

RESEARCH ARTICLE

Gene expression signatures of stepwise progression of Hepatocellular Carcinoma

Manisri Porukala, P. K. Vinod¹*

Centre for Computational Natural Sciences and Bioinformatics, IIIT, Hyderabad, India

* vinod.pk@iiit.ac.in

Abstract

The molecular pathogenesis of Hepatocellular Carcinoma (HCC) is a complex process progressing from premalignant stages to cancer in a stepwise manner. Mostly, HCC is detected at advanced stages, leading to high mortality rates. Hence, characterising the molecular underpinnings of HCC from normal to cancer state through precancerous state may help in early detection and improve its prognosis and treatment. In this work, we analysed the transcriptomic profile of tumour and premalignant samples from HCC or chronic liver disease patients, who had undergone either total or partial hepatectomy. The normal samples from patients with metastatic cancer/polycystic liver disease/ cholangiocarcinoma were also included. A gene co-expression network approach was applied to identify hierarchical changes: modules, pathways, and genes related to different trajectories of HCC and patient survival. Our analysis shows that the progression from premalignant conditions to tumour is accompanied by differences in the downregulation of genes associated with HNF4A activity and the immune system and upregulation of cell cycle genes, bringing about variability in patient outcomes. However, an increase in immune and cell cycle activity is observed in premalignant samples. Interestingly, co-expression modules and genes from premalignant stages are associated with survival. THBD, a classical marker for dendritic cells, is a predictor of survival at the premalignant stage. Further, genes linked to microtubules, kinetochores, and centromere are altered in both premalignant and tumour conditions and are associated with survival. Our analysis revealed a three-way molecular axis of liver function, immune pathways, and cell cycle driving HCC pathogenesis.

OPEN ACCESS

Citation: Porukala M, Vinod PK (2023) Gene expression signatures of stepwise progression of Hepatocellular Carcinoma. PLoS ONE 18(12): e0296454. <https://doi.org/10.1371/journal.pone.0296454>

Editor: Amy McCart Reed, The University of Queensland Faculty of Medicine, AUSTRALIA

Received: August 17, 2023

Accepted: December 11, 2023

Published: December 29, 2023

Copyright: © 2023 Porukala, Vinod. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its [Supporting Information](#) files.

Funding: PKV acknowledges financial support from the Department of Biotechnology (BT/PR31936/BID/7/861/2019) and IHub-Data, IIIT Hyderabad. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors declare no conflict of interest.

Introduction

HCC is the common form of primary liver tumour and the third-most leading cause of cancer-related deaths globally [1, 2]. Major risk factors leading to HCC include viral infections (Hepatitis B—HBV and Hepatitis C—HCV), excessive alcohol and tobacco consumption, exposure to fungal toxins, and Steatotic liver disease (SLD), with 90% of cases arising from the underlying chronic liver disease [3]. While HBV-driven HCC is prevalent in East Asia and Africa, HCV infections are most common in the US and Europe. SLD is emerging as the leading risk factor of HCC, especially in the West, owing to the rise in obesity and diabetes [4].

Despite continuous advances and management strategies designed to mitigate the incidence of HCC, its mortality rates have been rising over the last two decades. The major caveat in reducing the incidence of HCC is the detection at an early stage since more than 50% of HCC cases are diagnosed at advanced stages [5]. Therefore, a better understanding of HCC pathogenesis and its molecular underpinnings will help reduce the rising cases.

Most HCC cases develop in the background of unresolved chronic inflammation [6] that triggers a persistent healing response [7]. The unbalanced healing response disturbs the architecture of the liver, leading to fibrosis, followed by cirrhosis. The regenerating nodules produced during cirrhosis fuel the transformation of hepatocytes to premalignant lesions called dysplastic nodules. These premalignant lesions develop into early HCC (eHCC) and progressive HCC (pHCC). Although this stepwise progression from chronic liver disease to tumour state is widely prevalent in HCC, about 20% of cases arise from a non-cirrhotic background [8]. While most non-cirrhotic HCCs develop from metabolic syndrome [9], HBV or HCV infection can also lead to HCC from accelerated fibrosis without cirrhosis [8, 10]. Hence, it is crucial to consider the existence of multiple trajectories to HCC when developing diagnostic markers.

Due to the inherently complex nature of HCC development, managing patients is also quite challenging. Surgical resection is the primary treatment for HCC patients with preserved liver function but is prone to recurrence in about 70% of the cases within a few years [5]. Liver transplantation is another option for patients not eligible for resection but is limited by the availability of donors [11]. Under the circumstances where resection or liver transplantation is not amenable, liver-directed medication fails, or recurrence is seen post-resection, systemic therapy is chosen [5]. Systemic therapy in the form of tyrosine kinase inhibitors and immunotherapy targeting immune checkpoints have been developed for treating advanced-stage HCC [12]. Further, treatment strategies must also consider underlying liver disease along with tumour stage [13, 14], which may account for differences in the risk of HCC recurrence among patients [15].

The advancement in the high throughput techniques (next-generation sequencing) is helping to map the molecular changes of HCC at genomic, transcriptomic, and epigenetic levels [6, 16]. This information provides insights into the various signalling pathways involved in hepatocarcinogenesis. These include differentiation and development pathways (Wnt/ β -catenin, Notch Hedgehog signalling), genomic stability and cell cycle (TP53, RB1), telomerase (TERT), growth and cell proliferation (PI3K/AKT/mTOR, RAS/MAPK, EGF/EGFR), angiogenesis (VEGF/VEGFR, PDGF/PDGF α) and chromatin remodelling (ARID1A/ARID1B/ARID2 and MLL signalling) [6, 17]. However, the current understanding of the interplay of various signalling pathways in HCC is far from complete.

Molecular profiling distinguishes diverse subgroups of HCC that are otherwise indistinguishable by conventional histological methods [18]. Gene expression changes in tumour samples are used to predict recurrence and stratify patients into high and low-risk groups [19]. In liver cancer, genes that show a fold change in expression between the normal and tumour samples are a better predictor of survival than considering candidates based on tumour samples alone [20]. Gene expression profile of tumour-adjacent normal tissue is also reported to predict HCC recurrence [21]. Prediction models proposed in these studies are based on differentially expressed genes in tumours or pre-defined gene signatures.

A recent study on a comprehensive analysis of tumour samples, tumour-adjacent normal samples, and normal healthy samples showed that tumour-adjacent normal samples represent an intermediate transcriptomic state between the other two [22]. Therefore, there is a need to explore the multi-step progression of HCC through different trajectories to gain further insights into the molecular mechanisms and develop predictive models. Network-based

approaches provide a suitable platform to extract meaningful information from omics data, hypothesis generation, stratification of disease classes, and discovery of biomarkers [23]. In the present work, we aim to understand the molecular pathogenesis of HCC sequentially from normal to tumour through different premalignant stages. A network-level analysis of the transcriptomic profile of tumour samples and tumour-adjacent normal samples in different liver damage conditions was performed to obtain insights into the transition from normal to pre-cancerous to cancer state. The hierarchical changes: modules, pathways, and genes related to HCC progression and survival prediction were identified.

Methods

Dataset description

Bulk RNA-seq transcriptomics data of HCC progression was obtained from GEO with accession number GSE148355. The dataset consists of tumour and non-tumour samples from HCC patients or patients with the chronic liver disease treated with either total hepatectomy (TH) or partial hepatectomy (PH) at Seoul National University Hospital. Clinical information is available for 54 tumour samples (35 are from TH patients and 19 from PH patients). The dataset also comprises 47 premalignant and 15 normal samples. The normal samples were from patients with metastatic cancer/polycystic liver disease/or cholangiocarcinoma. All non-tumour liver tissues have no evidence of hepatic fibrosis or viral hepatitis. Out of these 47 premalignant samples, 24 are tumour-adjacent normal samples. The premalignant stages include Fibrosis Low (FL)– 10 samples, Fibrosis High (FH) -10 samples, Cirrhosis (CS)– 10, Dysplastic nodule low grade (DL)– 10 samples, and Dysplastic nodule high grade (DH)– 7 samples. All samples were collected after receiving written informed consent from the patients, and the original study was approved by the Institutional Review Board of Seoul National University Hospital. This dataset is referred to as the Korean cohort. The plot summarising clinicopathological features of tumour samples is given in Fig 1.

In addition, we used HCC datasets from TCGA (TCGA-LIHC) and GEO (GSE14520) with available clinical information. TCGA gene expression data and clinical data were obtained from UCSC Xena (<https://xena.ucsc.edu/>). TCGA-LIHC comprises 316 tumour samples with clinical information, and 39 of them have paired normal samples. GSE14520 is a microarray based (GPL3921 platform) gene expression profiling from HCC patients treated with surgical resection. The dataset includes gene expression data of 210 tumour and 210 adjacent normal samples with associated clinical information and is referred to as the Chinese cohort.

Network-based approach

A systems-level analysis was designed to study the pathogenesis of HCC at multiple levels: modules, pathways, and genes (Fig 2). The analysis pipeline was applied to three groups of

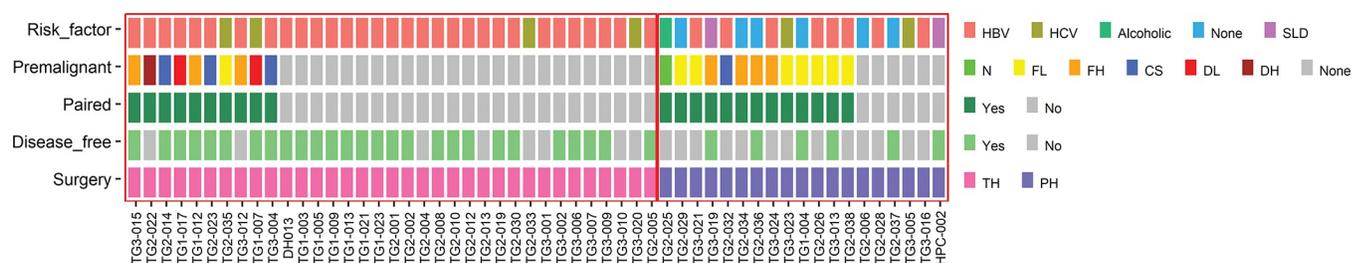


Fig 1. Clinicopathological features of tumour samples in the Korean cohort. This includes the surgery type (TH/PH), disease recurrence, whether the tumour sample has an adjacent normal sample, the premalignant stage of the tumour-adjacent normal, and the risk factor (HBV, HCV, Alcoholic, None, SLD).

<https://doi.org/10.1371/journal.pone.0296454.g001>

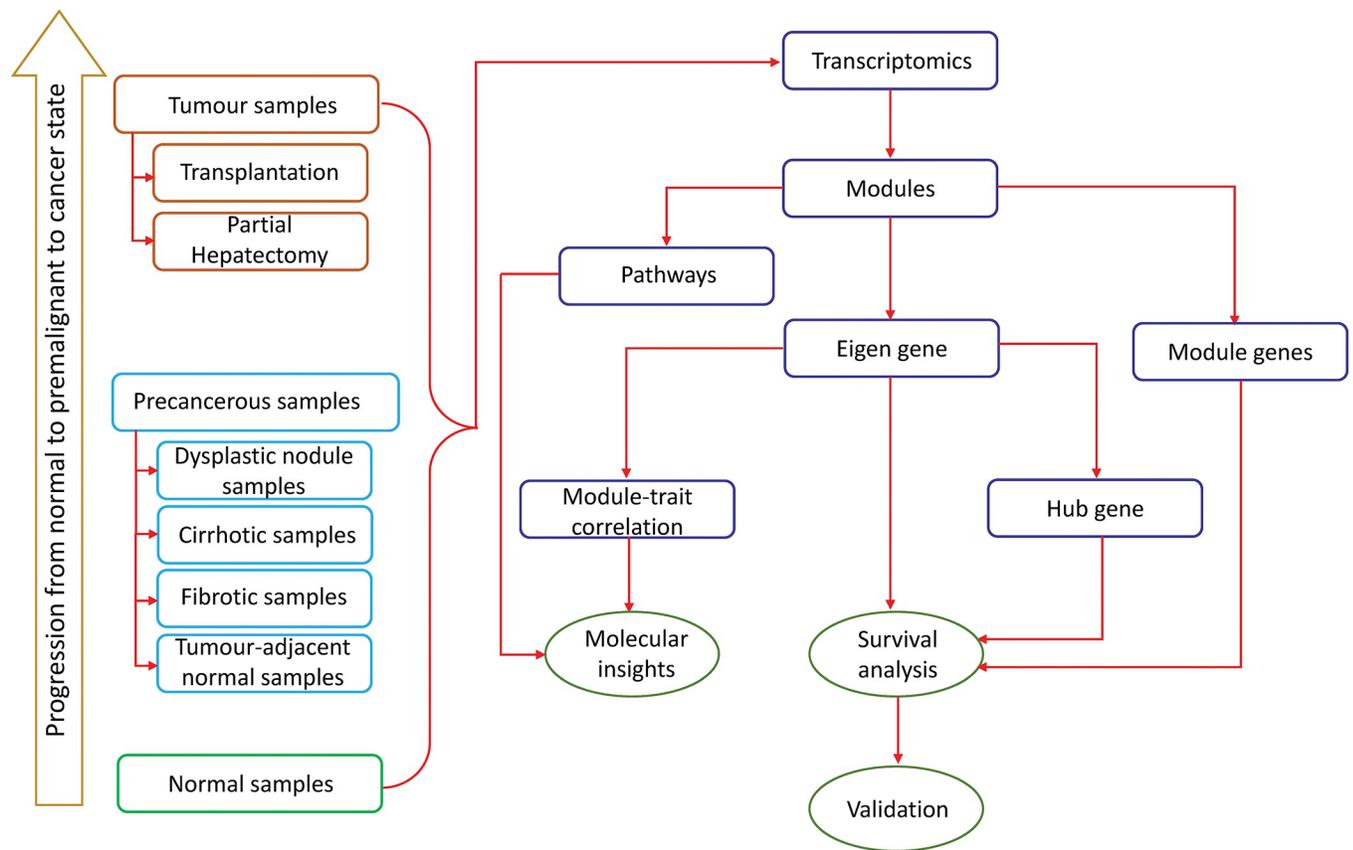


Fig 2. The workflow to study the progression from normal to precancerous to cancer state in HCC.

<https://doi.org/10.1371/journal.pone.0296454.g002>

samples: (a) only tumour samples, (b) adjacent normal and tumour samples, and (c) all normal and premalignant samples. To identify modules, the co-expression network was constructed from gene expression data of the Korean cohort using weighted gene co-expression network analysis (WGCNA) in R [24, 25]. FPKM values were transformed to $\log_2(\text{FPKM} + 1)$, and the top varying genes were selected using the rowVars function to construct the correlation (Pearson) matrix s_{ij} for WGCNA. A signed network was built by transforming the correlation matrix to an adjacency matrix (a_{ij}) using the power adjacency function and soft thresholding ($a_{ij} = s_{ij}^\beta$). Scale-free topology criteria was used to choose the power β . Subsequently, a Topological Overlap Matrix (TOM) was computed from the adjacency matrix, followed by dendrogram construction using $1 - \text{TOM}$. Modules were identified from the dendrogram using the dynamic cut tree algorithm, and module eigengene expression (ME) for each module was calculated using singular value decomposition (SVD).

Modules significantly correlating with disease-free survival (DFS) and other clinical traits were identified. Categorical traits such as surgery/treatment (TH and PH) and premalignant state were converted into continuous numerical values to compute correlation with different modules. For surgery, PH and TH were binarized as 1 and 2, respectively. The premalignant states were converted to numerical with 1, 2, 3, 4, 5, and 6 indicating normal, FL, FH, CS, DL, and DH, respectively. Hub genes from modules were extracted based on module membership ($\text{MM} > 0.8$) and intramodular connectivity.

Candidate genes were selected for univariate survival analysis based on the modules that correlated with DFS under each condition. Samples were dichotomised into two groups based

on the median gene expression profile of candidate genes, and survival analysis was performed using the survival R package [26]. Further, module preservation analysis [27] was carried out using TCGA data as the test set to access the biological relevance of modules identified from the Korean cohort. The Z_{summary} statistics was used to evaluate whether the module is preserved between the reference set (Korean cohort) and test set (TCGA) with the following cut-off:

$$\text{module preservation}(Z_{\text{summary}}) = \begin{cases} \text{moderate}, & 2 < Z_{\text{summary}} < 10 \\ \text{strong}, & Z_{\text{summary}} \geq 10 \end{cases}$$

Pathway enrichment analysis

The enrichment analysis of module genes was performed using Enrichr [28] to identify dysregulated pathways. The ClueGO Cytoscape plugin was used with default settings to visualise the interrelations of the GO biological terms associated with modules [29].

Results

Co-expressed modules of tumour samples

The co-expression pattern of genes within tumour samples (Korean cohort) was studied using top-varying genes. We found five modules (T1, T2, T3, T8, and T9) that significantly correlated with DFS (Fig 3A). Coincidentally, T1 and T9 modules are significantly correlated with the surgery/treatment (i.e., PH or TH) as well. An increase in DFS is associated with TH as the treatment. This is in agreement with the original study [30], which shows TH group has better DFS compared to patients undergoing PH treatment, although there were no differences in grade/stage of tumour in these two groups.

KEGG pathway enrichment of the T9 module showed that cell cycle-related pathways play an important role in governing the survival of a patient post-treatment (Fig 3B). The eigengene expression pattern of this module shows that DFS decreases with an increase in cell cycle activity. The T1 module includes cancer-related genes and pathways relevant for DFS prediction post-treatment. The T2 module that is positively correlated with DFS is enriched for amino acid metabolism, fatty acid degradation, and xenobiotic metabolism, indicating the capability of the liver to carry out its basic functions post-treatment, thus improving survival. It is also associated with complement and coagulation cascades. The T3 module is negatively correlated with DFS and is associated with ribosomes. T4 and T5 modules are associated with the treatment and are related to ECM and regulation of the Wnt signalling pathway, respectively.

Since some modules showed a significant correlation with DFS, we checked if their respective eigengene expression could be used to identify differences in survival probability (Fig 3C). We observed that the corresponding eigengene expression of DFS modules (median) also performs well in predicting the survival probability. Further, the hub genes of these modules also predicted the differences in survival probability and helped us to identify biomarkers. **S1 Data-set** shows the list of hub genes in each module and their association with the DFS of patients. The low expression of the macrophage scavenger receptor gene MARCO is associated with poor DFS in HCC patients. The evaluation of MACRO protein expression by immunostaining in HCC shows that its level decreases as the disease condition worsens [31]. CELC1B is a platelet-related gene, and its expression is related to immune cell infiltration [32]. CFP regulates the complement pathway, and its expression correlates with the infiltration of immune cells [33].

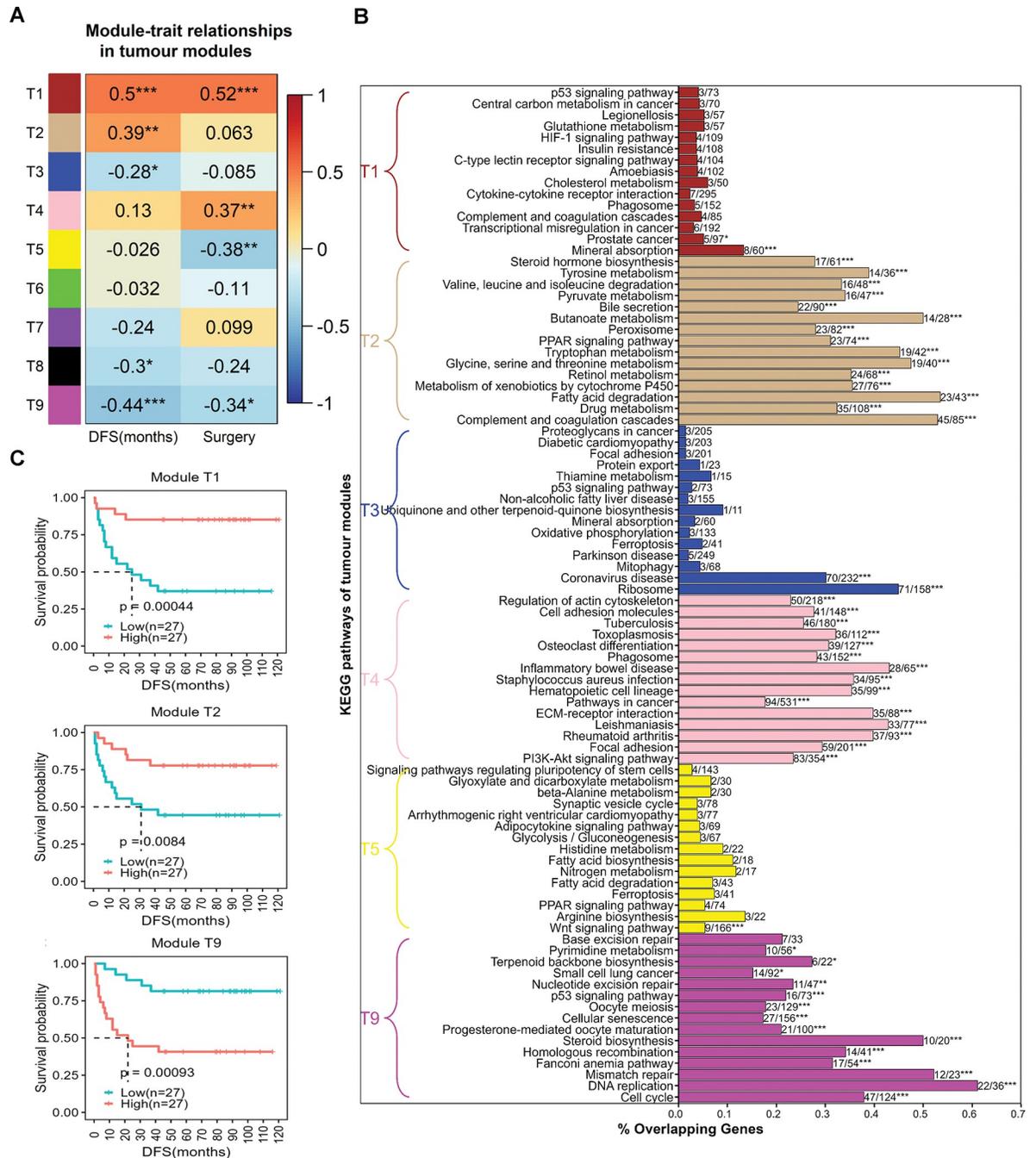


Fig 3. Co-expressed modules of HCC. (A) Module-trait correlations of tumour samples. DFS representing disease-free survival is a continuous variable. Surgery (treatment) is a binary variable with partial hepatectomy (PH) represented as 1 and total hepatectomy (TH) as 2. *** indicates p-value < 0.001, ** indicates 0.001 ≤ p-value < 0.01, * indicates 0.01 ≤ p-value < 0.05. (B) KEGG pathway enrichment of tumour modules. For each module, 15 most significant pathways sorted according to adjusted p-value are displayed (bottom to top within each module). *** indicates adjusted p-value < 0.001, ** indicates 0.001 ≤ adjusted p-value < 0.01, * indicates 0.01 ≤ adjusted p-value < 0.05. The number of overlapping genes and the total number of pathway genes are shown to the right of each bar. (C) Survival analysis based on eigengene expression of tumour modules. Samples are classified into high and low-expression groups based on the median of eigengene expression of each module. 'p' indicates the p-value of survival analysis.

<https://doi.org/10.1371/journal.pone.0296454.g003>

Genes related to the lectin pathway of complement activation (COLEC10, FCN2, FCN3) are also DFS hub genes of the T1 module. The expression of CRHBP, which mediates the reaction between the corticotropin-releasing hormone and its receptor, is also a predictor of DFS in HCC. The hub genes of the T2 module are related to metabolic processes, including TAT, a tumour suppressor gene. MTHDF1, involved in the interconversion of 1-carbon derivatives of THF, is also a DFS hub gene. Hub genes associated with microtubules and chromosomes from the T9 module are also good predictors of DFS in HCC. The modules identified from tumour samples of the Korean cohort are also preserved in tumour samples of TCGA (**Figure S1 in S1 Text**). Further, the genes from the above modules also show significant survival differences in TCGA tumour samples (**Table S1 in S1 Text**).

Progression from precancerous to cancerous state

Module-trait correlation with tumour samples revealed that modules significantly correlated with DFS also captured the differences in surgery a patient has undergone. Based on these observations, we hypothesised that there could be differences in the mechanism of precancerous to cancerous progression between two groups of patients undergoing either TH or PH. Therefore, we investigated the differences in the progression by identifying co-expression modules in each group from both tumour and tumour-adjacent normal samples and correlating them with disease conditions.

The progression from precancerous to cancerous state in both groups shows that liver function is affected in tumour samples. Tumour samples show a decrease in liver function (TH1 module in TH group, PH5 module in PH group) and compromised immune-related pathways (TH4 module in TH group, PH4 module in PH group) (**Fig 4**). The transcription factor enrichment of TH1 and PH5 modules based on ENCODE data shows HNF4A as an associated transcription factor. Modules capturing cell cycle changes in both groups (TH3 module in TH, PH2 module in PH) show a positive correlation in tumour samples. We observed significant correlations to these biological processes in PH compared to TH (**Fig 4A and 4B**). PH4 module shows a higher negative correlation compared to the TH4 module with respect to tumour samples, suggesting that immune response genes are significantly downregulated in PH compared to TH. Another feature difference in precancerous to cancer progression is that the cell cycle module shows a very high positive correlation in PH samples (PH2 module) compared to TH samples (TH3 module). A comparison of these modules in both groups shows some overlap, but the majority of genes are unique to a particular module in a group (**Figure S2A in S1 Text**). These sets of unique genes may account for the difference in the precancerous to cancer progression. This analysis gives a global picture of precancerous to cancer progression in both TH and PH groups fuelled by deviations in liver function, cell cycle, and immune response (**Fig 4**).

In addition to these observations, DEGs comparing tumour versus adjacent normal in both the groups also supports this stark difference in cell cycle and immune response genes between the two groups (**Figures S2B and S3 in S1 Text**). Further, we also observed that oxidative phosphorylation genes are downregulated in the TH group, while genes of choline metabolism in cancer and arginine biosynthesis are upregulated in PH. Genes of Th1 and Th2 cell differentiation, Th17 differentiation, and complement and coagulation cascades are downregulated in the PH group. We also performed module preservation analysis using TCGA samples (**Figure S4 in S1 Text**). The modules from the PH and TH groups show medium to high preservation in TCGA samples.

Co-expressed modules of normal and premalignant samples

The premalignant condition (47 samples) in the dataset ranges from fibrosis (low and high grade) to cirrhosis and dysplastic nodule (low and high). We also included 15 normal samples

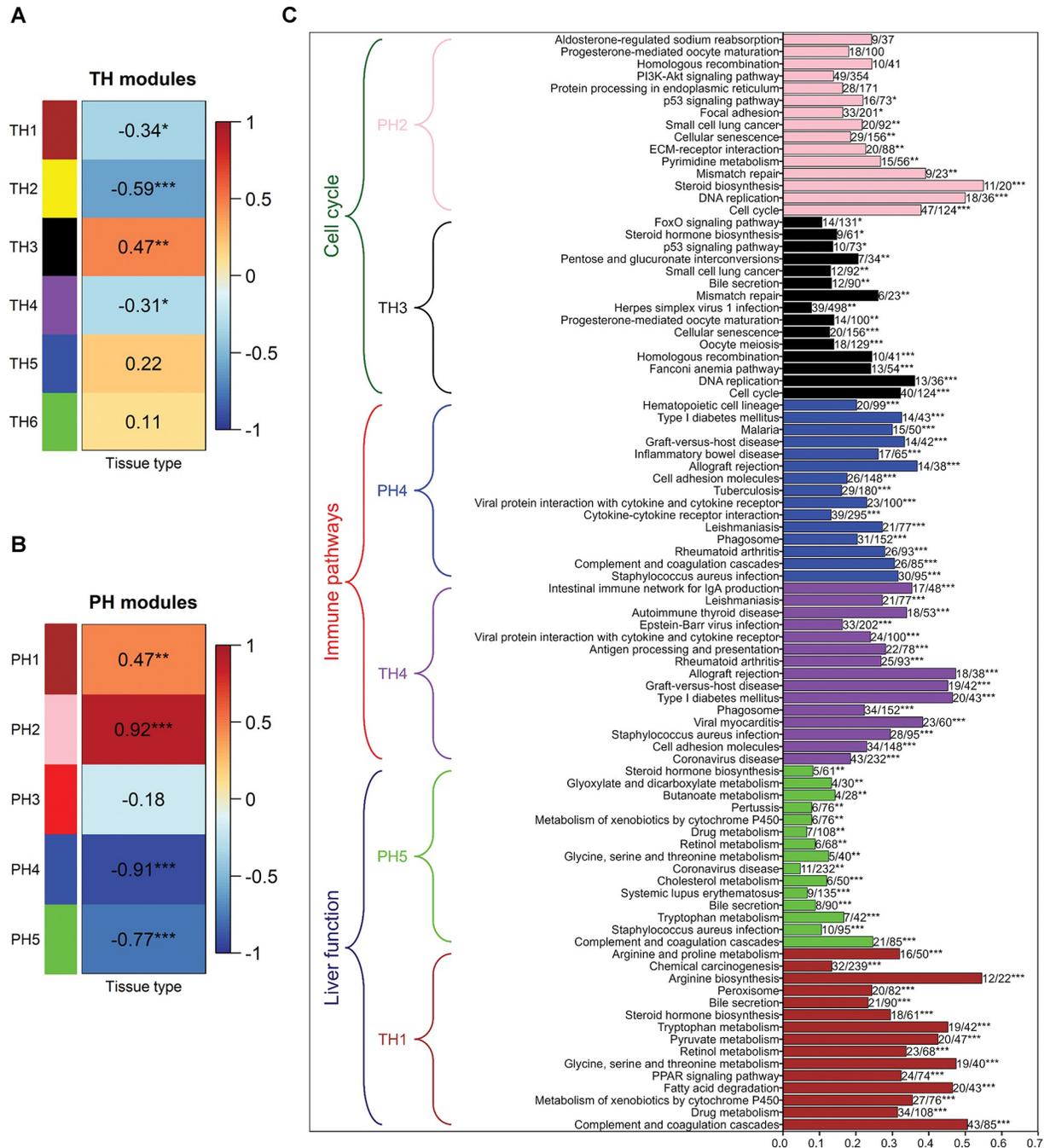


Fig 4. Progression from precancerous to cancer state. Module-trait correlations in (A) TH and (B) PH treatment groups. Tissue type is a binary variable with the precancerous state as 1 and cancer state as 2. *** indicates p-value < 0.001, ** indicates 0.001 ≤ p-value < 0.01, * indicates 0.01 ≤ p-value < 0.05. (C) KEGG pathway enrichment of precancerous to cancer modules. For each module, the 15 most significant pathways sorted according to adjusted p-value are displayed (bottom to top within each module). *** indicates adjusted p-value < 0.001, ** indicates 0.001 ≤ adjusted p-value < 0.01, * indicates 0.01 ≤ adjusted p-value < 0.05. The number of overlapping genes and the total number of pathway genes are shown to the right of each bar.

<https://doi.org/10.1371/journal.pone.0296454.g004>

to capture the changes from normal to premalignant lesions sequentially based on co-expression analysis. We found seven modules significantly correlating with premalignant stages (Fig 5A). The N5 module showed a significantly high correlation to premalignant stages with

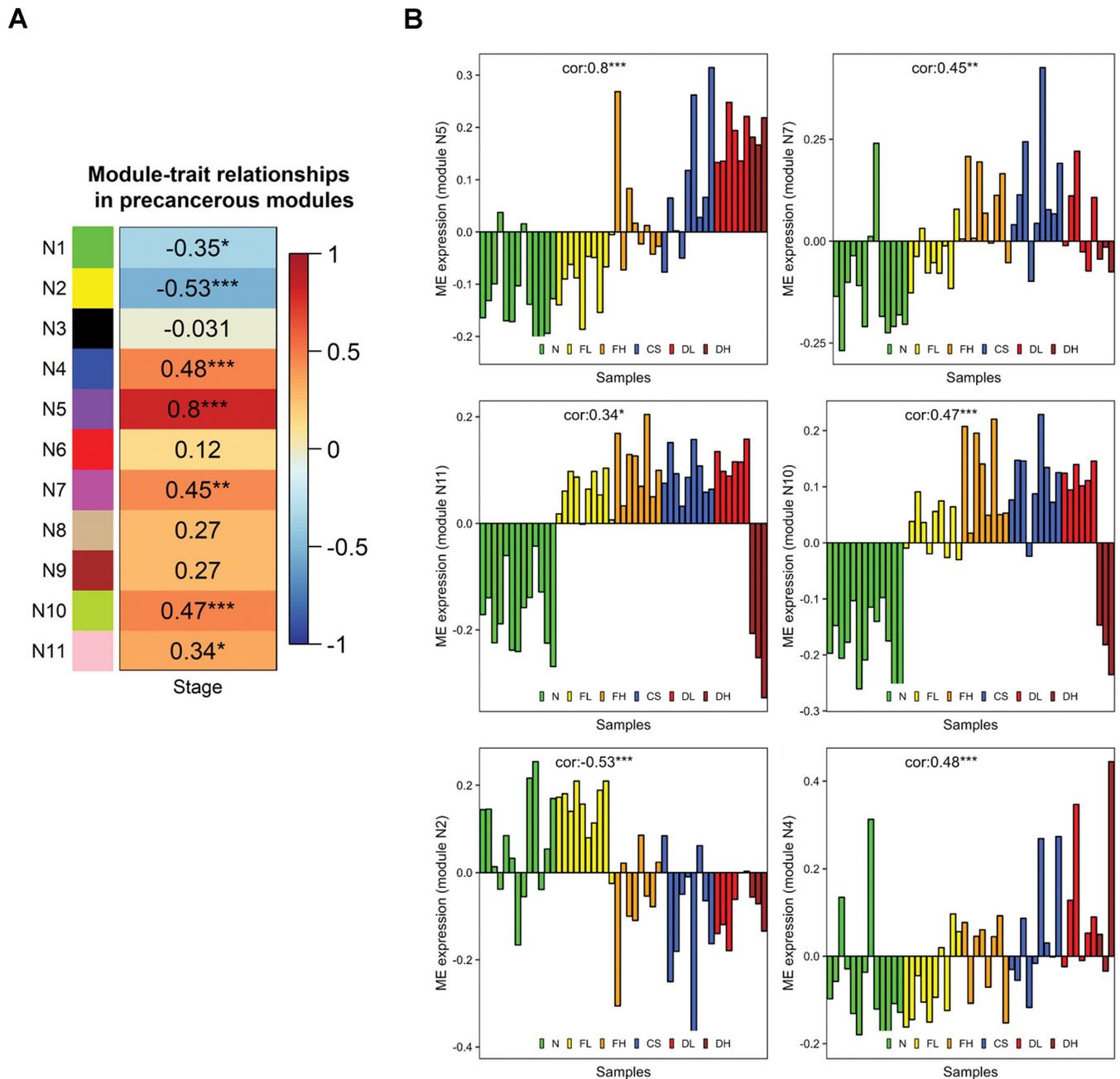


Fig 5. Co-expression modules of normal and premalignant conditions. (A) Module-trait correlations of premalignant samples. The stage represents different premalignant conditions. (B) Eigengene plots for individual premalignant modules showing correlation with premalignant stage. *** indicates p-value < 0.001, ** indicates 0.001 ≤ p-value < 0.01, * indicates 0.01 ≤ p-value < 0.05.

<https://doi.org/10.1371/journal.pone.0296454.g005>

low expression in normal and FL stages and high expression in FH, CS, DL, and DH stages (Fig 5B). This module is associated with cellular response to type 1 interferon, cytokine-mediated signalling, and defense response to the virus (Table S2 in S1 Text). The N7 module is related to neutrophil-mediated immunity and inflammatory response. The N10 module is also positively correlated to premalignant stages, capturing the changes in gene expression that occurred early in the FL stage. These early changes are associated with the complement

coagulation cascade, ribosome machinery, and lipid metabolic process. The N11 module showed a gene expression pattern similar to the N10 module and is associated with mitochondrial oxidative phosphorylation.

The N2 module is negatively correlated with the premalignant stage and is enriched for metabolic pathways linked to liver function and HNF4 transcriptional activity. The eigengene expression shows that liver function is compromised in the late premalignant stages (Fig 5B). Intriguingly, the N4 module that is positively correlated with premalignant stages is associated with cell cycle pathways, showing the onset of the tumourigenesis process.

A previous study on HCC showed that hepatic injury and regeneration (HIR) signature (233 genes) is a good predictor of DFS using premalignant samples from the Chinese cohort [34]. We verified the overlap of the HIR signature with premalignant modules identified through our analysis. We observed that only the N3 module, which is not associated with premalignant states, showed a significant overlap of 61 genes with the HIR signature (Figure S5 in S1 Text). The N3 module is associated with immune pathways and cellular senescence.

We also tested the ability of individual modules to predict the DFS. For this purpose, the Chinese cohort was chosen due to the large sample size with clinical information compared to the Korean cohort. For each module (N1–N11) identified in the Korean cohort, we calculated the corresponding eigengene from paired normal samples of the Chinese cohort, followed by survival analysis based on eigengene expression. The eigengene expression of the N3 module predicts DFS (p-value = 0.004) with a high expression value associated with poor survival (Fig 6). It was observed that 29 out of the 61 intersecting genes between the N3 module and the HIR signature performed well in predicting the DFS in univariate Cox regression analysis (S1 Dataset). These include genes, PLK2, ODC1, WWC1, MYC, DDX21, SOCS3. The high expression of these genes is associated with poor survival. 13 genes out of 96 non-intersecting genes also showed good predictability of DFS (S1 Dataset). Interestingly, we also observed that eigengene expression of modules associated with premalignant stages (N10, N7 and N5) predicted DFS based on normal/premalignant samples (Fig 6). The N5 module yielded the best p-value of 0.0022 in the DFS analysis. THBD (p-value = 0.00035) and BCL2L1 (p-value = 0.0007) are top candidate DFS genes from the N7 module (Table S3 in S1 Text). THBD is a classical marker for dendritic cells (DCs). Increased DCs are associated with early relapse of HCC [35]. BCL2L1 promotes invasion and inhibits apoptosis of liver cancer cells [36]. High expression of FOS (p-value = 0.005) and JUN (p-value = 0.0015) in the N5 module are also associated with poor DFS (Table S3 in S1 Text). Thus, by extracting modules of co-expressed genes from premalignant samples, we identified biomarkers for DFS prediction.

Cell cycle-related pathways change in progression from normal to precancer to HCC

It was observed that cell cycle-related pathways were enriched in premalignant samples (N4 module). Similarly, TH3 and PH2 modules from the precancerous to the cancer stage of TH and PH samples were also associated with the cell cycle. The overlap of these module genes with the cell cycle-related genes obtained from the GO term (S1 Dataset) showed that 55 genes are common and found in precancerous stages (Fig 7A). We observed an increase in cell cycle-related genes with progression from precancer to cancer in TH3 and PH2 modules, having 222 and 365 genes, respectively. There are 169 cell cycle genes unique to the PH2 module.

To gain further insights into the cell cycle processes, genes of the individual modules (N4, TH3, PH2) overlapping with cell cycle genes (56, 222, 365 genes, respectively) were visualised using GO biological processes with ClueGO Cytoscape plugin. The 55 common genes map to biological processes related to the kinetochore, microtubule and chromosome (Fig 7B). GO

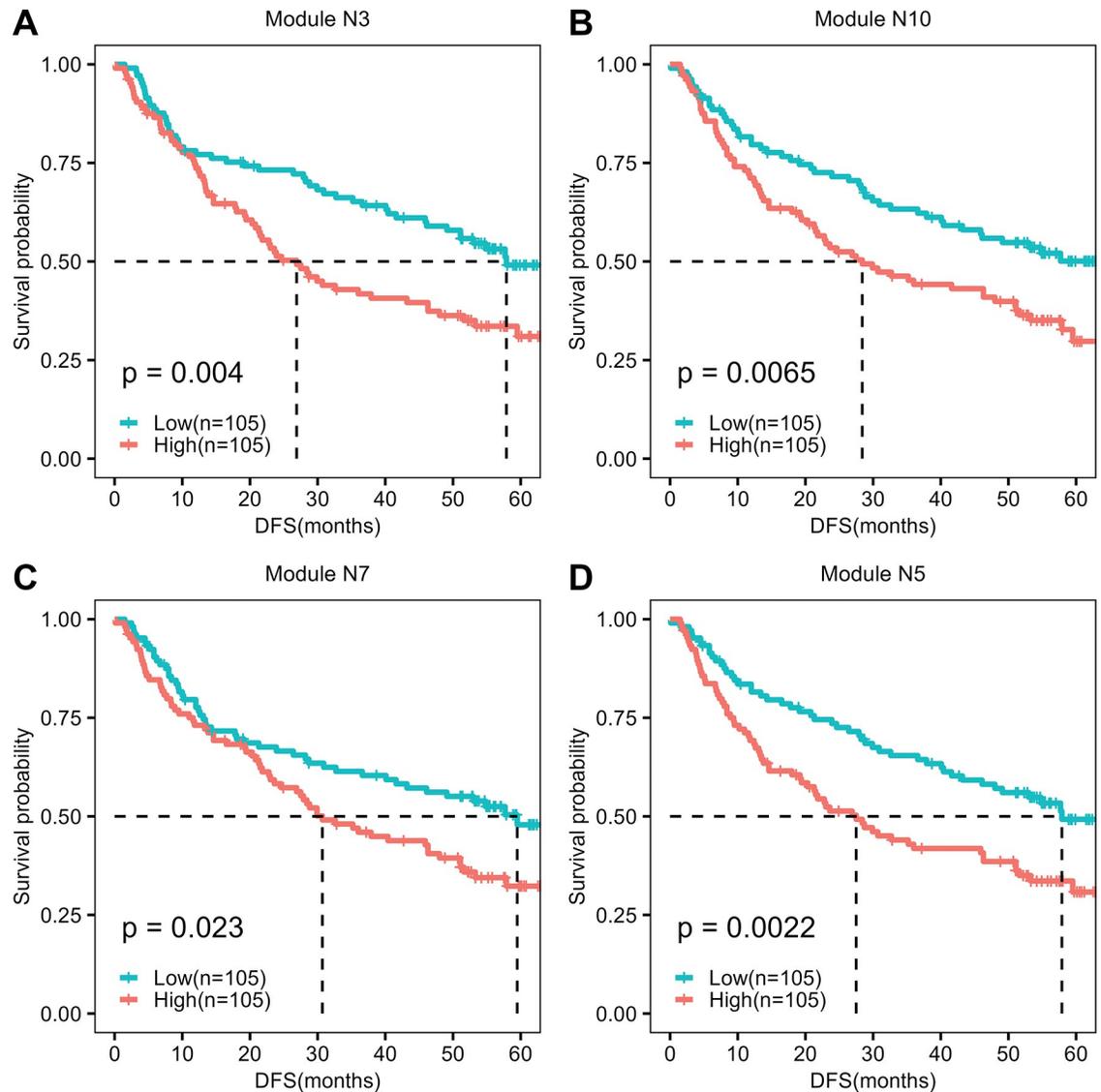


Fig 6. Eigengene-based survival analysis using tumour-adjacent normal samples in the Chinese cohort. The eigengene expression of each premalignant module (obtained from the Korean cohort) was calculated using the tumour-adjacent normal samples in the Chinese cohort. Samples are classified into high and low-expression groups based on the median of eigengene expression of each module. 'p' indicates the p-value of survival analysis.

<https://doi.org/10.1371/journal.pone.0296454.g006>

terms unique to TH3 and PH2 modules suggest the progression differences from precancerous to cancer state between TH and PH conditions. Checkpoint signalling, negative regulation of the cell cycle process, DNA repair process, and regulation of exit from mitosis are observed in the TH3 module but not in the PH2 module (Figs 7C and 8). On the other hand, the PH2 module shows positive regulation of cell cycle, proliferation, cell division, and cytokinesis, along with positive regulation of protein metabolic processes. There is an increase in the number of genes related to microtubule spindle organization compared to N4 and TH3 modules. Further, DFS cell cycle genes related to microtubules, kinetochores, and centromere also overlap with genes of the N4 module, suggesting some of these changes are associated with premalignant stages.

The network analysis of patients who have undergone PH and TH was performed independently, including the tumour and corresponding tumour-adjacent normal samples, to understand the differences in progression. We observed that the same biological processes are affected to a different extent in TH and PH groups. Both groups show a decrease in liver function and immune system and an increase in cell cycle activity. However, the tumour samples in the PH group show a very high correlation to these biological processes (Fig 4). This indicates that the extent of immune suppression and decrease in liver function is related to cell cycle activity in tumour samples, bringing about the variability in the outcomes. This view contrasts with our observations from modules identified from normal and premalignant samples. We observed an increase in immune activity and cell cycle gene expression and a decrease in liver function. An increase in immune activity may be associated with the antiviral mechanism (most patients have HBV infection) by interferon signalling. Genes of immune modules (N5, N7, N10) show some overlap with downregulated immune modules (TH4 and PH4) specific to tumour samples. FOS and JUN are part of the upregulated module in premalignant samples and downregulated modules in tumour samples. This suggests a shift in immune activity from a premalignant state to tumour state. Pro-inflammatory M1 marker CCL2 decreases in tumour modules TH4 and PH4 but increases in the premalignant state. The expression of fibrotic genes EGR1, JUND, KLF2 and TAGLN also decreases in tumour samples (PH4 module).

On the other hand, genes related to liver function decrease in premalignant and tumour samples. HNF4A, which controls liver function, is known to be inhibited by increased inflammation (immune activity) in liver fibrosis [39, 40]. The expression of HNF4A leads to the restoration of metabolic function and reversing (attenuation) of liver fibrosis and cirrhosis via controlling macrophages and hepatic stellate cells [41]. HNF4 drives the transition of macrophages to the M2 phenotype. We hypothesise that the mutual antagonism between HNF4A and immune activity plays a role in HCC progression. An increase in inflammation may result in the inhibition of HNF4A with an increase in cell cycle activity. A progressive loss of HNF4A activity is observed in liver diseases (SLD) compared to HCC [42]. We observed that the transition from normal to premalignant to tumour state is also characterised by an increase in cell cycle activity. Genes related to mitotic spindle organisation are present in the premalignant state, and some of them are also DFS genes in tumour samples. An increase in the expression of genes involved in the maintenance of genomic integrity is associated with chromosomal instability (CIN), which is a prognostic factor in multiple cancers [43, 44]. There is also an emerging link between CIN and tumour immunity.

We observed that multiple (N3, N7, N5, N10) modules from premalignant samples are good predictors of DFS (Fig 6). The N3 module showed some overlap with the HIR signature, which was earlier proposed for DFS prediction. However, we identified three more modules that can be used for the prognostic task. These modules are associated with the immune system. We obtained the best performance (p value = 0.0022) with the eigengene expression of the N5 module (Fig 6D). These modules are associated with premalignant conditions (fibrosis, cirrhosis), and an increase in the eigengene expression is associated with poor survival. This suggests that early relapse can also be predicted based on tumour-adjacent normal immune environment. Most studies on HCC relapse are based on immune cell recruitment in tumour samples. Early-relapse HCC cases have increased recruitment of dendritic cells (DC) and CD8⁺ T cells compared with primary tumours [30, 35]. However, our study showed that the gene expression of tumour-adjacent normal samples of HCC patients contains multiple signatures relevant to predicting DFS.

Limitations

This study on HCC progression was performed solely based on the transcriptomics data. A comprehensive view of progression should also account for alterations at the whole genome

(somatic mutations, copy number variations) or epigenome (histone modifications, methylation) level that may drive transcriptional changes. Further, proteomic profiling during the stepwise progression of HCC is required to confirm the changes at the transcriptomic level. The current analysis is performed on the HCC cohort, which is mostly driven by HBV infection. Given the increasing incidence of SLD-driven HCC, it would be relevant to study the SLD cohort using a similar strategy. Further validation of the work is required to study the transition from different premalignant conditions using larger-size cohorts.

Conclusion

Overall, the network-level analysis of gene expression of HCC in different scenarios, including only tumour samples, tumour and tumour-adjacent normal samples, and normal and premalignant samples, showed that pathways relating to liver function, cell cycle, and immune system are at the interplay of the molecular pathogenesis of HCC. The analysis of precancerous to cancer transition revealed that similar pathways are affected to different extents in PH and TH treatment groups. The characterisation of cell cycle changes showed that the TH group is associated with the negative regulation of cell cycle in contrast to positive regulation of the cell cycle in the PH group. Further, we showed that the gene expression profile of premalignant conditions serves as early biomarkers of HCC and its recurrence. We conclude that this may be due to dynamic changes in gene expression of the biological processes during the stepwise progression of HCC.

Supporting information

S1 Text. Supplementary figures and tables.

(PDF)

S1 Dataset. List of genes. Survival genes based on modules from tumour samples, and normal and premalignant samples; List of genes in each module; HIR signature; Cell cycle genes.

(XLSX)

Author Contributions

Conceptualization: Manisri Porukala, P. K. Vinod.

Formal analysis: Manisri Porukala.

Funding acquisition: P. K. Vinod.

Investigation: Manisri Porukala.

Methodology: Manisri Porukala, P. K. Vinod.

Supervision: P. K. Vinod.

Writing – original draft: Manisri Porukala.

Writing – review & editing: Manisri Porukala, P. K. Vinod.

References

1. Villanueva A. Carcinoma Hepatocellular. Longo DL, editor. *New England Journal of Medicine*. 2019; 380: 1450–1462. <https://doi.org/10.1056/NEJMRA1713263>
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. *Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries*. *CA Cancer J Clin*. 2021; 71: 209–249. <https://doi.org/10.3322/caac.21660> PMID: 33538338

3. Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, et al. Hepatocellular carcinoma. *Nature Reviews Disease Primers* 2021 7:1. 2021; 7: 1–28. <https://doi.org/10.1038/s41572-020-00240-3> PMID: 33479224
4. Hepatocellular carcinoma. *Nature Reviews Disease Primers* 2021 7:1. 2021; 7: 1–1. <https://doi.org/10.1038/s41572-021-00245-6>
5. Leowattana W, Leowattana T, Leowattana PT. Systemic treatment for unresectable hepatocellular carcinoma. *World J Gastroenterol*. 2023; 29: 1551. <https://doi.org/10.3748/wjg.v29.i10.1551> PMID: 36970588
6. Ho DWH, Lo RCL, Chan LK, Ng IOL. Molecular Pathogenesis of Hepatocellular Carcinoma. *Liver Cancer*. 2016; 5: 290. <https://doi.org/10.1159/000449340> PMID: 27781201
7. Montanari NR, Anugwom CM, Boonstra A, Debes JD. The Role of Cytokines in the Different Stages of Hepatocellular Carcinoma. *Cancers* 2021, Vol 13, Page 4876. 2021;13: 4876. <https://doi.org/10.3390/cancers13194876> PMID: 34638361
8. Desai A, Sandhu S, Lai JP, Sandhu DS. Hepatocellular carcinoma in non-cirrhotic liver: A comprehensive review. *World J Hepatol*. 2019; 11: 1. <https://doi.org/10.4254/wjh.v11.i1.1> PMID: 30705715
9. Schütte K, Schulz C, Poranzke J, Antweiler K, Bornschein J, Bretschneider T, et al. Characterization and prognosis of patients with hepatocellular carcinoma (HCC) in the non-cirrhotic liver. *BMC Gastroenterol*. 2014; 14: 117. <https://doi.org/10.1186/1471-230X-14-117> PMID: 24990270
10. Yoo S, Wang W, Wang Q, Fiel MI, Lee E, Hiotis SP, et al. A pilot systematic genomic comparison of recurrence risks of hepatitis B virus-associated hepatocellular carcinoma with low- and high-degree liver fibrosis. *BMC Med*. 2017; 15: 214. <https://doi.org/10.1186/s12916-017-0973-7> PMID: 29212479
11. Chieh Kow AW. Transplantation versus liver resection in patients with hepatocellular carcinoma. *Trans Gastroenterol Hepatol*. 2019;4. <https://doi.org/10.21037/TGH.2019.05.06> PMID: 31231700
12. Sangro B, Sarobe P, Hervás-Stubbs S, Melero I. Advances in immunotherapy for hepatocellular carcinoma. *Nature Reviews Gastroenterology & Hepatology* 2021 18:8. 2021; 18: 525–543. <https://doi.org/10.1038/s41575-021-00438-0> PMID: 33850328
13. Maass T, Sfakianakis I, Staib F, Krupp M, R. Galle P, Teufel A. Microarray-Based Gene Expression Analysis of Hepatocellular Carcinoma. *Curr Genomics*. 2010; 11: 261. <https://doi.org/10.2174/138920210791233063>
14. Alqahtani A, Khan Z, Alloghbi A, Ahmed TSS, Ashraf M, Hammouda DM. Hepatocellular Carcinoma: Molecular Mechanisms and Targeted Therapies. *Medicina* 2019, Vol 55, Page 526. 2019; 55: 526. <https://doi.org/10.3390/medicina55090526> PMID: 31450841
15. Teufel A, Weinmann A, Centner C, Piendl A, Lohse AW, Galle PR, et al. Hepatocellular carcinoma in patients with autoimmune hepatitis. *World Journal of Gastroenterology: WJG*. 2009; 15: 578. <https://doi.org/10.3748/wjg.15.578> PMID: 19195059
16. Ally A, Balasundaram M, Carlsen R, Chuah E, Clarke A, Dhalla N, et al. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell*. 2017; 169: 1327–1341.e23. <https://doi.org/10.1016/j.cell.2017.05.046> PMID: 28622513
17. Badwei N. Hepatocellular Carcinoma Gene Expression: The New Era, Where It goes? *iLIVER*. 2023; 2: 36–40. <https://doi.org/10.1016/J.ILIVER.2022.12.001>
18. Wu Y, Liu Z, Xu X. Molecular subtyping of hepatocellular carcinoma: A step toward precision medicine. *Cancer Commun*. 2020; 40: 681. <https://doi.org/10.1002/cac2.12115> PMID: 33290597
19. Son JA, Ahn HR, You D, Baek GO, Yoon MG, Yoon JH, et al. Novel Gene Signatures as Prognostic Biomarkers for Predicting the Recurrence of Hepatocellular Carcinoma. *Cancers (Basel)*. 2022; 14: 865. <https://doi.org/10.3390/cancers14040865> PMID: 35205612
20. Huang X, Stern DF, Zhao H. Transcriptional Profiles from Paired Normal Samples Offer Complementary Information on Cancer Patient Survival—Evidence from TCGA Pan-Cancer Data. *Scientific Reports* 2016 6:1. 2016; 6: 1–9. <https://doi.org/10.1038/srep20567> PMID: 26837275
21. Hoshida Y, Villanueva A, Kobayashi M, Peix J, Chiang DY, Camargo A, et al. Gene Expression in Fixed Tissues and Outcome in Hepatocellular Carcinoma. *New England Journal of Medicine*. 2008; 359: 1995–2004. <https://doi.org/10.1056/NEJMoa0804525> PMID: 18923165
22. Aran D, Camarda R, Odegaard J, Paik H, Oskotsky B, Krings G, et al. Comprehensive analysis of normal adjacent to tumor transcriptomes. *Nature Communications* 2017 8:1. 2017; 8: 1–14. <https://doi.org/10.1038/s41467-017-01027-z> PMID: 29057876
23. Galindez G, Sadegh S, Baumbach J, Kacprowski T, List M. Network-based approaches for modeling disease regulation and progression. *Comput Struct Biotechnol J*. 2023; 21: 780–795. <https://doi.org/10.1016/j.csbj.2022.12.022> PMID: 36698974
24. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. 2005;4. <https://doi.org/10.2202/1544-6115.1128> PMID: 16646834

25. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008; 9: 1–13. <https://doi.org/10.1186/1471-2105-9-559/FIGURES/4>
26. Therneau T. A package for survival analysis in R. 2023.
27. Langfelder P, Luo R, Oldham MC, Horvath S. Is My Network Module Preserved and Reproducible? *PLoS Comput Biol*. 2011; 7: e1001057. <https://doi.org/10.1371/journal.pcbi.1001057> PMID: 21283776
28. Kuleshov M V., Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016; 44: W90–W97. <https://doi.org/10.1093/nar/gkw377> PMID: 27141961
29. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. 2009; 25: 1091. <https://doi.org/10.1093/bioinformatics/btp101> PMID: 19237447
30. Yoon SH, Choi SW, Nam SW, Lee KB, Nam JW. Preoperative immune landscape predisposes adverse outcomes in hepatocellular carcinoma patients with liver transplantation. *npj Precision Oncology* 2021 5:1. 2021; 5: 1–13. <https://doi.org/10.1038/s41698-021-00167-2> PMID: 33772139
31. Sun H, Song J, Weng C, Xu J, Huang M, Huang Q, et al. Association of decreased expression of the macrophage scavenger receptor MARCO with tumor progression and poor prognosis in human hepatocellular carcinoma. *J Gastroenterol Hepatol*. 2017; 32: 1107–1114. <https://doi.org/10.1111/jgh.13633> PMID: 27806438
32. Liang X, Song F, Fang W, Zhang Y, Feng Z, Chen Z, et al. CLEC1B is a Promising Prognostic Biomarker and Correlated with Immune Infiltration in Hepatocellular Carcinoma. *Int J Gen Med*. 2022; 15: 5661. <https://doi.org/10.2147/IJGM.S363050> PMID: 35734199
33. Qian X, Yang Z, Gao L, Liu Y, Yan J. The role of complement in the clinical course of hepatocellular carcinoma. *Immun Inflamm Dis*. 2022; 10: e569. <https://doi.org/10.1002/iid3.569> PMID: 34813686
34. Kim JH, Sohn BH, Lee HS, Kim SB, Yoo JE, Park YY, et al. Genomic predictors for recurrence patterns of hepatocellular carcinoma: model derivation and validation. *PLoS Med*. 2014; 11. <https://doi.org/10.1371/journal.pmed.1001770> PMID: 25536056
35. Sun Y, Wu L, Zhong Y, Zhou K, Hou Y, Wang Z, et al. Single-cell landscape of the ecosystem in early-relapse hepatocellular carcinoma. *Cell*. 2021; 184: 404–421.e16. <https://doi.org/10.1016/j.cell.2020.11.041> PMID: 33357445
36. Zhang N, Wang F, Zhu L, Chang R, Mok SRS, Peixoto RD, et al. Molecular mechanism of the miR-7/BCL2L1/P53 signaling axis regulating the progression of hepatocellular carcinoma. *Ann Transl Med*. 2023; 11: 12–12. <https://doi.org/10.21037/atm-22-5929> PMID: 36760243
37. Porukala M, Vinod PK. Systems-level analysis of transcriptome reorganization during liver regeneration. *Mol Omics*. 2022; 18: 315–327. <https://doi.org/10.1039/d1mo00382h> PMID: 35072185
38. Xie X, Guo P, Yu H, Wang Y, Chen G. Ribosomal proteins: insight into molecular roles and functions in hepatocellular carcinoma. *Oncogene*. 2018; 37: 277–285. <https://doi.org/10.1038/ncr.2017.343> PMID: 28945227
39. HatziaPOSTOULOU M, PolyTARCHOU C, Aggelidou E, Drakaki A, Poultsides GA, Jaeger SA, et al. An HNF4 α -miRNA inflammatory feedback circuit regulates hepatocellular oncogenesis. *Cell*. 2011; 147: 1233–1247. <https://doi.org/10.1016/j.cell.2011.10.043> PMID: 22153071
40. Goldman O, Adler LN, Hajaj E, Croese T, Darzi N, Galai S, et al. Early infiltration of innate immune cells to the liver depletes HNF4a and promotes extra-hepatic carcinogenesis. *Cancer Discov*. 2023 [cited 25 May 2023]. <https://doi.org/10.1158/2159-8290.CD-22-1062/718964/EARLY-INFILTRATION-OF-INNATE-IMMUNE-CELLS-TO-THE>
41. Yang T, Poenisch M, Khanal R, Hu Q, Dai Z, Li R, et al. Therapeutic HNF4A mRNA attenuates liver fibrosis in a preclinical model. *J Hepatol*. 2021; 75: 1420–1433. <https://doi.org/10.1016/j.jhep.2021.08.011> PMID: 34453962
42. Gunewardena S, Huck I, Walesky C, Robarts D, Weinman S, Apte U. Progressive loss of hepatocyte nuclear factor 4 alpha activity in chronic liver diseases in humans. *Hepatology*. 2022; 76: 372–386. <https://doi.org/10.1002/hep.32326> PMID: 35006629
43. Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet*. 2006; 38: 1043–1048. <https://doi.org/10.1038/ng1861> PMID: 16921376
44. Zhang W, Mao JH, Zhu W, Jain AK, Liu K, Brown JB, et al. Centromere and kinetochore gene misexpression predicts cancer patient survival and response to radiotherapy and chemotherapy. *Nature Communications* 2016 7:1. 2016; 7: 1–15. <https://doi.org/10.1038/ncomms12619> PMID: 27577169