Identification of a hub gene in hepatocellular carcinoma induced by nonalcoholic fatty liver disease using bioinformatics analysis

Juan Li
Jinan University
Chutian Wu
Jinan University
Xiongxiu Liu
Jinan University
Linjing Long
Guangzhou Medical University
Lixian Zhong
Jinan University
Junlong Lai
Jinan University
Junqin Lin
Jinan University
Yuting Li
Jinan University
Qiuting Zeng
Jinan University
Shaohui Tang (tangshaohui206@163.com)
Jinan University

Research Article

Keywords: Hepatocellular carcinoma, Nonalcoholic fatty liver disease, Bioinformatics analysis, Lipid metabolism, Prognosis

Posted Date: April 7th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1523897/v1

Abstract

Background: Hepatocellular carcinoma (HCC) is the fourth most common cause of cancer mortality worldwide and nonalcoholic fatty liver disease (NAFLD)-related HCC is related to poorer survival than viral hepatitis-related HCC. However, the pathogenetic process of NAFLD to HCC is still poorly understood.

Material and methods: A series of bioinformatics analyses were performed in two public datasets (GSE164760, GSE37031), The Cancer Genome Atlas (TCGA) database, and the International Cancer Genome Consortium (ICGC) database to explore potential hub genes in NAFLD-related HCC.

Results: Weighted Gene Co-Expression Network Analysis (WGCNA) and LASSO logistic regression were performed to filter out 8 hub genes, then *TMEM126A* was selected for further analysis for its better diagnostic, stratification, and prognostic value in HCC. The results of Gene Set Enrichment Analysis (GSEA) indicated that *TMEM126A* was highly enriched in lipid metabolism. Next, *TMEM126A* was significantly correlated with the fatty acid biosynthesis and oxidation pathways. Further validation in other datasets confirmed our results.

Conclusion: *TMEM126A* is firstly screened out for a hub gene in the pathogenesis of NAFLD-related HCC. Although further experiments are supposed to verify these results, we provide useful and novel information to explore the underlying mechanism in NAFLD-related HCC.

Introduction

Hepatocellular carcinoma (HCC), accounting for 90% of all primary hepatoma, is one of the most common types of cancer and is estimated to be the fourth most common cause of cancer mortality worldwide[28,15]. More than 90% of patients with HCC have preexisting chronic liver disease caused most commonly by chronic hepatitis B virus (HBV) infection and chronic hepatitis C virus (HCV) infection[23]. However, due to the global epidemics of obesity and diabetes, nonalcoholic fatty liver disease (NAFLD)-related HCC is now proliferating, with a prevalence ranging from 0.44 per 1,000 person-years among NALFD patients and 9–26 per 1,000 person-years among nonalcoholic steatohepatitis (NASH) cirrhosis patients, and is related to poorer survival than viral hepatitis-related HCC[2,10,29,27]. Although surgical resection has the highest recovery rate, only 15% of patients are qualified, and the 5-year recurrence rate is about 70%[3]. NAFLD-related HCC is typically characterized by activation of fatty acid (FA) synthesis and suppression of fatty acid oxidation (FAO), which is beneficial for HCC development[22]. Recently, Steatohepatitic HCC (SH-HCC), characterized by tumor cells with steatosis, pericellular fibrosis, and inflammatory infiltrates, has been recognized as a histological variant of HCC [5]. These indicate that lipid metabolism may become a new therapeutic target for HCC. However, the pathogenetic process of NAFLD to HCC is still poorly understood, even if NAFLD and HCC partially share a common pathogenetic process.

Several genes from genome-wide association studies have been identified as having impacts on the pathogenesis of NAFLD-HCC, essentially intensifying the level of fat and the severity of NAFLD. It is

reported that *PNPLA3*, *GCKR*, *MBOAT7*, *TM6SF2*, *STAT4*, *KIF1B,* and *HSD17B13* are important genetic modifiers played important roles in the pathogenesis and progression of HCC-NAFLD[7,18]. Recent studies from bioinformatics analysis show that *CYP7A1*, *GINS2*, *PDLIM3*, *HMMR*, *SPP1*, and *AKR1B10*, were associated with the pathogenetic process in NAFLD-HCC[4,26,31]. Although many studies have been spared efforts to explore the pathogenesis and progression of NAFLD-HCC, there are still no effective drugs for NAFLD and NAFLD-related HCC. In addition, NAFLD-related HCC may be less responsive to immunotherapy, and anti-programmed death-1 (PD1) treatment may shorten the survival rate of patients in NAFLD-related HCC[20], which indicates that NAFLD-related HCC is different from hepatitis-related HCC. Therefore, it is necessary to further explore new potential targets, especially in the lipid metabolic process, for diagnosis, stratifications of patients, and therapy.

Transmembrane protein 126A (*TMEM126A*), a mitochondrial inner membrane protein enriched in the cristae, is encoded by a gene located on chromosome 11[8,9]. It is reported that defects and mutations in *TMEM126A* cause autosomal recessive optic atrophy[13,11]. Bae et al. [1] found that *TMEM126A* was a binding partner of CD137 ligand (CD137L), inducing reverse signaling in macrophages and knockdown of *TMEM126A* abolished CD137L-induced tyrosine phosphorylation and cell adherence. The loss of *TMEM126A* induces mitochondrial dysfunction and subsequently metastasis by activating extracellular matrix remodeling and epithelial-to-mesenchymal transition (EMT) in breast cancer[24]. However, its role in the pathogenesis of NAFLD-related HCC has not been reported.

Thus, one public dataset about NAFLD-HCC was analyzed to identify differentially expressed genes (DEGs) among healthy control (HC), NASH, and NASH-HCC group. Then, Weighted Gene Co-Expression Network Analysis (WGCNA) and LASSO logistic regression were performed to filter out trait-survival-related genes. Subsequently, the relationship between *TMEM126A* and clinicopathologic features was further explored. Then, Gene Set Enrichment Analysis (GSEA) was conducted to explore the potential underlying mechanism of *TMEM126A*, and the correlation analysis of *TMEM126A* and genes related to the FA biosynthesis and oxidation pathways was also performed. This study aimed to screen out potential therapeutic genes of NAFLD-related HCC.

Materials And Methods

Data Retrieving and Processing

The gene expression profile of GSE164760 [21] and GSE37031 [17] were downloaded from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo). 6 healthy control (HC) samples, 8 cirrhotic liver samples, 74 NASH samples, and 53 NASH-HCC samples in the GSE164760, and 7 HC samples and 8 NASH samples were included in this study. HCC data were downloaded from The Cancer Genome Atlas (TCGA) database, the Genotype-Tissue Expression (GTEx) database, and the International Cancer Genome Consortium (ICGC) database, including 374 HCC samples, 50 normal adjacent samples, and 110

normal samples in TCGA and GTEx database (**Table 1**), and 212 HCC samples and 177 normal adjacent samples in ICGC database.

Subsequently, the DEGs were identified using two R packages (GEOquery and limma). The threshold for the DEGs was set as *p*-value <0.05 and $|\log_2$ fold change (FC) $| \geq 0.263$. Next, the Venn diagram was constructed using "ggplot2" R packages, and the overlaps represented the genes significantly altered in both NASH and HCC. **Figure 1** illustrated the overall research design.

Diagnostic Methods of Different States

All the samples in GSE164760 and GSE37031 were validated using histological examination by at least two board-certified pathologists before molecular analysis, and hematoxylin and eosin (H&E) staining were used for histological analysis. NASH samples were diagnosed using criteria from NAFLD Activity Score (NAS). Samples with a METAVIR score of F4 were considered cirrhotic.

Weighted Gene Co-Expression Network Analysis (WGCNA)

WGCNA was performed using "WGCNA"[14] R package and carried out on all genes using GSE164760 dataset. The scale-free topology of the networks was assessed for various values of the β shrinkage parameter, and we chose β = 6 based on scale-free topology criterion. Finally, the dynamic tree cut algorithm was applied to the dendrogram for module identification with the mini-size of module gene numbers set as 50 and similar modules were merged following a height cutoff of 0.25. In the module-trait analysis, gene-trait significance (GS) value > 0.3 and module membership (MM) value > 0.55 were defined as trait-related genes[26]. Subsequently, the intersection of DEGs, trait-related genes, and survival-related genes in HCC was selected for further analysis.

The LASSO Logistic Regression

Logistic regression was performed with Lasso regularization using "glmnet" R package. Regularization parameters were determined by 10-fold cross-validation analysis. All Lasso models were run with the same seed (2021). Subsequently, the risk score plot was generated from "ggplot2" R package.

Gene Set Enrichment Analysis (GSEA)

The median of transmembrane protein 126A (*TMEM126A*) expression was selected as the cut-off value to performed gene set enrichment analysis (GSEA). Gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) analysis were performed to identify potential biological processes and pathways in high-risk groups using "cluster Profiler"[30] and "org.Hs.eg.db" R package. Gene sets with |Normalized enrichment score (NES)| > 1, adjusted-*p* < 0.05 and false discovery rate (FDR) < 0.05 were considered as statistical significance.

Statistical Analysis

Statistical analysis was performed using R software (Version 4.1.0).

Statistical comparisons between groups of continuous parametric variables were performed using the t-test or Wilcoxon rank sum test according to the test condition, and categorical variables were evaluated using Chi-square test or Fisher's exact test. A receiver operating characteristic (ROC) curve was performed to assess the diagnostic value of candidate genes in normal samples of GTEx combined adjacent HCC tissues and HCC samples. The expression of *TMEM126A* in NASH and HCC patients was assessed and visualized in box plots. Subsequently, the association between *TMEM126A* expression and clinical features in HCC were analyzed. The survival analysis was collected from Gene Expression Profiling Interactive Analysis (GEPIA, http://gepia.cancer-pku.cn/) or performed using "survival" R package. Considering the large sample size in TCGA database and missing data were at random, data with missing values were removed, and univariate and multivariate Cox analyses of TCGA dataset were utilized to explore potential prognostic factors. Then, a nomogram was constructed to predict 1-, 3- and 5-year OS for patients with HCC. The immunohistochemical pictures were collected from the Human Protein Atlas (HPA, https://www.proteinatlas.org/). The correlation between *TMEM126A* and genes related to FAO was performed using the Spearman correlation. A difference with $p < 0.05$ was considered significant.

Results

Identification of DEGs in NAFLD-related HCC

The DEGs among HC, NASH, and NAFLD-related HCC in GSE164760 dataset were identified, respectively (Figure 2A-C). Next, the intersection among the three groups represented 1334 DEGs associated with disease progression (Figure 2D). These results illustrated that the cross DEGs had effects on the pathogenesis from NASH to NAFLD-related HCC.

Identification of Trait-related Genes by WGCNA

WGCNA was performed to identify key modules related to clinical traits in GSE164760 dataset. (**Figure 3A**). The power of β = 6 (scale-free R^2=0.87) was selected as the soft thresholding parameter to construct a scale-free network (**Figure 3B**). Similar module clustering was constructed by using dynamic hybrid cutting (threshold = 0.25, **Figure 3C, Supplementary Table S1-2**). A total of 19 modules were identified (**Figure 2D**). The results in **Figure 3D** showed that the red module was the highest correlation module to NASH (NAS, R^2 = 0.67, $p < 1e^{-200}$) and HCC (R^2 = 0.60, $p = 3.1e^{-116}$). **Figure 3E-F** showed gene significance for NASH and HCC in red modules. Then, 116 trait-related genes obtained from WGCNA, 1334 DEGs from GSE164760 dataset, and survival-related genes from TCGA database were intersected, and finally 42 trait-survival-related genes were selected for further analysis. The results indicated that these genes were not only highly associated with NAFLD-related HCC but also with prognosis.

Construction of LASSO Logistic Regression and Diagnostic Value of *TMEM126A*

To further filtrate key genes, LASSO logistic regression was performed based on 42 trait-survival-related genes in TCGA database, and 8 hub genes were screened out for further analysis (**Figure 4A, Supplementary Table S3**). **Figure 4B** showed the risk score of selected genes(**Supplementary Table S4**). Subsequently, we chose genes with the top 3 risk score to construct the ROC curve, and *TMEM126A* had the highest diagnostic value compared with *SRRT* and *GPR180*. The area under curve (AUC) of *TMEM126A* was 0.866 showing favorable practical functions for the diagnosis (**Figure 4C**). Considering *TMEM126A* had the most diagnosis value and the highest risk score, it was chosen for further exploration. **Figure 4D-E** demonstrated that high *TMEM126A* expression was significantly correlated with overall survival (OS, $p = 0.002$) rate and disease-free survival (DFS, $p = 0.016$) rate in the GEPIA database.

High *TMEM126A* Expression is Correlate with Clinicopathologic Features and poor prognosis in patients with HCC

Next, further exploration was performed to explore the relationship between *TMEM126A* expression and HCC in TCGA and GTEx databases. The median of *TMEM126A* expression was selected as the cut-off value to divide the patients into two groups, and high *TMEM126A* was associated with T stage ($p=0.024$) and histologic grade ($p < 0.001$) (**Table 1)**. Surprisingly, *TMEM126A* expression was significantly higher in HCC tissues than in adjacent HCC samples ($p < 0.001$) by using TCGA database (**Figure 4F**). Meanwhile, the different expression of *TMEM126A* in normal samples of GTEx combined adjacent HCC tissues and HCC samples were analyzed and found that *TMEM126A* was high-expressed in HCC ($p < 0.001$) (**Figure 4G**). Then, among 50 HCC samples and matched adjacent samples, *TMEM126A* expression was significantly increased in tumor samples ($p<0.001$) (**Figure 4H**). In the HPA database, immunofluorescence staining of the subcellular distribution showed that *TMEM126A* was located in the cytosol (green) and nucleoplasm (blue), and the expression of *TMEM126A* was also abnormally elevated in HCC (**Figure 4I-J**).

High *TMEM126A* expression was significantly associated with pathologic stage, histologic grade, vascular invasion, and T stage, while it was not associated with other features (**Figure 5A-C and Figure S1**). Subsequently, univariate Cox analyses showed that high *TMEM126A* expression was significantly correlated with poor OS (Hazard ratio [HR] =2.154, 95% CI= 1.494-3.105, $p < 0.001$). Further estimation using multivariate Cox analysis using 271 complete data demonstrated that high *TMEM126A* expression might be an independent risk factor correlated with poor OS (HR =2.222, 95% CI= 1.387-3.559, $p < 0.001$, **Table 2**). Then, the nomogram used age, T, M, N classification, pathologic stage, histologic grade, vascular invasion, and *TMEM126A* to predict the 1, 3, 5-year OS in the TCGA dataset (**Figure 5D**).

TMEM126A-Related signaling Pathways Based on GSEA

GSEA was performed to identify the signaling pathways activated in HCC by comparing data sets that had low and high expression of *TMEM126A.* **Figure 5E** showed the results of GO(**Supplementary Table S5**). The results indicated that cell cycle, complement and coagulation cascades, and DNA replication were enriched in *TMEM126A* high expression phenotype, and primary bile acid biosynthesis, oxidative

phosphorylation, and fatty acid degradation were enriched in *TMEM126A* low expression phenotype (**Figure 5F, Supplementary Table S6**).

Activation of the FA synthesis and suppression of the FAO are typically featured in NAFLD-related HCC[22]. Considering that *TMEM126A* was highly associated with lipid metabolism, the correlation analysis was performed between *TMEM126A* and genes related to the FA biosynthesis and oxidation pathways (**Figure 6A**). In the FA biosynthesis pathway, *TMEM126A* was significantly correlated with triosephosphate isomerase (*TPI1*, p <0.001), acetyl-CoA carboxylase 1 (*ACACA*, p <0.001), NADH: ubiquinone oxidoreductase subunit AB1 (*NDUFAB1*, p <0.001)**, malonyl-CoA-acyl carrier protein transacylase (*MCAT*,** p <0.001**),** and ATP-citrate lyase (*ACLY*, p <0.001). In the FAO pathway, *TMEM126A* was negatively correlated with enoyl-CoA hydratase domain-containing protein 2 (*ECHDC2*, p <0.001) and enoyl-CoA hydratase and 3-hydroxyacyl CoA dehydrogenase (*EHHADH*, p <0.001), short-chain specific acyl-CoA dehydrogenase (*ACADS*, p <0.001), long-chain-fatty-acid–CoA ligase 1 (*ACSL1*, p <0.001), and long-chain specific acyl-CoA dehydrogenase (ACADL, p <0.001). These results indicated that TMEM126A might be involved in lipid metabolism reprogramming.

Further Validation of TMEM126A in Diagnosis and Prognosis

Subsequently, further validation was performed in GSE37031 and ICGC datasets. Consistent with our results, TMEM126A expression was abnormally up-regulated in NASH and HCC patients (**Figure 6B-C**). In addition, high *TMEM126A* expression was also significantly correlated with OS (p = 0.0017, **Figure 6D**). These outcomes suggested that TMEM126A had superior diagnostic and prognostic value.

Discussion

The morbidity of NAFLD-related HCC is predicted to increase dramatically by 2030, with increases of 82, 117, and 122% from 2016 in China, France, and the USA, respectively[10]. However, the pathogenesis of NAFLD-related HCC is still unclear. In the present study, WGCNA and LASSO logistic regression were performed to filter out 8 hub genes, then *TMEM126A* was selected for further analysis for its better diagnostic, stratification, and prognostic value in HCC. The results of GSEA indicated that *TMEM126A* was highly enriched in lipid metabolism. Subsequently, *TMEM126A* was positively correlated with FA biosynthesis and negatively associated with FAO pathway, which was consistent with the features of NAFLD-related HCC.

Cancer research has recently focused on the role of metabolic reprogramming in tumors, particularly lipid metabolism. Metabolic reprogramming helps cancer cells adapt to their local environment and is recognized as an important characteristic of tumor cells during tumorigenesis and metastasis[25]. Cancer cells often increase reliance on FA biosynthesis and generate energy through FAO, but excessive FAO in high intracellular fat levels results in metabolic stress and eventually lipotoxic cell death[6,12]. In patients with NASH, FAO is required to eliminate lipotoxic hepatocytes. The lipotoxic effect plays an indispensable role in the pathogenesis of NAFLD-related HCC, and cancer cells must adapt to a lipid-rich environment to survive[19]. Therefore, suppression of FAO in NASH contributes to the accumulation of FA and thereby

enhances the lipotoxic effect and avoids excessive metabolic stress as well as the clearance of lipotoxic hepatocytes. Once NASH progresses to HCC, cancer cells need appropriate FAO to generate energy. Meanwhile, mitochondria are the center of metabolic and biosynthetic activity and are associated with oxidative phosphorylation and lipid metabolism[16]. In the present study, *TMEM126A* expression was significantly increased in NASH and NAFLD-related HCC. However, it was expressively down-expressed in NAFLD-related HCC tissues compared with NASH tissues, which could be explained by lipid reprogramming. Therefore, we speculate that *TMEM126A* may be involved in lipid reprogramming and promote carcinogenesis in NAFLD-related HCC via altering mitochondrial morphology and function.

To the best of our knowledge, *TMEM126A* is firstly screened out for a new hub gene in the pathogenesis of NAFLD-related HCC. Although further experiments are supposed to verify these results, we provide useful and novel information to explore the underlying mechanism in NAFLD-related HCC.

Declarations

Funding

This research was supported by Livelihood Science and Technology project of Liaoning Province, China (No.2021JH2/10300046).

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

Juan Li, Chutian Wu, Xiongxiu Liu, and Linjing Long contributed equally to this paper. Juan Li, Chutian Wu, Xiongxiu Liu, and Linjing Long analyzed the study data, helped draft the manuscript, made critical revisions of the manuscript. Lixian Zhong, Junlong Lai, Junqin Lin, Yuting Li, and Qiuting Zeng assisted with data collection and the analysis. Shaohui Tang supervised the research and edited the manuscript. All authors contributed to the article and approved the submitted version.

Data Availability

Publicly available datasets were analyzed in this study. This data

can be found here: GEO data base, accession number: GSE164760 and GSE37031.

Ethics approval

Not applicable

Consent to participate

Not applicable

Consent to publish

Not applicable

References

1. Bae JS, Choi JK, Moon JH, Kim EC, Croft M, Lee HW(2012)Novel transmembrane protein 126A (TMEM126A) couples with CD137L reverse signals in myeloid cells.CELL SIGNAL 24 (12):2227-2236. https://doi.org/10.1016/j.cellsig.2012.07.021
2. Desai A, Sandhu S, Lai JP, Sandhu DS(2019)Hepatocellular carcinoma in non-cirrhotic liver: A comprehensive review.World J Hepatol 11 (1):1-18. https://doi.org/10.4254/wjh.v11.i1.1
3. El-Serag HB(2011)Hepatocellular carcinoma.N Engl J Med 365 (12):1118-1127. https://doi.org/10.1056/NEJMra1001683
4. Feng G, Li XP, Niu CY, Liu ML, Yan QQ, Fan LP, Li Y, Zhang KL, Gao J, Qian MR, He N, Mi M(2020)Bioinformatics analysis reveals novel core genes associated with nonalcoholic fatty liver disease and nonalcoholic steatohepatitis.GENE 742:144549. https://doi.org/10.1016/j.gene.2020.144549
5. Fujiwara N, Nakagawa H, Enooku K, Kudo Y, Hayata Y, Nakatsuka T, Tanaka Y, Tateishi R, Hikiba Y, Misumi K, Tanaka M, Hayashi A, Shibahara J, Fukayama M, Arita J, Hasegawa K, Hirschfield H, Hoshida Y, Hirata Y, Otsuka M, Tateishi K, Koike K(2018)CPT2 downregulation adapts HCC to lipid-rich environment and promotes carcinogenesis via acylcarnitine accumulation in obesity.GUT 67 (8):1493-1504. https://doi.org/10.1136/gutjnl-2017-315193
6. Fujiwara N, Nakagawa H, Enooku K, Kudo Y, Hayata Y, Nakatsuka T, Tanaka Y, Tateishi R, Hikiba Y, Misumi K, Tanaka M, Hayashi A, Shibahara J, Fukayama M, Arita J, Hasegawa K, Hirschfield H, Hoshida Y, Hirata Y, Otsuka M, Tateishi K, Koike K(2018)CPT2 downregulation adapts HCC to lipid-rich environment and promotes carcinogenesis via acylcarnitine accumulation in obesity.GUT 67 (8):1493-1504. https://doi.org/10.1136/gutjnl-2017-315193
7. Geh D, Anstee QM, Reeves HL(2021)NAFLD-Associated HCC: Progress and Opportunities.J Hepatocell Carcinoma 8:223-239. https://doi.org/10.2147/JHC.S272213
8. Hanein S, Garcia M, Fares-Taie L, Serre V, De Keyzer Y, Delaveau T, Perrault I, Delphin N, Gerber S, Schmitt A, Masse JM, Munnich A, Kaplan J, Devaux F, Rozet JM(2013)TMEM126A is a mitochondrial located mRNA (MLR) protein of the mitochondrial inner membrane.Biochim Biophys Acta 1830 (6):3719-3733. https://doi.org/10.1016/j.bbagen.2013.02.025
9. Hanein S, Perrault I, Roche O, Gerber S, Khadom N, Rio M, Boddaert N, Jean-Pierre M, Brahimi N, Serre V, Chretien D, Delphin N, Fares-Taie L, Lachheb S, Rotig A, Meire F, Munnich A, Dufier JL, Kaplan J, Rozet JM(2009)TMEM126A, encoding a mitochondrial protein, is mutated in autosomal-recessive nonsyndromic optic atrophy.AM J HUM GENET 84 (4):493-498. https://doi.org/10.1016/j.ajhg.2009.03.003

10. Huang DQ, El-Serag HB, Loomba R(2021)Global epidemiology of NAFLD-related HCC: trends, predictions, risk factors and prevention.Nat Rev Gastroenterol Hepatol 18 (4):223-238. https://doi.org/10.1038/s41575-020-00381-6
11. Kloth K, Synofzik M, Kernstock C, Schimpf-Linzenbold S, Schuettauf F, Neu A, Wissinger B, Weisschuh N(2019)Novel likely pathogenic variants in TMEM126A identified in non-syndromic autosomal recessive optic atrophy: two case reports.BMC MED GENET 20 (1):62. https://doi.org/10.1186/s12881-019-0795-x
12. Koundouros N, Poulogiannis G(2020)Reprogramming of fatty acid metabolism in cancer.Br J Cancer 122 (1):4-22. https://doi.org/10.1038/s41416-019-0650-z
13. La Morgia C, Caporali L, Tagliavini F, Palombo F, Carbonelli M, Liguori R, Barboni P, Carelli V(2019)First TMEM126A missense mutation in an Italian proband with optic atrophy and deafness.Neurol Genet 5 (3):e329. https://doi.org/10.1212/NXG.0000000000000329
14. Lazar C, Gatto L, Ferro M, Bruley C, Burger T(2016)Accounting for the Multiple Natures of Missing Values in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies.J PROTEOME RES 15 (4):1116-1125. https://doi.org/10.1021/acs.jproteome.5b00981
15. Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, Lencioni R, Koike K, Zucman-Rossi J, Finn RS(2021)Hepatocellular carcinoma.NAT REV DIS PRIMERS 7 (1):6. https://doi.org/10.1038/s41572-020-00240-3
16. Longo M, Paolini E, Meroni M, Dongiovanni P(2021)Remodeling of Mitochondrial Plasticity: The Key Switch from NAFLD/NASH to HCC.INT J MOL SCI 22 (8). https://doi.org/10.3390/ijms22084173
17. Lopez-Vicario C, Gonzalez-Periz A, Rius B, Moran-Salvador E, Garcia-Alonso V, Lozano JJ, Bataller R, Cofan M, Kang JX, Arroyo V, Claria J, Titos E(2014)Molecular interplay between Delta5/Delta6 desaturases and long-chain fatty acids in the pathogenesis of non-alcoholic steatohepatitis.GUT 63 (2):344-355. https://doi.org/10.1136/gutjnl-2012-303179
18. McGlynn KA, Petrick JL, El-Serag HB(2021)Epidemiology of Hepatocellular Carcinoma.HEPATOLOGY 73 Suppl 1:4-13. https://doi.org/10.1002/hep.31288
19. Olofson AM, Gonzalo DH, Chang M, Liu X(2018)Steatohepatitic Variant of Hepatocellular Carcinoma: A Focused Review.Gastroenterology Res 11 (6):391-396. https://doi.org/10.14740/gr1110
20. Pfister D, Nunez NG, Pinyol R, Govaere O, Pinter M, Szydlowska M, Gupta R, Qiu M, Deczkowska A, Weiner A, Muller F, Sinha A, Friebel E, Engleitner T, Lenggenhager D, Moncsek A, Heide D, Stirm K, Kosla J, Kotsiliti E, Leone V, Dudek M, Yousuf S, Inverso D, Singh I, Teijeiro A, Castet F, Montironi C, Haber PK, Tiniakos D, Bedossa P, Cockell S, Younes R, Vacca M, Marra F, Schattenberg JM, Allison M, Bugianesi E, Ratziu V, Pressiani T, D'Alessio A, Personeni N, Rimassa L, Daly AK, Scheiner B, Pomej K, Kirstein MM, Vogel A, Peck-Radosavljevic M, Hucke F, Finkelmeier F, Waidmann O, Trojan J, Schulze K, Wege H, Koch S, Weinmann A, Bueter M, Rossler F, Siebenhuner A, De Dosso S, Mallm JP, Umansky V, Jugold M, Luedde T, Schietinger A, Schirmacher P, Emu B, Augustin HG, Billeter A, Muller-Stich B, Kikuchi H, Duda DG, Kutting F, Waldschmidt DT, Ebert MP, Rahbari N, Mei HE, Schulz AR, Ringelhan M, Malek N, Spahn S, Bitzer M, Ruiz DGM, Lujambio A, Dufour JF, Marron TU, Kaseb A, Kudo M, Huang

YH, Djouder N, Wolter K, Zender L, Marche PN, Decaens T, Pinato DJ, Rad R, Mertens JC, Weber A, Unger K, Meissner F, Roth S, Jilkova ZM, Claassen M, Anstee QM, Amit I, Knolle P, Becher B, Llovet JM, Heikenwalder M(2021)NASH limits anti-tumour surveillance in immunotherapy-treated HCC.NATURE 592 (7854):450-456. https://doi.org/10.1038/s41586-021-03362-0

21. Pinyol R, Torrecilla S, Wang H, Montironi C, Pique-Gili M, Torres-Martin M, Wei-Qiang L, Willoughby CE, Ramadori P, Andreu-Oller C, Taik P, Lee YA, Moeini A, Peix J, Faure-Dupuy S, Riedl T, Schuehle S, Oliveira CP, Alves VA, Boffetta P, Lachenmayer A, Roessler S, Minguez B, Schirmacher P, Dufour JF, Thung SN, Reeves HL, Carrilho FJ, Chang C, Uzilov AV, Heikenwalder M, Sanyal A, Friedman SL, Sia D, Llovet JM(2021)Molecular characterisation of hepatocellular carcinoma in patients with non-alcoholic steatohepatitis.J HEPATOL 75 (4):865-878. https://doi.org/10.1016/j.jhep.2021.04.049
22. Sangineto M, Villani R, Cavallone F, Romano A, Loizzi D, Serviddio G(2020)Lipid Metabolism in Development and Progression of Hepatocellular Carcinoma.Cancers (Basel) 12 (6). https://doi.org/10.3390/cancers12061419
23. Siegel R, Ma J, Zou Z, Jemal A(2014)Cancer statistics, 2014.CA Cancer J Clin 64 (1):9-29. https://doi.org/10.3322/caac.21208
24. Sun HF, Yang XL, Zhao Y, Tian Q, Chen MT, Zhao YY, Jin W(2019)Loss of TMEM126A promotes extracellular matrix remodeling, epithelial-to-mesenchymal transition, and breast cancer metastasis by regulating mitochondrial retrograde signaling.CANCER LETT 440-441:189-201. https://doi.org/10.1016/j.canlet.2018.10.018
25. Ward PS, Thompson CB(2012)Metabolic reprogramming: a cancer hallmark even warburg did not anticipate.CANCER CELL 21 (3):297-308. https://doi.org/10.1016/j.ccr.2012.02.014
26. Wu C, Zhou Y, Wang M, Dai G, Liu X, Lai L, Tang S(2021)Bioinformatics Analysis Explores Potential Hub Genes in Nonalcoholic Fatty Liver Disease.FRONT GENET 12. https://doi.org/10.3389/fgene.2021.772487
27. Yang JD, Ahmed F, Mara KC, Addissie BD, Allen AM, Gores GJ, Roberts LR(2020)Diabetes Is Associated With Increased Risk of Hepatocellular Carcinoma in Patients With Cirrhosis From Nonalcoholic Fatty Liver Disease.HEPATOLOGY 71 (3):907-916. https://doi.org/10.1002/hep.30858
28. Yang JD, Hainaut P, Gores GJ, Amadou A, Plymoth A, Roberts LR(2019)A global view of hepatocellular carcinoma: trends, risk, prevention and management.Nat Rev Gastroenterol Hepatol 16 (10):589-604. https://doi.org/10.1038/s41575-019-0186-y
29. Younossi ZM, Otgonsuren M, Henry L, Venkatesan C, Mishra A, Erario M, Hunt S(2015)Association of nonalcoholic fatty liver disease (NAFLD) with hepatocellular carcinoma (HCC) in the United States from 2004 to 2009.HEPATOLOGY 62 (6):1723-1730. https://doi.org/10.1002/hep.28123
30. Yu G, Wang LG, Han Y, He QY(2012)clusterProfiler: an R package for comparing biological themes among gene clusters.OMICS 16 (5):284-287. https://doi.org/10.1089/omi.2011.0118
31. Zhang D, Liu J, Xie T, Jiang Q, Ding L, Zhu J, Ye Q(2020)Oleate acid-stimulated HMMR expression by CEBPalpha is associated with nonalcoholic steatohepatitis and hepatocellular carcinoma.INT J BIOL SCI 16 (15):2812-2827. https://doi.org/10.7150/ijbs.49785

Tables

Table 1 Correlation Between *TMEM126A* Expression and Clinicopathologic Characteristics of Patients with HCC

Characteristic	Low expression of *TMEM126A*	High expression of *TMEM126A*	p
n	187	187	
Year of initial diagnosis (IQR)	2011 (2008, 2012)	2011 (2008, 2013)	
Gender, n (%)			0.825
Female	62 (16.6%)	59 (15.8%)	
Male	125 (33.4%)	128 (34.2%)	
Age, meidan (IQR)	62 (53.5, 69)	60 (51, 68)	0.258
BMI, n (%)			0.098
<=25	77 (22.8%)	100 (29.7%)	
>25	85 (25.2%)	75 (22.3%)	
T stage, n (%)			0.024
T1	104 (28%)	79 (21.3%)	
T2	41 (11.1%)	54 (14.6%)	
T3	37 (10%)	43 (11.6%)	
T4	3 (0.8%)	10 (2.7%)	
N stage, n (%)			0.624
N0	121 (46.9%)	133 (51.6%)	
N1	1 (0.4%)	3 (1.2%)	
M stage, n (%)			0.628
M0	123 (45.2%)	145 (53.3%)	
M1	1 (0.4%)	3 (1.1%)	
Pathologic stage, n (%)			0.056
Stage I	98 (28%)	75 (21.4%)	
Stage II	38 (10.9%)	49 (14%)	
Stage III	35 (10%)	50 (14.3%)	
Stage IV	2 (0.6%)	3 (0.9%)	
Tumor status, n (%)			0.078
Tumor free	110 (31%)	92 (25.9%)	

Characteristic	Low expression of *TMEM126A*	High expression of *TMEM126A*	p
With tumor	68 (19.2%)	85 (23.9%)	
Histologic grade, n (%)			< 0.001
G1	38 (10.3%)	17 (4.6%)	
G2	95 (25.7%)	83 (22.5%)	
G3	49 (13.3%)	75 (20.3%)	
G4	3 (0.8%)	9 (2.4%)	
Child-Pugh grade, n (%)			0.310
A	108 (44.8%)	111 (46.1%)	
B	8 (3.3%)	13 (5.4%)	
C	1 (0.4%)	0 (0%)	
Fibrosis ishak score, n (%)			0.234
0	43 (20%)	32 (14.9%)	
1/2	18 (8.4%)	13 (6%)	
3/4	10 (4.7%)	18 (8.4%)	
5/6	42 (19.5%)	39 (18.1%)	
Vascular invasion, n (%)			0.102
No	116 (36.5%)	92 (28.9%)	
Yes	50 (15.7%)	60 (18.9%)	

IQR, interquartile range.

Table 2 Univariate and multivariate Cox regression analysis in TCGA database

Characteristics	Total (N)	Univariate analysis		Multivariate analysis	
		Hazard ratio (95% CI)	P	Hazard ratio (95% CI)	P
Age (>60 vs ≤60)	196/177	1.205 (0.850-1.708)	0.295		
T stage (T3&T4 vs T1&T2)	93/277	2.598 (1.826-3.697)	**<0.001**	2.597 (1.676-4.024)	**<0.001**
N stage (N1 vs N0)	4/254	2.029 (0.497-8.281)	0.324		
M stage (M1 vs M0)	4/268	4.077 (1.281-12.973)	**0.017**	1.560 (0.473-5.148)	0.465
Pathologic stage (Stage III & Stage IV vs Stage I & Stage II)	90/259	2.504 (1.727-3.631)	**<0.001**		
Histologic grade (G3&G4 vs G2&G1)	135/233	1.091 (0.761-1.564)	0.636		
Vascular invasion (Yes vs No)	109/208	1.344 (0.887-2.035)	0.163		
Fibrosis ishak score (3/4&5/6 vs 0&1/2)	108/106	0.740 (0.445-1.232)	0.247		
Child-Pugh grade (B&C vs A)	22/218	1.643 (0.811-3.330)	0.168		
TMEM126A	373	2.154 (1.494-3.105)	**<0.001**	2.222 (1.387-3.559)	**<0.001**

Figures

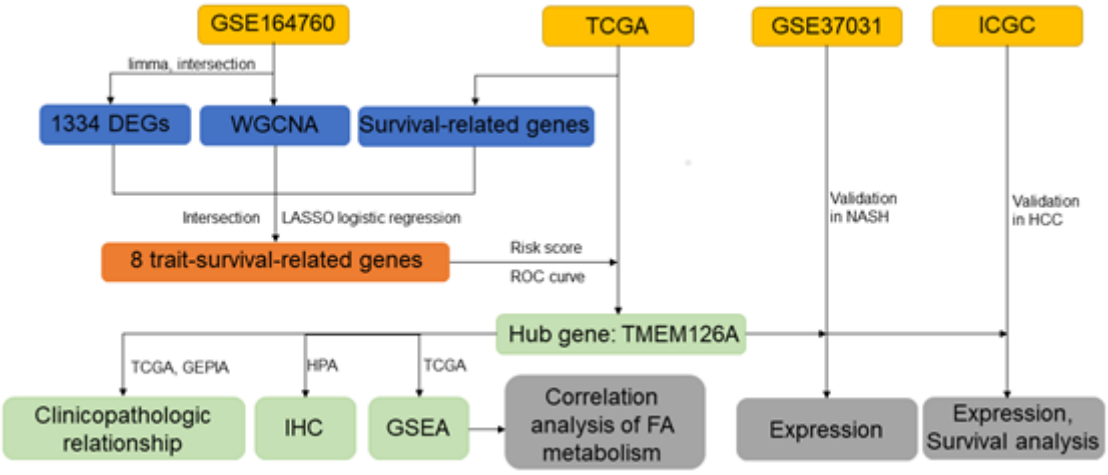

Figure 1

The overall research designs. The data were downloaded from GEO, TCGA, and ICGC databases. GSE164760 dataset was used to identify DEGs and perform WGCNA. Subsequently, the intersection genes of DEGs, WGCNA, and survival-related genes from TCGA database were applied to LASSO logistic regression. Then, *TMEM126A* was selected as the hub gene for further exploration of clinicopathologic features and underlying pathways. Eventually, further validation of *TMEM126A* was performed in GSE37031 and ICGC database.

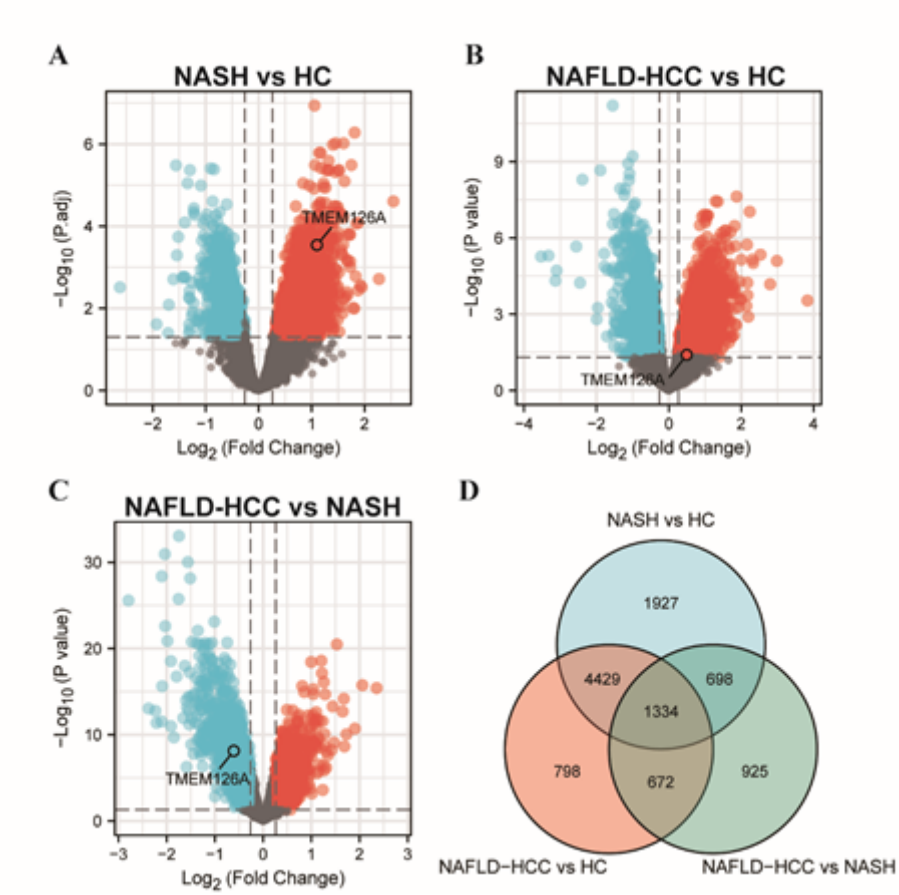

Figure 2

(**A-C**) Identification of differentially expressed genes (DEGs) among HC, NASH, and NAFLD-related HCC; (**D**) Venn diagrams displayed the overlapping DEGs among NASH vs HC, NAFLD-related HCC vs HC, and NAFLD-related HCC vs NASH groups.

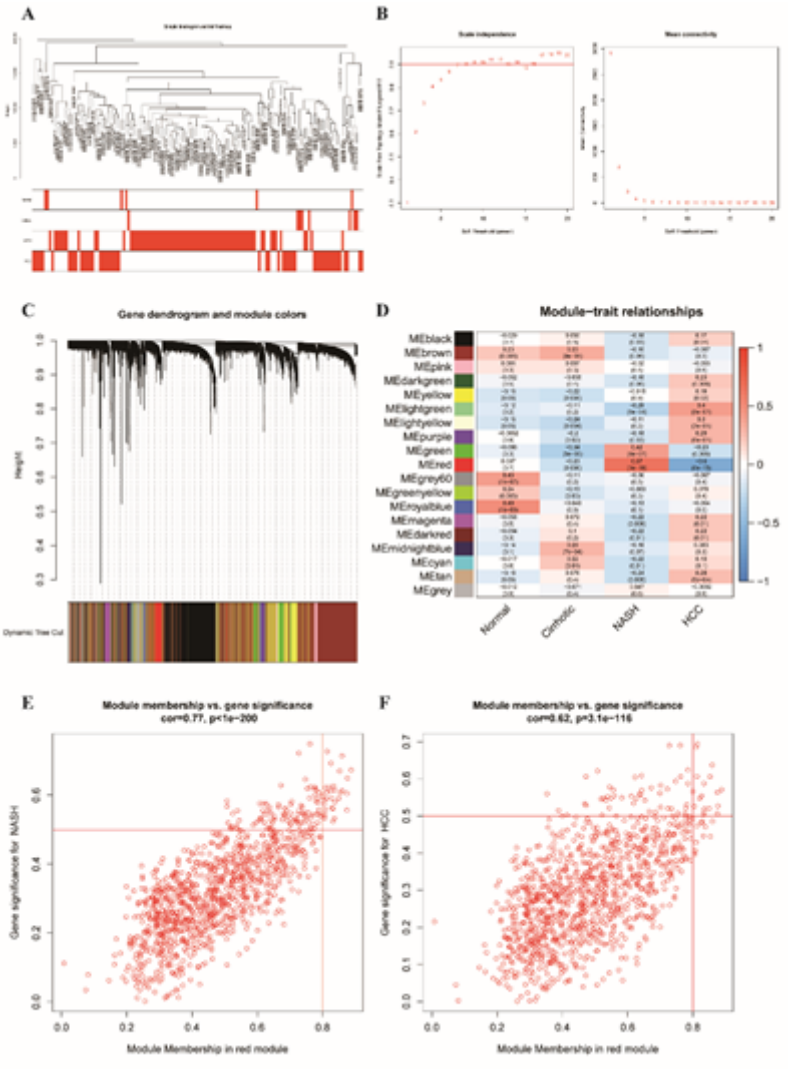

Figure 3

WGCNA to identify trait-related modules and genes. (**A**) Sample dendrogram and trait heat map; (**B**) Calculating soft-thresholding power; Left: scale-free fit indices using different soft-thresholding powers; Right: mean connectivity using different soft-thresholding powers; (**C**) The dendrogram clustered by Dynamic Tree Cut algorithm; (**D**) The heatmap profiling the correlations between module eigengenes and the clinical characteristics; (**E-F**) Scatter plot of gene significance for NASH and HCC in red module.

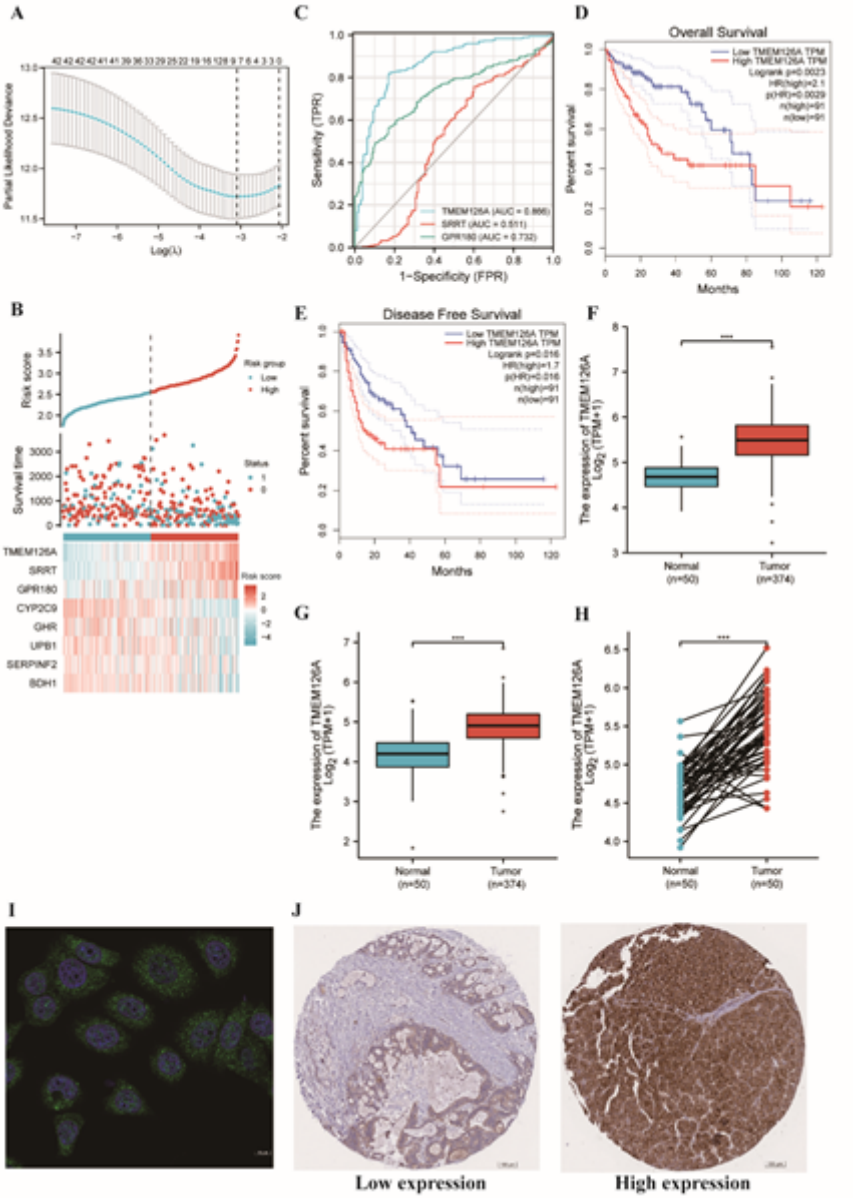

Figure 4

Construction LASSO logistic regression and the value of *TMEM126A* in diagnosis and prognosis (ns, not significant; **p* < 0.05; ***p* < 0.01; *****p** < 0.001). (A) Parameter selection in the LASSO mode; **(B)** Risk score plot for the candidate genes in LASSO mode; **(C)** ROC curve for *TMEM126A*, *SRRT* and *GPR180* in normal samples of GTEx combined adjacent HCC tissues and HCC samples; (**D-E**) Survival plots of *TMEM126A* in overall survival and disease-free survival; (**F**) The difference expression of *TMEM126A* in HCC tissues and adjacent HCC tissues; (**G**) The different expression of *TMEM126A* in normal samples of GTEx combined adjacent HCC tissues and HCC samples; (**H**) The different expression of SEC14L3 in HCC samples and matched adjacent samples; (I) Immunofluorescence staining of the subcellular distribution of *TMEM126A* within cytosol (green) and nucleoplasm (blue) in breast cancer; (**J**) Protein expression of *TMEM126A* between low-expression HCC patients and high-expression HCC patients in the HPA database

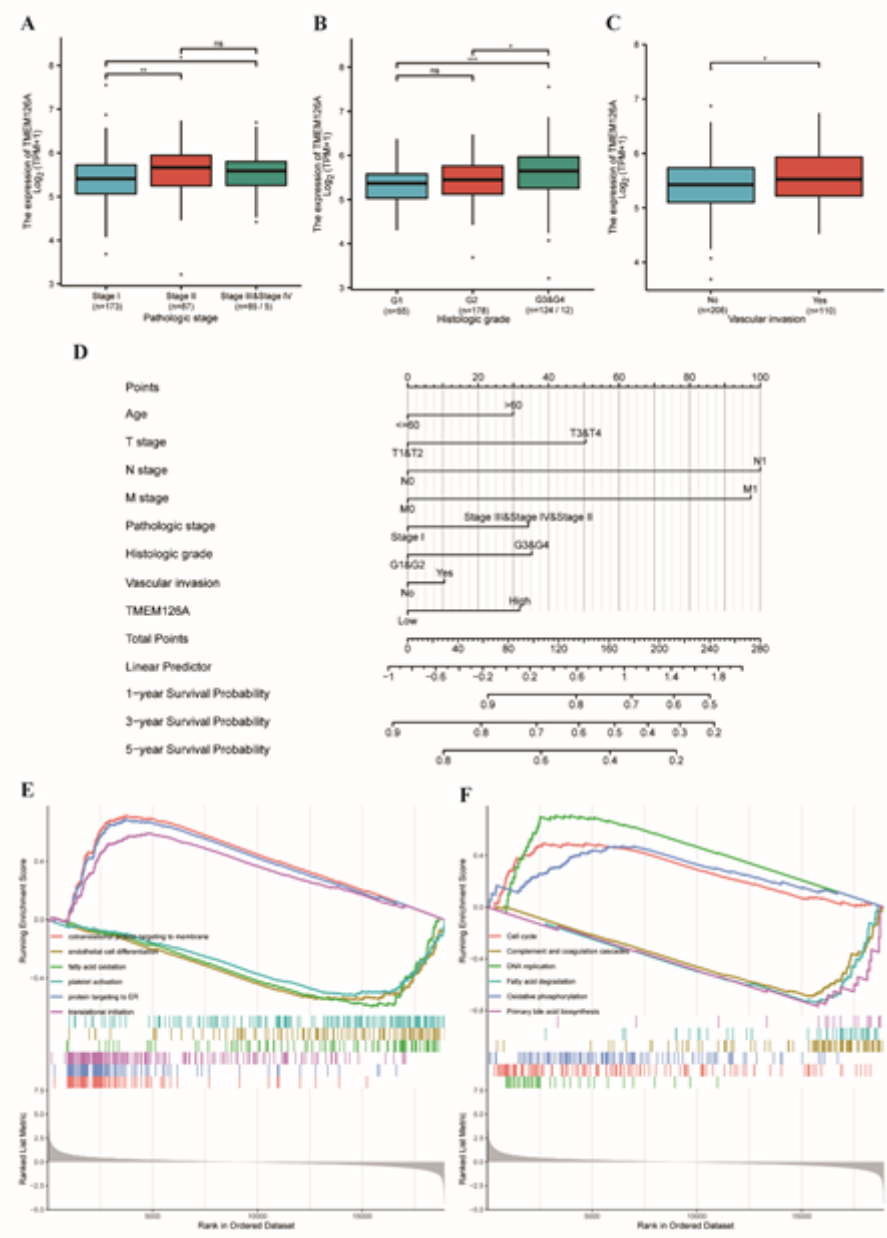

Figure 5

(**A-C**) Box plot assessing *TMEM126A* expression of patients with HCC according to different clinical characteristics, Pathologic stage (**A**); Histologic grade (**B**); Vascular invasion (**C**); (**D**) A nomogram for predicting probability of patients with 1-, 3- and 5-year overall survival; (**E-F**) Enrichment plots form GSEA showing the *TMEM126A* -related signaling pathways in GO (**E**) and KEGG (**F**).

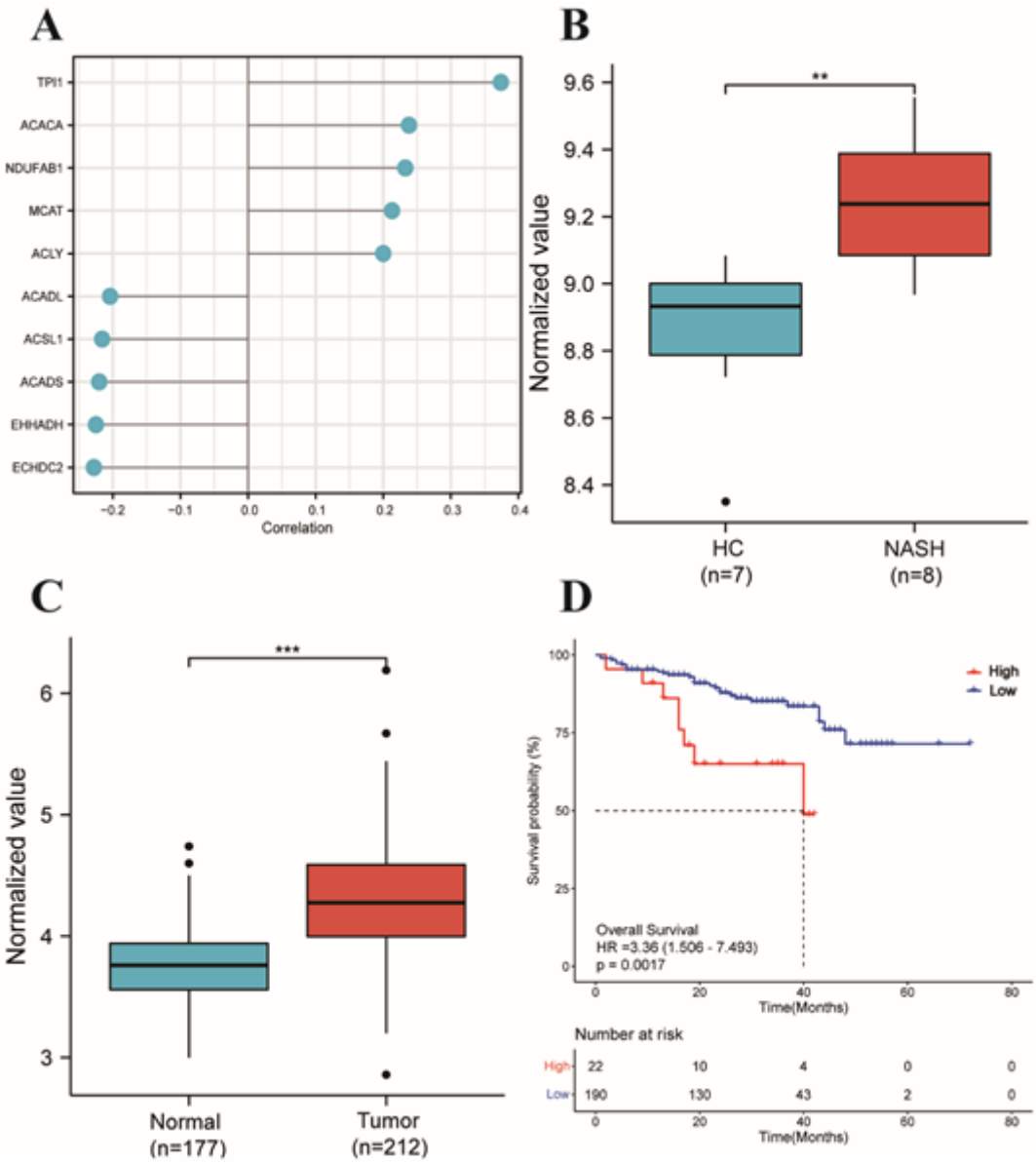

Figure 6

(**A**) Correlation analysis of *TMEM126A* with genes related to lipogenic enzymes and fatty acid oxidation; (**B**) The expression level of *TMEM126A* in GSE37031 dataset; (**C-D**) The expression level and survival analysis of *TMEM126A* in the ICGC dataset. (**$p < 0.01$; ***$p < 0.001$).

Supplementary Files

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

- FigureS1.docx
- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx

- TableS4.xlsx
- TableS5.xlsx
- TableS6.xlsx