Transcription Factors Linked to the Molecular Signatures in the Development of HCC on a Cirrhotic Background

Jamshid Motalebzadeh (motalebzade@gmail.com)
Medical Informatics and Genetics Center (MedInfoGene) https://orcid.org/0000-0002-7780-0933

Elaheh Eskandari
Medical Informatics and Genetics Center (MedInfoGene)

Research Article

Keywords: Gene regulatory network, Hepatocellular carcinoma, Liver cirrhosis, Molecular signatures, Transcription factors

Posted Date: May 25th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-531734/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Mechanisms underlying the regulation of gene expression in cancer have been surveyed for decades to find novel prognostic factors and new targets for molecular targeted therapies in cancer. Because most cases of liver cancer are associated with liver cirrhosis, we aimed to analyze the gene expression signatures and the gene regulatory mechanism in hepatocellular carcinoma (HCC) on a cirrhotic background using high-throughput data analysis. In the present study, three valid array-based datasets containing HCC and liver cirrhosis samples were obtained to identify common differentially expressed genes (DEGs). Moreover, a comprehensive data analysis was conducted based on RNA-Seq data and using Kaplan-Meier curve analysis to find molecular signatures that reduce patients' overall survival rate. Furthermore, we proposed a gene regulatory network (GRN) to explore the possible regulatory mechanism of these molecular signatures by transcription factors in HCC progression from cirrhosis. Besides, we analyzed protein-protein interactions, gene ontology (GO), and pathway enrichment to elucidate the cellular and molecular function of the GRN elements in HCC. In this way, we found a list of 231 molecular signatures in HCC derived from cirrhosis. We also found the importance of TCF4, RUNX1, HINFP, KDM2B, MAF, JUN, NR5A2, NFYA, and AR as key differentially expressed transcription factors (DETFs) in the progression of HCC from cirrhosis. In conclusion, the identified molecular signatures and their transcription factors propose candidate prognostic markers and possible molecular targets in the progression of HCC.

1. Introduction

With a less than 20% chance of survival, liver cancer is categorized among the deadliest cancers [1]. Hepatocellular carcinoma (HCC) is one type of liver cancer which is originated from the main liver cells called hepatocytes. This type of cancer is the third most common cancer-related deaths in the world [2, 3]. Liver cancer is a malignant growth of liver cells and tends to occur in livers damaged by birth defects, alcohol abuse, chronic infection with diseases such as hepatitis B and hepatitis C, metabolic liver disease, hemochromatosis, and cirrhosis [4]. Among these risk factors, cirrhosis is considered as the major cause of hepatocellular carcinoma (HCC) in the United States, Europe, and Asia and comes at the top of the list [5]. Noticeably, more than half of all people that suffer from liver cancer have cirrhosis. Cirrhosis occurs very slowly and can take many years to develop damage to the liver hardly. Both diseases are prevalent in the general population increases the possibility of catching both diseases concurrently [6].

Nonalcoholic fatty liver disease refers to a group of diseases related to liver fat deposition. When the deposition of fat coincides with inflammation and moderate fibrosis, nonalcoholic steatohepatitis (NASH) occurs (Fig. 1A). With raising in inflammation, fibrosis, and regenerative nodules, liver cirrhosis develops [7, 8].

Cirrhosis is the most important factor for the incidence of HCC but, molecular processes especially regulatory mechanisms of the genes in the development of HCC from cirrhosis have been poorly
understood. Despite remarkable progress in the knowledge and management of liver cirrhosis in the past decades, it is still one of the most important reasons for humans mortality. The development of HCC from cirrhosis is mostly affected by genetic factors [9]. Then, we can consider these factors related to liver cell malignancy. Thus, it is important to understand the mechanisms causing the transition of cirrhosis to HCC.

The incidence and mortality of hepatocellular carcinoma (HCC) have remained a major public health issue as one of the most common malignant tumors worldwide. Hepatocarcinogenesis is a very complex biological process associated with many environmental risk factors and factors in heredity, including abnormal activation of cellular and molecular signaling pathways such as Wnt/β-catenin, AKT, MAPK, and ERK signaling pathways [10].

Therefore, the mechanism of gene regulation in cancer has been an important issue for decades and oncologists have come up with new and practical breakthroughs in the fight against cancer. For a basic understanding of this process, it is necessary to identify the mechanism of gene expression patterns in cancer by using high throughput techniques [11]. To elucidate gene expression mechanisms underlying the development of HCC we focused on the transition between cirrhosis and HCC. Studies have indicated that this transition is mostly affected by genetic/epigenetic alterations [9]. Thus, first of all, we found genes associated with HCC derived from cirrhosis and then, we found genes associated with HCC patients' survival rate. Therefore, our study aimed to revealed potential prognosis markers and molecular targeted therapy in HCC.

Molecular biologists require gene regulatory networks (GRNs) because it is a useful way to elucidate the cellular process in living cells. The alteration between genes and their products plays an important role in many molecular processes and induces some abnormalities [12]. Of all regulatory factors in the cell, transcription factors (TFs) are considered as one of the most important elements in the gene expression alteration processes. Then, in our current survey, we found the key differentially transcription factors (DETFs) in the development of HCC on a cirrhotic background. Therefore, in this study we have tried to find the regulatory mechanism by DETFs on the molecular signatures in HCC. The workflow of this study is presented in Fig. 1B.

2. Materials And Methods

2.1 Analysis of DEGs in HCC on a cirrhotic background

Cirrhotic and HCC tissue sample datasets for mRNA expression with patients' pathological information were obtained from the GEO database. We searched for datasets consist of both HCC and cirrhosis samples. Datasets including a small number of samples, cell lines, patients with neoadjuvant therapy and patients without necessary clinical data were excluded from this study. In this regard, we selected three reliable array-based datasets consist of GSE25097 [13], GSE46444 [14], and GSE63898 [15]. To identify DEGs involved in HCC on a cirrhotic background, we compared the gene expression of HCC
samples with cirrhotic tissues. In this regard, each dataset was analyzed separately using limma package in R to assess DEGs [16]. A gene with |FC| > 1.5 and p-value < 0.05 were selected as a DEG in each dataset. Finally, we selected all common DEGs among the three datasets for further analysis. In this way, we used totally 840 cirrhotic and HCC patients' samples for this study. Details are indicated in Table 1.

| Accession number | Contributor | Platform | Number of cirrhosis samples | Number of HCC samples | Total number of patients |
|------------------|-------------|----------|----------------------------|-----------------------|------------------------|
| GSE25097         | Zhang C     | Rosetta/Merck Human RSTA Affymetrix 1.0 microarray | 40 | 268 | 308 |
| GSE46444         | Chen X      | Illumina Human Whole-Genome DASL HT | 48 | 88 | 136 |
| GSE63898         | Villanueva A | Affymetrix Human Genome U219 Array | 168 | 228 | 396 |

2.2 Identification of molecular signatures for HCC patients

Using Kaplan-Meier curve analysis we estimated each common DEG as a potential prognostic marker for HCC [17]. In this regard, we used RNA-Seq data from 364 HCC patients with liver cancer using the KM-plotter database [18]. In a comprehensive data analysis we selected all common up-regulated DEGs with HR > 1 with 95% confidence intervals (CIs) and p-value < 0.05, and also all down-regulated DEGs with HR < 1 with 95% CIs and p-value < 0.05 for HCC. In this way, we selected all common DEGs, which changes in their expression are associated with a lower chance of the patients to be survived with liver cancer. Then, we combined all data to reach a set of molecular signatures in HCC on a cirrhotic background.

2.3 Identification of DETFs and their target genes

Transcription factors with |FC| > 1.5 and p-value < 0.05 were selected as DETFs in comparison of gene expression in HCC derived from liver cirrhosis and then, to identify the acquired DETFs target genes we utilized ChEA and TRANSFAC databases which are designed based on experimental evidence for the identification of TFs target genes [19]. We considered a p-value of less than 0.05 as a significant result.

2.4 Gene regulatory network analysis

Understanding relationships between the architectural properties of gene-regulatory networks (GRNs) has been one of the critical goals in systems biology and bioinformatics. GRNs were constructed and visualized in Cytoscape3.4.0 software based on DETFs and their target DEGs (common DEGs which are associated with patients survival rate) in HCC patients. GRNs were analyzed using Centiscape [20], CytoCtrlAnalyser [21], and Minimum Connected Dominating Set (MCDS) [22] plugins. We created a
network that allows us to predict the impact of molecular signatures and their regulatory mechanism by DETFs in HCC on a cirrhotic background. Based on mentioned points, we used nodes degree, betweenness, classification, control centrality, and MCDS algorithms to find the GRN key regulators. Node degree is determined based on the number of other nodes that are directly connected to a node. Betweenness measures the number of regulators for a node that controls the node over the interactions in the whole network. The node classification was used to identify the controllability of a node in the network, accordingly, the classification for a node is defined as "indispensable", "neutral" or "dispensable" which are correlated with increasing, no effect, or decreasing the number of the minimum number of driver nodes needed to control the network when a specific node is absent. MCDS classifies a node as "dominator", "connector" or "none", a dominator node is a node that provides full control over the network and a connector is a node that connects the dominators in the GRN.

2.5 Protein-Protein interaction network

We investigated protein-protein interactions (PPIs) between components of each GRN using the STRING database [23]. In this database, we used experimentally determined interaction then displayed data with a score of more than 0.4 in the Cytoscape software.

2.6 GO and pathway enrichment analysis

The Database for Annotation, Visualization and Integrated Discovery (DAVID) [24], an essential tool for systematically extracting biological information from numerous genes, was used to perform gene ontology (GO) and pathway enrichment analysis [25, 26]. pathway enrichment analysis was performed using the KEGG database and here, we developed network-construction models to better understanding this relationship [27]. P-value less than 0.01 was considered to indicate a significant difference.

3. Results

3.1 Common DEGs involved in HCC on a cirrhotic background

To reach a set of common DEGs involved in HCC on a cirrhotic background, we performed expression data analysis using three available datasets. We carried out expression data analysis of 256 cirrhotic tissues and 584 HCC samples. According to our criteria, all DEGs were divided into Up- and Down-regulated genes. Noticeably, we selected all common DEGs among the three datasets. In this way, we found 427 common genes which differentially expressed in HCC derived from liver cirrhosis. Our results showed that 85 genes had increased expression and 342 genes had decreased expression.

3.2 Molecular signatures as potential prognostic markers in HCC

Using patients' survival rate analysis from 364 tumors samples of HCC patients, we detected all common DEGs whose expression level is associated with a lower chance of the patients to be survived. Therefore,
we showed a list of 231 molecular signatures in HCC that should be considered as potential prognostic markers for this type of liver cancer. Of all identified molecular signatures, 47 genes are up-regulated and 184 genes are down-regulated in HCC compared to the cirrhotic tissues. In Fig. 2, FC and HR for the identified molecular signatures are illustrated.

### 3.3 Potential DETFs that target the identified molecular signatures

Transcription factors are considered as one of the most important factors in the gene expression process. TFs have either inhibitory or activatory effects on gene expression. In the present study, we found 29 DETFs that regulate the expression of the candidate molecular signatures in HCC patients. These DETFs are consist of AR, AUTS2, BACH1, CTNNB1, E2F7, EOMES, FOXM1, GATA6, HINFP, HMGA1, IRF1, IRF8, JDP2, JUN, KDM2B, MAF, MYCN, NFATC2, NFE2L1, NFYA, NR5A2, PGR, PIA1, POU3F2, PRDM1, RUNX1, SALL4, TCF4, and VDR. In Fig. 3A, a heatmap was presented to show the expression pattern of 29 identified DETEs.

### 3.4 The GRN mechanization

Based on DETFs and their target genes as molecular signatures for HCC on a cirrhotic background, we constructed a GRN. The GRN shown in Fig. 3B is a part of the whole GRN which contains 9 key DETFs and their target genes. The names of the key DETFs, as the most important regulators of the GRN are TCF4, RUNX1, NR5A2, HINFP, NFYA, AR, KDM2B, MAF, JUN. The contribution of each identified key DETFs in the whole GRN is shown in Fig. 3C. All criteria and also the expression pattern for these DETFs are indicated in Table 2.
Table 2
Characteristics of the key transcription factors based on analysis of six parameters.

| Name   | Control centrality | Betweenness | Degree | Classification | MCDS | Expression Pattern |
|--------|--------------------|-------------|--------|----------------|------|--------------------|
| TCF4   | 2                  | 12730.32    | 80     | 1              | Dominator | Down-regulated     |
| RUNX1  | 2                  | 11297.91    | 76     | 1              | Dominator | Down-regulated     |
| NR5A2  | 2                  | 8172.80     | 63     | 1              | Dominator | Up-regulated       |
| HINFP  | 2                  | 6210.87     | 50     | 1              | Dominator | Down-regulated     |
| NFYA   | 2                  | 3593.92     | 42     | 1              | Dominator | Up-regulated       |
| AR     | 2                  | 4348.92     | 39     | 1              | Dominator | Up-regulated       |
| KDM2B  | 2                  | 4423.68     | 38     | 1              | Dominator | Down-regulated     |
| MAF    | 2                  | 3694.37     | 34     | 1              | Dominator | Down-regulated     |
| JUN    | 2                  | 2825.29     | 33     | 1              | Dominator | Down-regulated     |

3.5 The PPI network

Based on the Protein-Protein Interactions (PPI) among the elements of the constructed GRN some protein complexes were identified. The constructed PPI network among all GRN elements was shown in Fig. 4. According to the experimental evidence interaction score of more than 0.4, we found CDC20, JUN, and CTTNB1 as three hub nodes for the constructed network.

3.6 GO term enrichment analysis

To investigate the function of the identified molecular signatures, we analyzed them using DAVID (Database for Annotation, Visualization and Integrated Discovery) which is a web-based tool. In this way, all acquired genes were significantly enriched in biological process, molecular function, and cellular component. Genes enrichment analysis in cellular components showed cytoplasmic part, extracellular region, and vesicle (Fig. 5A) as significant results. Genes were also showed cellular response to chemical stimulus, response to oxygen-containing compound, and single-organism cellular process as significant processes (Fig. 5B). According to Fig. 5C, genes were also enriched in tetrapyrrole, collagen, heme, and calcium ion binding as significant molecular function processes.

3.7 Pathway enrichment analysis
KEGG pathway analysis was used to identify pathway enrichment for identified molecular signatures. In this regard, we found 5 significant pathways involved in molecular signatures in HCC on a cirrhotic background. As indicated in Fig. 5D the most significantly enriched pathways are consist of linoleic acid metabolism, chemical carcinogenesis, cell cycle, drug metabolism-cytochrome p450, metabolism of xenobiotics by cytochrome p450.

4. Discussion

Patients with cirrhosis are at risk of a variety of complex diseases like ascites, esophageal or gastric varices, hepatic encephalopathy, and hepatocellular carcinoma [28]. Cirrhosis is the most important risk factor for the development of HCC and about 80% of HCC cases are caused by liver cirrhosis.

One of the effective methods in the management of HCC on a cirrhotic background is to understand the molecular mechanisms underlying the transition level of cirrhosis to HCC. Studies show genetic factors affect the development of HCC from cirrhosis. Therefore, it is important to understand how HCC develops from cirrhosis to decipher factors involved in cell malignancy. Noticeably, identifying these factors is crucial for screening and molecular targeted therapy for cancer. HCC development is closely linked to cirrhosis, an inflammatory liver disease, in which normal liver tissue is replaced by scar tissue and reconstructive nodes after prolonged damage caused by various causes such as hepatitis B or C viruses [29].

Like many other cancers, HCC develops slowly after the progressive accumulation of alterations in genetic and epigenetic factors, therefore, to decipher novel strategies for the treatment of HCC, the molecular networks and pathways need to be further investigated.

At first glance, it seems that enough studies have been done on HCC, however, there is an urgent need for further research on this subject especially given the path we took in this study. The used algorithms and methods for this study are valuable and efficient due to their lower cost and earlier efficiency. Because of using experimental evidence algorithms, we could access important and key genes along this path with greater confidence and a lower probability of error [30].

As studies have shown, one advantage of compiling large numbers of high throughput data is that the results of different studies can be compared directly [31]. In this study, we had an effort to find a set of cancer-associated genes in HCC on a cirrhotic background using a comprehensive data analysis and systems biological viewpoint. Regarding the analysis of three datasets, we found totally 427 common DEGs in HCC derived from cirrhotic tissues. Then we analyzed the association of patients' overall survival rate for each acquired common DEG and we found a list of 231 molecular signatures which could be considered as prognostic markers for HCC patients. In other words, common DEGs were filtered to identify a list of prognostic markers for HCC on a cirrhotic background whose expression changes reduce the overall survival rate of HCC patients (Fig. 2).
Moreover, using experimental evidence we constructed a GRN consist of 231 genes as molecular signatures and 29 DETFs which affect the expression level of the molecular signatures in HCC. After analyzing the constructed GRN, we found 9 DETFs as the key regulators for the molecular signatures. The DETFs by the names of TCF4, RUNX1, HINFP, KDM2B, MAF, and JUN are down-regulated and NR5A2, NFYA, and AR are up-regulated in HCC (Table 2). This approach shows these 9 transcription factors as one of the most important elements to control the expression levels of genes that are associated with patients' survival rate in HCC on a cirrhotic background which could be considered as targets for molecular target therapies in this type of cancer. The impact of some of these transcription factors has been previously reported in cancer progression and tumorigenesis. The transcription factor TCF4 causes the epithelial to mesenchymal transition and enhances the cancer cell invasion [32]. RUNX1 is a member of the RUNX family which plays important role in the development of cancer and tumorigenesis [33, 34]. KDM2B roles in the alteration of the gene expression as a histone lysine demethylase by epigenetics changes. KDM2B increases cancer cell proliferation and enhances cellular migration by affecting the migration-associated genes [35]. NFYA is reported to be up-regulated in HCC, breast, lung and other types of cancers [36]. The value of NR5A2 as a key regulator for colorectal cancer metastasis was studied [37]. Besides, it was shown that NR5A2 expression level is up-regulated in glioma. Also, the expression level of NR5A2 is considered a poor prognostic factor in glioma patients. In addition, this factor plays a role in cell proliferation, migration, and invasion in malignant glioblastoma cells [38]. We previously showed that the nuclear receptor AR is an important factor for breast cancer development and progression [22]. The value of AR has been shown in HCC as a useful molecule in the molecular targeted therapy for hard-to-treat cancers [39].

In addition, we found protein interaction networks based on experimental evidence that convey some important information about cellular pathways and developing effective therapies for the treatment of cancer (Fig. 4). Proteins are vital components that act as molecular machines, sensors, transporters, and structural elements, with interactions between proteins, being key to their function [40]. In this study, PPI network analysis showed that CDC20 plays as a hub node. CDC20 (Cell division cycle 20) is known as a key element that is remarkably suppressed by p53 introduction and is up-regulated in a wide variety of human cancer tissues. CDC20 is known as a potential cancer therapeutic target that is negatively regulated by p53. We also found JUN and CTTNB1 (Catenin beta 1), both of which play pivotal roles in a variety of cancers, as hub genes in PPI network. The products of the Jun family genes are essential components of the activating protein-1 transcription factor complexes that are critically important in the control of cell growth, differentiation, and neoplastic transformation. It should be noted that JUN plays a crucial role in signal transduction pathways and is involved in cell division, motility, adhesion, and survival in both normal and cancer cells [41], and affects the expression of catenin beta 1 in gastroenteropancreatic endocrine tumors[42].

With regards to the pathway enrichment analysis, we found linoleic acid metabolism, chemical carcinogenesis, and cell cycle as the most significant pathways for HCC patients' molecular signatures. The linoleic acid pathway regulates many physiological processes, then, its metabolic pathway is vital for metabolisms in cancer cells. According to the previous studies, defects in this pathway have been
observed in HCC patients, therefore, our results confirm the previous ones [43, 44]. Besides, pathways involved with chemical carcinogenesis and cell cycle are associated with cancer development. Also, according to the results from the GO analysis, we found cytoplasmic, extracellular region, and vesicle as cellular components associated with HCC on cirrhotic tissue molecular signatures. In addition, GO analysis at the level of molecular function showed the value of some binding functions such as tetrapyrrole, collagen, and heme-binding proteins (Fig. 5).

In conclusion, the validation of these findings is valuable in clinical and pathological research in HCC patients. In the present study, we proposed a comprehensive transcriptome data analysis to find a set of molecular signatures associated with the transition between HCC and cirrhotic tissue. Then we constructed a GRN to highlight the possible regulatory mechanism of these DEGs by DETFs. In addition, we used systems biology approaches to find PPI network, pathways and GO associated with the acquired molecular signatures. The results of this study suggested some key elements that could be used as potential prognostic markers and/or therapeutic targets in cirrhotic and HCC patients to prevent malignancy.

Declarations

Authorship contribution statement

Jamshid Motalebzadeh: Conceptualization, Investigation, Formal analysis, Writing original draft, Writing review and editing, Supervision.

Elaheh Eskandari: Investigation, Formal analysis, Writing original draft.

Conflict of Interest

None

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

All data will be made available on request.

Acknowledgment

This research was conducted using the authors’ personal money.

References

1. Siegel, R.L., et al., Cancer Statistics, 2021. CA: a Cancer Journal for Clinicians, 2021. 71(1): p. 7-33.
2. Sahu, S.K., et al., Rupture of hepatocellular carcinoma: a review of literature. Journal of clinical and experimental hepatology, 2019. 9(2): p. 245-256.
3. Sung, H., et al., *Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries*. CA: A Cancer Journal for Clinicians, 2021. 71(3): p. 209-249.

4. Rees, C.J. and S. Koo, *Artificial intelligence—upping the game in gastrointestinal endoscopy?* Nature reviews Gastroenterology & hepatology, 2019. 16(10): p. 584-585.

5. Janevska, D., V. Chaloska-Ivanova, and V. Janevski, *Hepatocellular carcinoma: risk factors, diagnosis and treatment*. Open access Macedonian journal of medical sciences, 2015. 3(4): p. 732.

6. Fan, F., et al., *Withdrawal of immunosuppressive therapy in allogeneic bone marrow transplantation reactivates chronic viral hepatitis C*. Bone marrow transplantation, 1991. 8(5): p. 417-420.

7. Rocio, G., et al., *Insulin-like growth factor-1 deficiency and cirrhosis establishment*. Journal of clinical medicine research, 2017. 9(4): p. 233.

8. Oliveira, C.P., et al., *Nutrition and physical activity in nonalcoholic fatty liver disease*. Journal of diabetes research, 2016. 2016.

9. Ramakrishna, G., et al., *From cirrhosis to hepatocellular carcinoma: new molecular insights on inflammation and cellular senescence*. Liver cancer, 2013. 2(3-4): p. 367-383.

10. Chen, C. and G. Wang, *Mechanisms of hepatocellular carcinoma and challenges and opportunities for molecular targeted therapy*. World journal of hepatology, 2015. 7(15): p. 1964.

11. Balmain, A., *Cancer genetics: from Boveri and Mendel to microarrays*. Nature reviews cancer, 2001. 1(1): p. 77-82.

12. de Sousa Abreu, R., et al., *Global signatures of protein and mRNA expression levels*. Molecular BioSystems, 2009. 5(12): p. 1512-1526.

13. Sung, W.-K., et al., *Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma*. Nature genetics, 2012. 44(7): p. 765-769.

14. Kojima, K., et al., *Transcriptome profiling of archived sectioned formalin-fixed paraffin-embedded (AS-FFPE) tissue for disease classification*. PloS one, 2014. 9(1): p. e86961.

15. Moeini, A., et al., *An immune gene expression signature associated with development of human hepatocellular carcinoma identifies mice that respond to chemopreventive agents*. Gastroenterology, 2019. 157(5): p. 1383-1397. e11.

16. Smyth, G.K., *Limma: linear models for microarray data*, in Bioinformatics and computational biology solutions using R and Bioconductor. 2005, Springer. p. 397-420.

17. Gong, Q. and L. Fang, *Asymptotic properties of mean survival estimate based on the Kaplan–Meier curve with an extrapolated tail*. Pharmaceutical statistics, 2012. 11(2): p. 135-140.

18. Menyhárt, O., Á. Nagy, and B. Győffy, *Determining consistent prognostic biomarkers of overall survival and vascular invasion in hepatocellular carcinoma*. Royal Society open science, 2018. 5(12): p. 181006.

19. Lachmann, A., et al., *ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments*. Bioinformatics, 2010. 26(19): p. 2438-2444.
20. Scardoni, G., et al., *Biological network analysis with CentiScaPe: centralities and experimental dataset integration*. F1000Research, 2014. 3.

21. Zhou, Z., et al., *Transcriptome Analysis of the Cytokinin Response in Medicago truncatula*. Journal of Plant Biology, 2020: p. 1-14.

22. Eskandari, E. and J. Motalebzadeh, *Transcriptomics-based screening of molecular signatures associated with patients overall survival and their key regulators in subtypes of breast cancer*. Cancer genetics, 2019. 239: p. 62-74.

23. Szklarczyk, D., et al., *STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets*. Nucleic acids research, 2019. 47(D1): p. D607-D613.

24. Lu, Y., et al., *Analysis of long non-coding RNA expression profiles identifies functional IncRNAs associated with the progression of acute coronary syndromes*. Experimental and therapeutic medicine, 2018. 15(2): p. 1376-1384.

25. Long, T., et al., *Identification of differentially expressed genes and enriched pathways in lung cancer using bioinformatics analysis*. Molecular medicine reports, 2019. 19(3): p. 2029-2040.

26. Consortium, G.O., *The Gene Ontology (GO) database and informatics resource*. Nucleic acids research, 2004. 32(suppl_1): p. D258-D261.

27. Abdelzaher, A.F., et al., *Transcriptional network growing models using motif-based preferential attachment*. Frontiers in bioengineering and biotechnology, 2015. 3: p. 157.

28. Pinter, M., et al., *Cancer and liver cirrhosis: implications on prognosis and management*. ESMO open, 2016. 1(2): p. e000042.

29. Bosveld, F., Z. Wang, and Y. Bellaïche, *Tricellular junctions: a hot corner of epithelial biology*. Current opinion in cell biology, 2018. 54: p. 80-88.

30. Wang, E., A. Lenferink, and M. O'Connor-McCourt, *Cancer systems biology: exploring cancer-associated genes on cellular networks*. arXiv preprint arXiv:0712.3753, 2007.

31. Rhodes, D.R., et al., *Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles*. Neoplasia, 2007. 9(2): p. 166-180.

32. Sánchez-Tilló, E., et al., *β-catenin/TCF4 complex induces the epithelial-to-mesenchymal transition (EMT)-activator ZEB1 to regulate tumor invasiveness*. Proceedings of the National Academy of Sciences, 2011. 108(48): p. 19204-19209.

33. Sood, R., Y. Kamikubo, and P. Liu, *Role of RUNX1 in hematological malignancies*. Blood, 2017. 129(15): p. 2070-2082.

34. Hong, D., et al., *RUNX1-dependent mechanisms in biological control and dysregulation in cancer*. Journal of cellular physiology, 2019. 234(6): p. 8597-8609.

35. Yan, M., et al., *The critical role of histone lysine demethylase KDM2B in cancer*. American journal of translational research, 2018. 10(8): p. 2222.
36. Bezzecchi, E., et al., *NF-Y Overexpression in Liver Hepatocellular Carcinoma (HCC)*. International journal of molecular sciences, 2020. **21**(23): p. 9157.

37. Eskandari, E., F. Mahjoubi, and J. Motalebzadeh, *An integrated study on TFs and miRNAs in colorectal cancer metastasis and evaluation of three co-regulated candidate genes as prognostic markers*. Gene, 2018. **679**: p. 150-159.

38. Yang, Q., et al., *NR5A2 Promotes Cell Growth and Resistance to Temozolomide Through Regulating Notch Signal Pathway in Glioma*. OncoTargets and therapy, 2020. **13**: p. 10231.

39. Kanda, T., X. Jiang, and O. Yokosuka, *Androgen receptor signaling in hepatocellular carcinoma and pancreatic cancers*. World Journal of Gastroenterology: WJG, 2014. **20**(28): p. 9229.

40. Snider, J., et al., *Fundamentals of protein interaction network mapping*. Molecular systems biology, 2015. **11**(12): p. 848.

41. Cai, W., et al., *PMP22 regulates self-renewal and chemoresistance of gastric cancer cells*. Molecular cancer therapeutics, 2017. **16**(6): p. 1187-1198.

42. Bacon, C.M., M.-Q. Du, and A. Dogan, *Mucosa-associated lymphoid tissue (MALT) lymphoma: a practical guide for pathologists*. Journal of clinical pathology, 2007. **60**(4): p. 361-372.

43. Cui, H., et al., *Gamma linolenic acid regulates PHD2 mediated hypoxia and mitochondrial apoptosis in DEN induced hepatocellular carcinoma*. Drug design, development and therapy, 2018. **12**: p. 4241.

44. Santoro, N., S. Caprio, and A.E. Feldstein, *Oxidized metabolites of linoleic acid as biomarkers of liver injury in nonalcoholic steatohepatitis*. Clinical lipidology, 2013. **8**(4): p. 411-418.

**Figures**
Figure 1

Schematic representation of pathological changes associated with HCC development on a cirrhotic background and the workflow used in this study to identify molecular signatures and their regulatory mechanism by transcription factors in HCC derived from cirrhotic tissue. (A) Nonalcoholic steatohepatitis (NASH) occurs when fat accumulates and moderate fibrosis occurs in a healthy liver. If NASH gets worse, scars occur in the liver and inflammatory signals are emitted. Then, fibroblasts get stimulated for healing the scars and fibrosis happens. Over time, deposition of fibrosis increases and it leads to more inflammation and cell shrinkage in the liver. Then, cirrhosis is developed which is known as the most important factor for HCC. HCC is the most common type of liver cancer and among the patients with HCC, most of them developed HCC on a cirrhosis background. (B) GRN was constructed based on DETFs
and common DEGs associated with patients' survival rate. Gene ontology, pathway enrichment, and protein-protein interactions were analyzed and the key regulators were assessed.

Figure 2

Molecular signatures for HCC on a cirrhotic tissue. (A) Heatmap shows the patterns of gene expression for the identified molecular signatures in three used datasets. In this figure up-regulated genes are shown at the top and down-regulated genes are shown beneath them. (B) The forest plot depicted the overall burden of HR for 231 identified molecular signatures using KM plot analysis. As shown in this figure, all genes were divided into two groups consist of the up-regulated genes with HR>1 and down-regulated genes with HR<1 with 95% confidence intervals. A p-value<0.05 was considered a significant result for each analysis.
Figure 3

DETFs identification and GRN construction. (A) Heatmap shows the expression pattern of 29 DETFs in comparison of gene expression in three used datasets. (B) GRN among 9 key DETFs and their neighbor target genes. In this figure, nodes represented DEGs and DETFs, DEGs are shown as circular nodes, DETFs as square nodes, red and blue colors represent up and down-regulated genes respectively. (C) Contribution of 29 DETFs in the whole GRN according to their interactions with the target genes. The key DETFs involved in more than half of all interactions in the whole GRN.
Figure 4

PPI analysis shows some protein complexes among the elements of the constructed GRN. The network shows CTNNB1, CDC20 as two up-regulated hubs and JUN as a down-regulated hub in this network. In this network, protein-protein interactions are depicted as edges and nodes represented DEGs and DETFs. DEGs: Circular nodes; DETFs: Square nodes; Up-regulated genes: Red color font; Down-regulated genes: Blue color font.
Figure 5

Gene ontology and functional analysis of the identified molecular signatures in HCC on a cirrhotic background. (A) Most significant results for cellular components analysis. (B) Significant results for biological processes. (C) Significant results for molecular function analysis. (D) Top 5 most significant pathways involved in the identified molecular signatures.