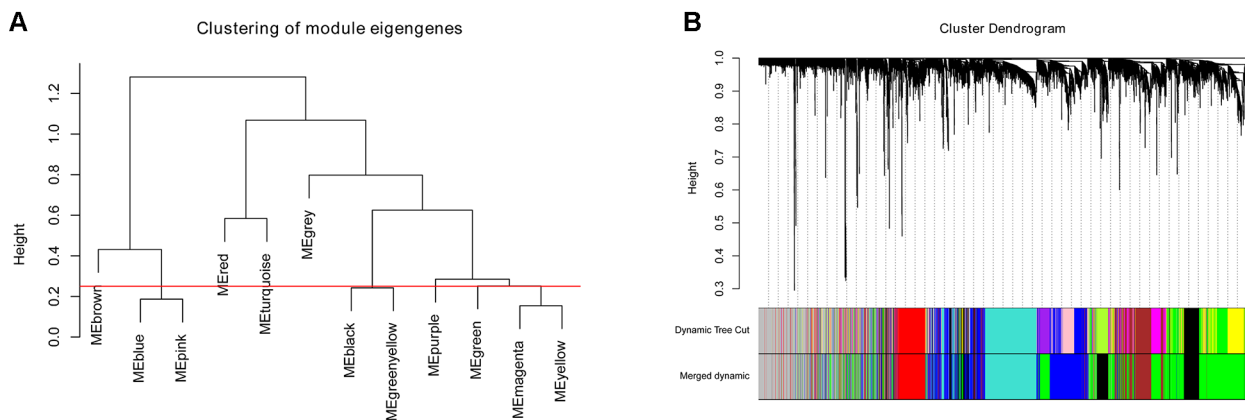


Fig. 2 Co-expression module construction using the WGCNA package in R. **A** The cluster dendrogram of module eigengenes. **B** The cluster dendrogram of genes in E-MEXP-3291. Each branch in the Fig. represents one gene, and each color represents one coexpression module

by calculating eigengenes and clustering them based on the correlation, and two main clusters were obtained (Fig. 3B). In addition, the heatmap of driving gene network, depicted based on adjacencies, reveals similar results (Fig. 3C). In the present study, age, sex and stage were included as clinical traits. Pearson correlation analysis was performed on modules and clinical traits, where modules (clinical traits) were expressed as rows and the status was expressed as columns. Values in the modules represent the correlation and p -value. As shown in Fig. 3D, the blue module was positively correlated with stage, and the turquoise module was negatively correlated with stage, suggesting that the blue module could promote the progression of NASH and that the turquoise module could inhibit the progression of NASH. The correlations between module membership and GS in the blue and turquoise modules were shown in supplementary Fig. 1. Therefore, blue and turquoise modules were ultimately selected for the following analysis.

Function enrichment analysis

As the blue module was positively correlated with disease stage, the gene in the blue was enrolled for further KEGG and GO analysis, p value < 0.05 and FDR < 0.05 were considered statistically significant. A total of 62 biological processes and 57 pathways were enriched in blue module. The blue module was mainly enriched in the regulation of the Wnt, MAPK and AMPK pathways, etc. (Fig. 4A and B). However, we did not identify significantly enriched pathways in the turquoise module.

Identification of driving genes in the blue module

The GSE89632 dataset was downloaded from the GEO database to identify the expression of driving genes. DEGs were screened out ($\log_2FC > |1|$ and $P < 0.05$) as described [26, 27], which are depicted as volcano plots (Fig. 5A) and shown in a hierarchical clustering heatmap

(Fig. 5B). Subsequently, a Venn diagram was constructed for the overlapping upregulated genes and genes in the blue module, and 8 overlapping driving genes (*ANXA9*, *FBXO2*, *ORAI3*, *NAGS*, *C/EBP α* , *CRYAA*, *GOLM1*, *TRIM14*) were obtained (Fig. 5C). The expression levels of the 8 driving genes in healthy controls and NASH cases from GSE89632 are shown in Fig. 5D, and *C/EBP α* was the most upregulated gene ($\log_2FC = 3.33$).

Likewise, the overlapping upregulated genes in GSE89632 and genes in the turquoise module were analyzed, and 8 overlapping driving genes (*GPR3*, *FOSL1*, *ETNK1*, *C2CD4B*, *CSF3*, *FOXCl*, *SOX17*, *IER5L*) were obtained (Fig. 5E).

Experimental validation of driving genes in the blue module

To verify our identifications, the expression of CCAAT/enhancer binding protein- α (*C/EBP α*), the most upregulated gene in blue module, was determined in the NASH model mouse and normal controls. In Fig. 6A, the brown color represents the positive staining of *C/EBP α* . The intensity was recognized and read using the image J software, and the value was then divided by the mean of that in normal group, resulting in the fold change of normal group (Fig. 6B). The immunohistochemistry result showed that the protein level of *C/EBP α* was upregulated in the NASH group compared with the control group in this mouse model, consistent with our bioinformatics analysis.

Discussion

NAFLD has emerged as the leading cause of chronic liver disease in many countries worldwide. NAFLD represents a spectrum of disease severity, ranging from simple steatosis to NASH, cirrhosis, and hepatocellular carcinoma (HCC) [28]. Compared with the general population, NAFLD patients are at increased risk of liver-related,

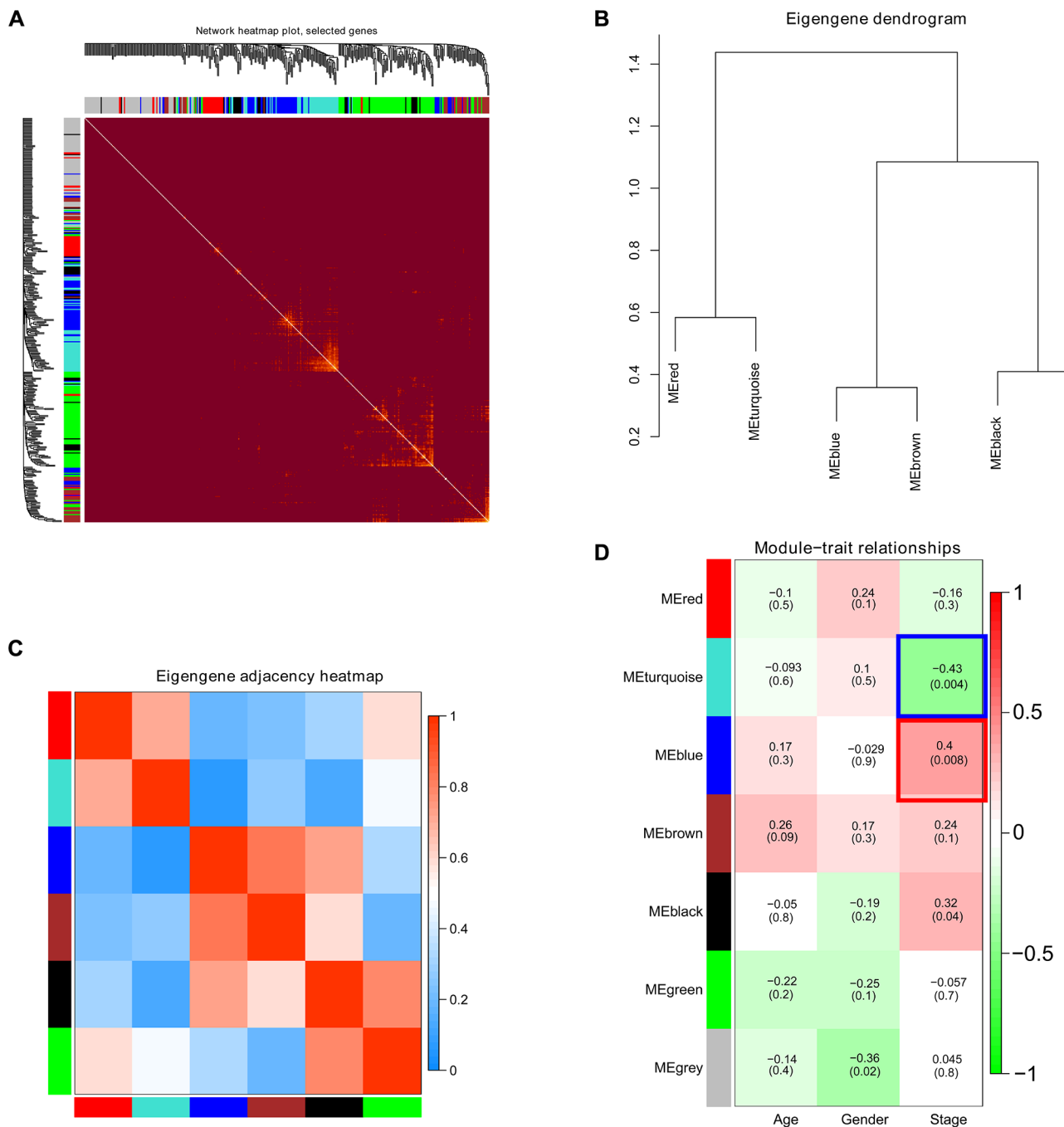


Fig. 3 Correlation between modules and key module identification. **A** Interaction relationship analysis of coexpressed genes. Different colors of the horizontal and vertical axes represent different modules. The brightness line in the middle represents the degree of connectivity of different modules. There was no significant difference in interactions among different modules, indicating a high-scale independence degree among these modules. **B** Hierarchical clustering of module genes that summarize the modules yielded in the clustering analysis. The modules with similarity over 0.2 was incorporated before clustering. **C** Heatmap plot of the adjacencies in the driving gene network. **D** Heatmap of the correlation between module eigengenes and the disease status of NAFLD. The turquoise module was the most negatively correlated with status, and the blue module was the most positively correlated with status

kidney-related, cardiovascular and all-cause mortality [29, 30]. However, with complex multifactorial pathogenesis, the genes or proteins related to progression of NAFLD remain obscure.

In recent years, the identification of key genes of a certain disease using WGCNA has become popular [31]. Establishing a WGCNA network contributes to screening and identifying key modules and genes that are

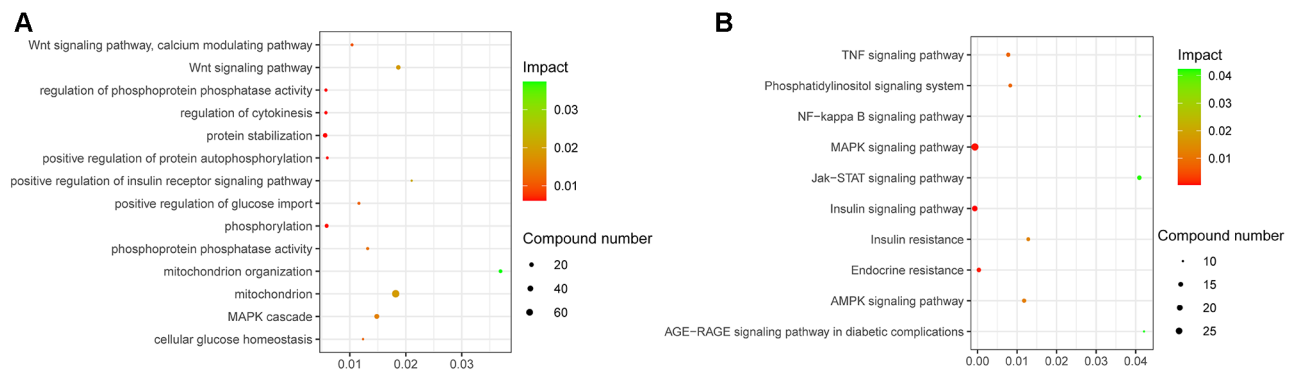


Fig. 4 GO and KEGG analysis of blue module. **A** GO analysis of the enriched genes in the blue module. **B** The KEGG pathways of the enriched genes in the blue module

responsible for specific features of a disease [32]. Traditional gene analysis mainly focuses on strong effector molecules rather than weak ones, although they are of significance as well [33, 34]. WGCNA is a supplemental method for data mining of weak effector molecules. It strengthens the correlation of strong effector molecules after power function processing and conversely weakens that of weak effector molecules in the same processing, thus leading to a scale-free topology criterion of networks [35].

In the current study, we apply WGCNA on two datasets, that is E-MEXP-3291 and GSE89632. Both of them contain samples of healthy controls, simple steatosis and NASH. Although simple steatosis and NASH are parts of, not identical to, NAFLD, they are important components of NAFLD, and studies of E-MEXP-3291 and GSE89632 also use the term NAFLD to include simple steatosis and NASH [36, 37], hence in the current study, we use expression of NAFLD to describe the clinical information/trait of the disease status in these datasets.

The RNA-seq dataset E-MEXP-3291 was downloaded from the ArrayExpress database. A total of 6,731 genes were screened out after excluding genes with an average expression level < 5, and these genes were subjected to WGCNA. Modules correlated with NAFLD were identified through cluster analysis. The data showed that the blue and turquoise modules were correlated with the stage of NAFLD. Afterwards, the two modules were subjected to GO and KEGG analysis, and they were determined to be mainly enriched in fat metabolism, insulin resistance and other biological processes that were of great significance in the development of NAFLD.

Then we investigated driving genes in each module. The term of driving genes is similar to hub genes and yet a bit different from the latter. The driving genes in the present study are the overlaps between modules and DEGs and related to the disease; however, in general, the hub genes refer to those genes participating in the transcriptional regulation [38, 39]. To ascertain driving genes

in each module, another dataset, GSE89632, was introduced to identify differentially expressed genes between healthy controls and NASH patients. The GSE89632 dataset serves as the external validation set to ensure the stability of the results. By overlapping upregulated differentially expressed genes and genes in the blue module, a total of 8 genes were obtained. Among them, *C/EBP α* was the top upregulated gene. To further validate our findings, we detected positive expression of *C/EBP α* in liver tissues of NAFLD mice by immunohistochemical staining. As expected, *C/EBP α* was significantly upregulated in NAFLD mice compared with mice fed a normal diet.

CCAAT/enhancer binding protein (*C/EBP*) is a eukaryotic transcription factor containing six family members (*C/EBP α* , *C/EBP β* , *C/EBP δ* , *C/EBP ϵ* , *C/EBP γ* and *C/EBP ζ*) [40]. These proteins are extensively distributed in various types of tissues, organs and cells. Functionally, *C/EBP α* is involved in hepatocyte proliferation and adipocyte differentiation [41]. *C/EBP β* is necessary for the immune function of macrophages [42]. *C/EBP δ* is synergistically involved in adipocyte differentiation [43]. And *C/EBP ϵ* is specifically expressed in bone marrow cells, serving as a vital mediator of granulocyte production [44]. But *C/EBP γ* and *C/EBP ζ* has been little studied.

Structurally, the basic leucine zipper (bZIP) at the C-terminus of *C/EBP* family members is highly conserved and is responsible for dimerization and DNA binding [45]. Its heterodimer or homodimer regulates gene transcription by binding the conserved sequence 5'-T(T/G)NNGNAA(T/G)-3', thereby participating in the immune and inflammatory responses [46]. *C/EBP* binding sites exist in promoter regions of many inflammation-related cytokines [47, 48]. Hence, *C/EBP* family plays a critical role in the inflammatory response. It has been reported that knockdown of *C/EBP β* in mouse type II alveolar epithelial cells downregulates IL-1 β -induced expression of IL-6 [49]. Stimulation of mouse BMDMs (mouse bone marrow-derived macrophages) by low-dose

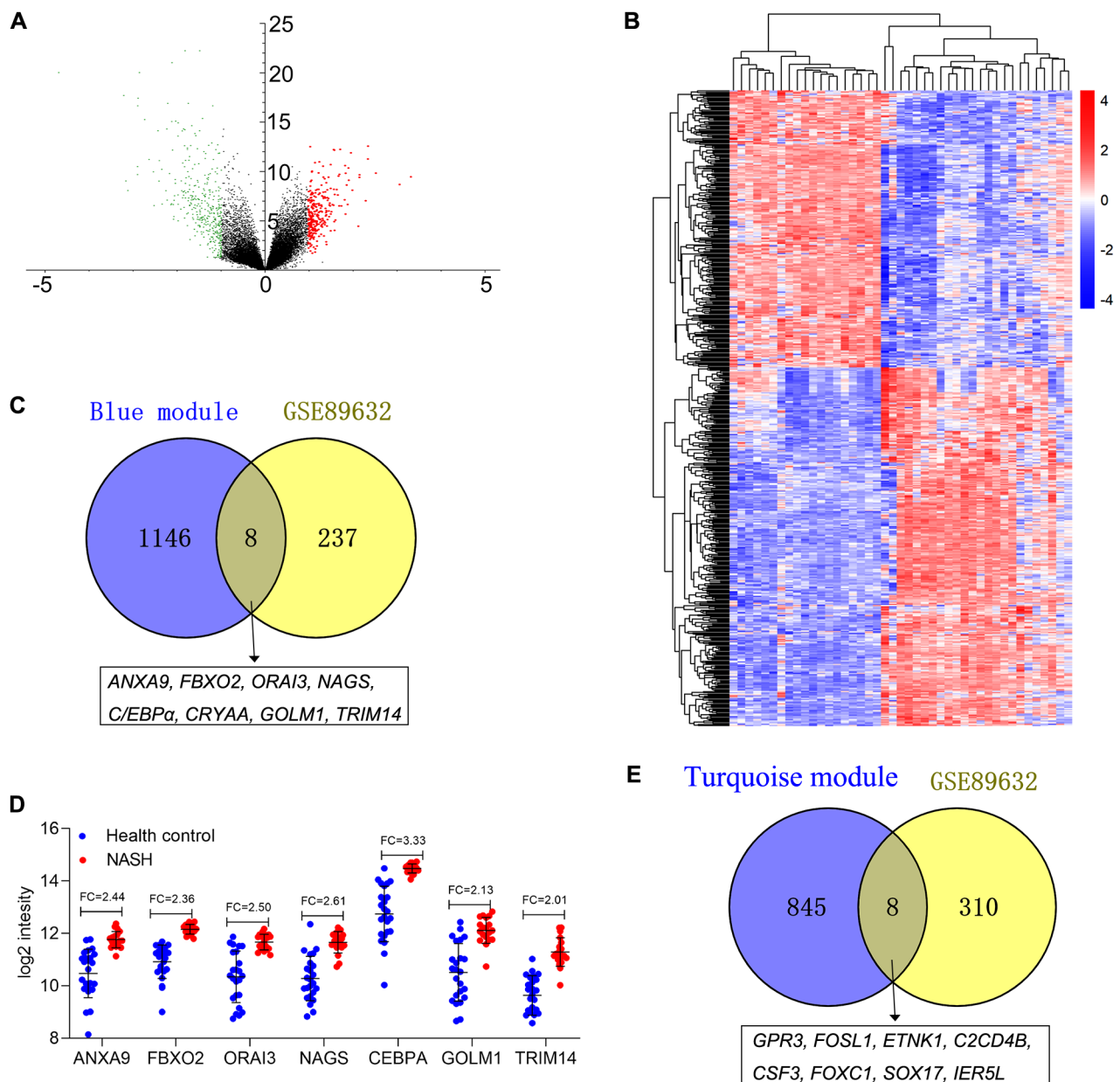


Fig. 5 Identification of driving genes in GSE89632. **A** Volcano plot visualizing DEGs in GSE89632 (19 with nonalcoholic steatohepatitis (NASH) and 24 healthy controls (HC)). The green nodes are downregulated genes, and the red nodes are upregulated genes (|fold change| > 2, p < 0.05). **B** Heatmap hierarchical clustering reveals dysregulated genes in the NASH groups compared with the healthy controls. **C-D** Identification of common genes between upregulated genes and the blue module by overlapping them; C/EBPα was determined to be the most upregulated gene. **E** Identification of common genes between upregulated genes and the turquoise module by overlapping them

LPS enhances the transcriptional activity of C/EBPδ [50]. Knockdown of C/EBPδ alleviates LPS-induced ALI/ARDS symptoms, mainly manifesting as decreased numbers of neutrophils in bronchoalveolar lavage fluid, albumin (reflection of vascular epithelial permeability of lung tissues) and cytokines [51].

It is reasonable to speculate that suppressing the transcriptional activity of C/EBPα may alleviate the inflammatory response. However, through a literature review,

the function of C/EBPα in the development of NAFLD has rarely been reported. Our bioinformatic analysis showed that C/EBPα was upregulated in the liver tissues of NAFLD, which was further confirmed in the NAFLD mouse model. Our result suggest that C/EBPα may aggravate the development of NAFLD.

In Conclusion, we identified two modules and 16 driving genes, including 8 genes positively-correlated and 8 genes negatively-correlated with the severity of NAFLD,

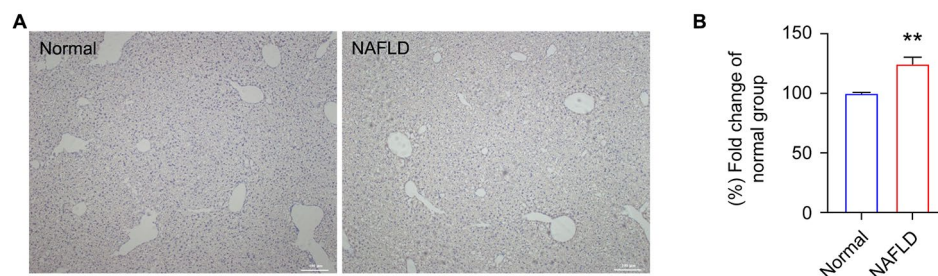


Fig. 6 Validation of C/EBP α in NAFLD mice. **A** Representative images of C/EBP α immunostaining. The left and right panels show the liver tissues of the normal and NAFLD mice, respectively. The brown color represents the positive staining of C/EBP α . **B** Fold changes of protein abundance. The blue and red bars indicate the fold change of the normal group. Standard errors of biological replications (N=6) are indicated. Asterisk indicates the statistical significance ($P < 0.01$)

which will advance the understanding of mechanism of NAFLD. Our findings provide novel directions and therapeutic targets of NAFLD.

Abbreviations

C/EBP α	CCAAT/enhancer binding protein α
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes.
NAFLD	Nonalcoholic fatty liver disease
WGCNA	The weighted gene co-expression network analysis

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09458-3>.

Supplementary Material 1: **Table 1.** Details of the data sets. **Fig. 1.** Correlations between module membership and GS in the blue and turquoise modules.

Acknowledgements

Not applicable.

Authors' contributions

All authors read and approved the final manuscript. Zhengmao Song, Yun Wang, Pingli Lin, Xilin Jiang, Junchen Dong and Shangjin Xie performed research and analyzed data; Lishan Cui, Feng Liu and Xuefeng Huang designed research and wrote the paper; Kaichun Yang and Rong Rao responded to the reviewers.

Funding

This research was funded by the Fujian Province Medical Innovation Project (2018-CXB-30), Fujian Province Science and Technology Plan Project (2019D027, 2023D028, 2023D029, 2023D030), Xiamen Medical and Health Guidance Project (3502Z20209225, 3502Z20209228, 3502Z20214ZD1257, 3502Z20224ZD1282, 3502Z20224ZD1293).

Data Availability

The datasets generated and analyzed in the present study are publicly available in the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) and ArrayExpress (<https://www.ebi.ac.uk/arrayexpress/>).

Declarations

Ethics approval and consent to participate

All animal work in the current study was approved by the Experimental Animal Ethics Committee of Xiamen University, and all methods were performed in accordance with the relevant guidelines and regulations. This

study was reported in accordance with ARRIVE guidelines for the reporting of animal experiment.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹The Fifth Hospital of Xiamen & Xiamen University, Xiamen, China

²Zhongshan Hospital, Xiamen University, Xiamen, China

³School of Medicine, Xiamen University, Xiamen, China

⁴Xiang'an Hospital, Xiamen University, Xiamen, China

Received: 7 September 2022 / Accepted: 16 June 2023

Published online: 24 July 2023

References

- Friedman SL, Neuschwander-Tetri BA, Rinella M, et al. Mechanisms of NAFLD development and therapeutic strategies. *Nat Med*. 2018 Jul;24(7):908–22. <https://doi.org/10.1038/s41591-018-0104-9>. PubMed PMID: 29967350; PubMed Central PMCID: PMC6553468.
- Fierbinteanu-Braticевич C, Sinescu C, Moldoveanu A et al. Nonalcoholic fatty liver disease: one entity, multiple impacts on liver health. *Cell Biol Toxicol*. 2017 Feb;33(1):5–14. doi: 10.1007/s10565-016-9361-x. PubMed PMID: 27680752.
- Younossi ZM, Corey KE, Lim JK. AGA clinical practice update on Lifestyle Modification using Diet and Exercise to achieve weight loss in the management of nonalcoholic fatty liver disease. *Expert Rev Gastroenterol*. 2021 Feb;16(3):912–8. PubMed PMID: 33307021.
- Holmer M, Melum E, Isoniemi H et al. Nonalcoholic fatty liver disease is an increasing indication for liver transplantation in the nordic countries. *Liver Int* 2018 Nov;38(11):2082–90. <https://doi.org/10.1111/liv.13751>. PubMed PMID: 29630771.
- Zezos P, Renner EL. Liver transplantation and non-alcoholic fatty liver disease. *World J Gastroenterol*. 2014 Nov 14;20(42):15532–8. <https://doi.org/10.3748/wjg.v20.i42.15532>. PubMed PMID: 25400437; PubMed Central PMCID: PMC4229518.
- Wang H, Mehal W, Nagy LE, et al. Immunological mechanisms and therapeutic targets of fatty liver diseases. *Cell Mol Immunol*. 2021 Jan;18(1):73–91. <https://doi.org/10.1038/s41423-020-00579-3>. PubMed PMID: 33268887; PubMed Central PMCID: PMC7852578.
- Guerra S, Gastaldelli A. The role of the liver in the modulation of glucose and insulin in non alcoholic fatty liver disease and type 2 diabetes. *Curr Opin Pharmacol*. 2020 Dec;55:165–74. <https://doi.org/10.1016/j.coph.2020.10.016>. PubMed PMID: 33278735.
- Paik J, Golabi P, Younoszai Z, et al. Chronic kidney disease is independently associated with increased mortality in patients with nonalcoholic fatty liver

- disease. *Liver Int.* 2019 Feb;39(2):342–52. <https://doi.org/10.1111/liv.13992>. PubMed PMID: 30347513.
9. Yin X, Guo X, Liu Z et al. Advances in the Diagnosis and Treatment of Non-Alcoholic Fatty Liver Disease. *Int J Mol Sci.* 2023 Feb 2;24(3). <https://doi.org/10.3390/ijms24032844>. PubMed PMID: 36769165; PubMed Central PMCID: PMCPCMC9917647.
 10. Liu F, Goh GB, Tiniakos D, et al. qFIBS: an automated technique for quantitative evaluation of fibrosis, inflammation, ballooning, and steatosis in patients with nonalcoholic steatohepatitis. *Hepatology.* 2020 Jun;71(6):1953–66. doi: 10.1002/hep.30986. PubMed PMID: 31600834.
 11. Kumar R, Priyadarshi RN, Anand U. Non-alcoholic Fatty Liver Disease: Growing Burden, Adverse Outcomes and Associations. *J Clin Transl Hepatol.* 2020 Mar 28;8(1):76–86. doi: 10.14218/JCTH.2019.00051. PubMed PMID: 32274348; PubMed Central PMCID: PMCPCMC7132013.
 12. Pavlopoulos GA, Wegener AL, Schneider R. A survey of visualization tools for biological network analysis. *BioData Min* 2008 Nov 28;1:12. <https://doi.org/10.1186/1756-0381-1-12>. PubMed PMID: 19040716; PubMed Central PMCID: PMCPCMC2636684.
 13. Wren JD. Bioinformatics programs are 31-fold over-represented among the highest impact scientific papers of the past two decades. *Bioinformatics.* 2016 Sep 1;32(17):2686–91. <https://doi.org/10.1093/bioinformatics/btw284>. PubMed PMID: 27153671.
 14. Brazas MD, Ouellette BF. Navigating the changing learning landscape: perspective from bioinformatics.ca. *Brief Bioinform.* 2013 Sep;14(5):556–62. <https://doi.org/10.1093/bib/bbt016>. PubMed PMID: 23515468; PubMed Central PMCID: PMCPCMC3771234.
 15. Ryall KA, Kim J, Klauck PJ, et al. An integrated bioinformatics analysis to dissect kinase dependency in triple negative breast cancer. *BMC Genomics.* 2015;16(Suppl 12):2. <https://doi.org/10.1186/1471-2164-16-S12-S2>. PubMed PMID: 26681397; PubMed Central PMCID: PMCPCMC4682411.
 16. Zhao W, Langfelder P, Fuller T, et al. Weighted gene coexpression network analysis: state of the art. *J Biopharm Stat.* 2010 Mar;20(2):281–300. 10543400903572753. PubMed PMID: 20309759.
 17. Wang P, Zheng H, Zhang J et al. Identification of key gene modules and genes in colorectal cancer by co-expression analysis weighted gene co-expression network analysis. *Biosci Rep.* 2020 Sep 30;40(9). <https://doi.org/10.1042/BSR20202044>. PubMed PMID: 32815531; PubMed Central PMCID: PMCPCMC7463304.
 18. Zhu Y, Li Z, Zhang J et al. Identification of crucial lncRNAs and mRNAs in liver regeneration after portal vein ligation through weighted gene correlation network analysis. *BMC Genomics.* 2022 Sep 21;23(1):665. <https://doi.org/10.1186/s12864-022-08891-0>. PubMed PMID: 36131263; PubMed Central PMCID: PMCPCMC9490934.
 19. Pei G, Chen L, Zhang W. *Methods Enzymol.* 2017;585:135–58. <https://doi.org/10.1016/bs.mie.2016.09.016>. PubMed PMID: 28109426. WGCNA Application to Proteomic and Metabolomic Data Analysis.
 20. Ai D, Wang Y, Li X et al. Colorectal Cancer Prediction Based on Weighted Gene Co-Expression Network Analysis and Variational Auto-Encoder. *Biomolecules.* 2020 Aug 20;10(9). <https://doi.org/10.3390/biom10091207>. PubMed PMID: 32825264; PubMed Central PMCID: PMCPCMC7563725.
 21. Wu R, Zhuang H, Mei YK et al. Systematic identification of key functional modules and genes in esophageal cancer. *Cancer Cell Int.* 2021 Feb 25;21(1):134. <https://doi.org/10.1186/s12935-021-01826-x>. PubMed PMID: 33632229; PubMed Central PMCID: PMCPCMC7905886.
 22. Feng S, Xu Y, Dai Z, et al. Integrative analysis from Multicenter Studies identifies a WGCNA-Derived Cancer-Associated Fibroblast signature for ovarian Cancer. *Front Immunol.* 2022;13:951582. <https://doi.org/10.3389/fimmu.2022.951582>. PubMed PMID: 35874760; PubMed Central PMCID: PMCPCMC9304893.
 23. Jiao M, Wang X, Ji Y et al. Potential key genes are expected to become biomarker for early diagnosis of colorectal cancer through bioinformatics analysis. *Biotechnol Genet Eng Rev* 2023 Mar 7:1–14. doi: 10.1080/02648725.2023.2186586. PubMed PMID: 36880415.
 24. Li JP, Liu LH, Li J et al. Microarray expression profile of long noncoding RNAs in human osteosarcoma. *Biochem Biophys Res Commun.* 2013 Apr 5;433(2):200–6. <https://doi.org/10.1016/j.bbrc.2013.02.083>. PubMed PMID: 23466354.
 25. Wang C, Yang Y, Yin L, et al. Novel potential biomarkers Associated with epithelial to mesenchymal transition and bladder Cancer prognosis identified by Integrated Bioinformatic Analysis. *Front Oncol.* 2020;10:931. <https://doi.org/10.3389/fonc.2020.00931>. PubMed PMID: 32695668; PubMed Central PMCID: PMCPCMC7338771.
 26. Han Z, Ren H, Sun J et al. Integrated weighted gene coexpression network analysis identifies Frizzled 2 (FZD2) as a key gene in invasive malignant pleomorphic adenoma. *J Transl Med.* 2022 Jan 5;20(1):15. <https://doi.org/10.1186/s12967-021-03204-7>. PubMed PMID: 34986855; PubMed Central PMCID: PMCPCMC8734245.
 27. Liang W, Sun F, Zhao Y et al. Identification of Susceptibility Modules and Genes for Cardiovascular Disease in Diabetic Patients Using WGCNA Analysis. *J Diabetes Res.* 2020;2020:4178639. doi: 10.1155/2020/4178639. PubMed PMID: 32455133; PubMed Central PMCID: PMCPCMC7238331 commercial or financial relationships that could be construed as a potential conflict of interest.
 28. Fujiwara N, Kubota N, Crouchet E et al. Molecular signatures of long-term hepatocellular carcinoma risk in nonalcoholic fatty liver disease. *Sci Transl Med.* 2022 Jun 22;14(650):eabo4474. <https://doi.org/10.1126/scitranslmed.abo4474>. PubMed PMID: 35731891; PubMed Central PMCID: PMCPCMC9236162.
 29. Cao Y, Deng Y, Wang J, et al. The association between NAFLD and risk of chronic kidney disease: a cross-sectional study. *Ther Adv Chronic Dis.* 2021;12:20406223211048649. : 10.1177/20406223211048649. PubMed PMID: 34777740; PubMed Central PMCID: PMCPCMC8586173.
 30. Mann JP, Carter P, Armstrong MJ, et al. Hospital admission with non-alcoholic fatty liver disease is associated with increased all-cause mortality independent of cardiovascular risk factors. *PLoS ONE.* 2020;15(10):e0241357. <https://doi.org/10.1371/journal.pone.0241357>. PubMed PMID: 33108366; PubMed Central PMCID: PMCPCMC7591046.
 31. Sanchez-Baizan N, Ribas L, Piferrer F. Improved biomarker discovery through a plot twist in transcriptomic data analysis. *BMC Biol.* 2022 Sep 24;20(1):208. <https://doi.org/10.1186/s12915-022-01398-w>. PubMed PMID: 36153614; PubMed Central PMCID: PMCPCMC9509653.
 32. Niu X, Pan Q, Zhang Q, et al. Weighted correlation network analysis identifies multiple susceptibility loci for low-grade glioma. *Cancer Med.* 2023 Mar;12(5):6379–87. <https://doi.org/10.1002/cam4.5368>. PubMed PMID: 36305248; PubMed Central PMCID: PMCPCMC10028094.
 33. Li Z, Feng J, Yuan Z. Key modules and hub genes identified by coexpression network analysis for revealing Novel biomarkers for Spina Bifida. *Front Genet.* 2020;11:583316. <https://doi.org/10.3389/fgene.2020.583316>. PubMed PMID: 33343629; PubMed Central PMCID: PMCPCMC7738565.
 34. Chen R, Ge T, Jiang W, et al. Identification of biomarkers correlated with hypertrophic cardiomyopathy with co-expression analysis. *J Cell Physiol.* 2019 Dec;234(12):21999–2008. :28762. PubMed PMID: 31059139.
 35. Zhou D, Wu Y, Jiang K et al. Identification of a risk prediction model for clinical prognosis in HER2 positive breast cancer patients. *Genomics* 2021 Oct 16;113(6):4088–97. <https://doi.org/10.1016/j.ygeno.2021.10.010>. PubMed PMID: 34666190.
 36. Lake AD, Novak P, Fisher CD, et al. Analysis of global and absorption, distribution, metabolism, and elimination gene expression in the progressive stages of human nonalcoholic fatty liver disease. *Drug Metab Dispos.* 2011 Oct;39(10):1954–60. <https://doi.org/10.1124/dmd.111.040592>. PubMed PMID: 21737566; PubMed Central PMCID: PMCPCMC3186211.
 37. Arendt BM, Comelli EM, Ma DW, et al. Altered hepatic gene expression in nonalcoholic fatty liver disease is associated with lower hepatic n-3 and n-6 polyunsaturated fatty acids. *Hepatology.* 2015 May;61(5):1565–78. <https://doi.org/10.1002/hep.27695>. PubMed PMID: 25581263.
 38. Wei S, Chen J, Huang Y, et al. Identification of hub genes and construction of transcriptional regulatory network for the progression of colon adenocarcinoma hub genes and TF regulatory network of colon adenocarcinoma. *J Cell Physiol.* 2020 Mar;235(3):2037–48. PubMed PMID: 31612481; PubMed Central PMCID: PMCPCMC6916361.
 39. Kaaij LT, van de Wetering M, Fang F et al. DNA methylation dynamics during intestinal stem cell differentiation reveals enhancers driving gene expression in the villus. *Genome Biol.* 2013 May 28;14(5):R50. <https://doi.org/10.1186/gb-2013-14-5-r50>. PubMed PMID: 23714178; PubMed Central PMCID: PMCPCMC4053812.
 40. Bull JJ, Muller-Rover S, Chronnell CM et al. Contrasting expression patterns of CCAAT/enhancer-binding protein transcription factors in the hair follicle and at different stages of the hair growth cycle. *J Invest Dermatol.* 2002 Jan;118(1):17–24. <https://doi.org/10.1046/j.0022-202x.2001.01629.x>. PubMed PMID: 11851871.
 41. Flodby P, Barlow C, Kylefjord H et al. Increased hepatic cell proliferation and lung abnormalities in mice deficient in CCAAT/enhancer binding protein alpha. *J Biol Chem.* 1996 Oct 4;271(40):24753–60. <https://doi.org/10.1074/jbc.271.40.24753>. PubMed PMID: 8798745.

42. Screpanti I, Romani L, Musiani P et al. Lymphoproliferative disorder and imbalanced T-helper response in C/EBP beta-deficient mice. *EMBO J*. 1995 May 1;14(9):1932–41. PubMed PMID: 7744000; PubMed Central PMCID: PMC398292.
43. Tanaka T, Yoshida N, Kishimoto T et al. Defective adipocyte differentiation in mice lacking the C/EBPbeta and/or C/EBPdelta gene. *EMBO J* 1997 Dec 15;16(24):7432–43. <https://doi.org/10.1093/emboj/16.24.7432>. PubMed PMID: 9405372; PubMed Central PMCID: PMC31170343.
44. Yamanaka R, Barlow C, Lekstrom-Himes J et al. Impaired granulopoiesis, myelodysplasia, and early lethality in CCAAT/enhancer binding protein epsilon-deficient mice. *Proc Natl Acad Sci U S A*. 1997 Nov 25;94(24):13187–92. <https://doi.org/10.1073/pnas.94.24.13187>. PubMed PMID: 9371821; PubMed Central PMCID: PMC324284.
45. Shimokawa T, Nunomura S, Fujisawa D, et al. Identification of the C/EBPalpha C-terminal tail residues involved in the protein interaction with GABP and their potency in myeloid differentiation of K562 cells. *Biochim Biophys Acta*. 2013 Nov;1829(11):1207–17. PubMed PMID: 24076158.
46. Nerlov C. The C/EBP family of transcription factors: a paradigm for interaction between gene expression and proliferation control. *Trends Cell Biol* 2007 Jul;17(7):318–24. <https://doi.org/10.1016/j.tcb.2007.07.004>. PubMed PMID: 17658261.
47. Nerlich A, Ruangkiattikul N, Laarmann K, et al. C/EBPbeta is a transcriptional key regulator of IL-36alpha in murine macrophages. *Biochim Biophys Acta*. 2015 Aug;1849(8):966–78. .002. PubMed PMID: 26066982.
48. Robinson KF, Narasipura SD, Wallace J et al. beta-catenin and TCFs/LEF signaling discordantly regulate IL-6 expression in astrocytes. *Cell Commun Signal* 2020 Jun 16;18(1):93. <https://doi.org/10.1186/s12964-020-00565-2>. PubMed PMID: 32546183; PubMed Central PMCID: PMC37296971.
49. Yan C, Wang X, Cao J, et al. CCAAT/enhancer-binding protein gamma is a critical regulator of IL-1beta-induced IL-6 production in alveolar epithelial cells. *PLoS ONE*. 2012;7(4):e35492. <https://doi.org/10.1371/journal.pone.0035492>. PubMed PMID: 22558159; PubMed Central PMCID: PMC3338717.
50. Maitra U, Gan L, Chang S et al. Low-dose endotoxin induces inflammation by selectively removing nuclear receptors and activating CCAAT/enhancer-binding protein delta. *J Immunol*. 2011 Apr 1;186(7):4467–73. <https://doi.org/10.4049/jimmunol.1003300>. PubMed PMID: 21357541.
51. Shu S, Xu Y, Xie L et al. The role of C/EBPbeta phosphorylation in modulating membrane phospholipids repairing in LPS-induced human lung/bronchial epithelial cells. *Gene* 2017 Sep 20;629:76–85. <https://doi.org/10.1016/j.gene.2017.07.076>. PubMed PMID: 28760550; PubMed Central PMCID: PMC37125708.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.