Article

Dynamic edge-based biomarker non-invasively predicts hepatocellular carcinoma with hepatitis B virus infection for individual patients based on blood testing

Yiyu Lu[1,†], Zhaoyuan Fang[2,†], Meiyi Li[2,6,†], Chen Qian[1], Tao Zeng[2], Lina Lu[2], Qilong Chen[1], Hui Zhang[1], Qianmei Zhou[1], Yan Sun[4], Xuefeng Xue[4], Yiyang Hu[5,*], Luonan Chen[2,3,7,8,*], and Shibing Su[1,*]

[1] Institute of Interdisciplinary Integrative Medicine Research, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China
[2] Key Laboratory of Systems Biology, Center for Excellence in Molecular Cell Science, Institute of Biochemistry and Cell Biology, Shanghai Institute of Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China
[3] Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China
[4] Qidong Liver Cancer Institute, Qidong People's Hospital, Qidong 226200, China
[5] Institute of Liver Disease, Shuguang Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China
[6] Minhang Branch, Zhongshan Hospital/Institute of Fudan-Minhang Academic Health System, Minhang Hospital, Fudan University, Shanghai 201199, China
[7] School of Life Science and Technology, Shanghai Tech University, Shanghai 201210, China
[8] Research Center for Brain Science and Brain-Inspired Intelligence, Shanghai 201210, China
[†] These authors contributed equally to this work.
*Correspondence to: Yiyang Hu, E-mail: yyhuliver@163.com; Luonan Chen, E-mail: lnchen@sibs.ac.cn; Shibing Su, E-mail: shibingsu07@163.com

Edited by Jiarui Wu

Hepatitis B virus (HBV)-induced hepatocellular carcinoma (HCC) is a major cause of cancer-related deaths in Asia and Africa. Developing effective and non-invasive biomarkers of HCC for individual patients remains an urgent task for early diagnosis and convenient monitoring. Analyzing the transcriptomic profiles of peripheral blood mononuclear cells from both healthy donors and patients with chronic HBV infection in different states (i.e. HBV carrier, chronic hepatitis B, cirrhosis, and HCC), we identified a set of 19 candidate genes according to our algorithm of dynamic network biomarkers. These genes can both characterize different stages during HCC progression and identify cirrhosis as the critical transition stage before carcinogenesis. The interaction effects (i.e. co-expressions) of candidate genes were used to build an accurate prediction model: the so-called edge-based biomarker. Considering the convenience and robustness of biomarkers in clinical applications, we performed functional analysis, validated candidate genes in other independent samples of our collected cohort, and finally selected *COL5A1*, *HLA-DQB1*, *MMP2*, and *CDK4* to build edge panel as prediction models. We demonstrated that the edge panel had great performance in both diagnosis and prognosis in terms of precision and specificity for HCC, especially for patients with alpha-fetoprotein-negative HCC. Our study not only provides a novel edge-based biomarker for non-invasive and effective diagnosis of HBV-associated HCC to each individual patient but also introduces a new way to integrate the interaction terms of individual molecules for clinical diagnosis and prognosis from the network and dynamics perspectives.

Keywords: hepatitis B virus, hepatocellular carcinoma, diagnosis and prognosis, edge-based biomarker, dynamic network biomarker

Introduction

Hepatocellular carcinoma (HCC) is predicted to be the sixth most commonly diagnosed cancer and the fourth leading cause of cancer death worldwide in 2018 (Bray et al., 2018); most HCC cases result from liver cirrhosis caused by chronic hepatitis B (CHB) or C virus infection (Yang et al., 2018). Notably, nearly half of new HCC cases come from China, which is also a major area

of hepatitis B virus (HBV) infection. An early diagnosis of HCC remains urgent for effective treatment of patients to prevent the progression of HBV-induced HCC and to reduce risk of mortality, especially in the early stage of carcinogenesis.

Currently, the conventional methods for HCC diagnosis include ultrasound, serum alpha-fetoprotein (AFP), computed tomography scanning, and magnetic resonance imaging. Ultrasound is mainly recommended for screening and surveillance; however, it exhibits only a moderate sensitivity of 60% (Singal et al., 2009) and an extreme reliance on operator experience. Although an AFP level of >400–500 ng/ml is considered a gold standard diagnostic criterion for HCC at present, 30% of the patients with a low AFP level are already in an advanced stage (Waidely et al., 2015). These methods have their own limitations in terms of accuracy, sensitivity, and timeliness.

To address these issues, numerous studies have developed and provided novel strategies from different viewpoints, especially for molecular biomarkers (He et al., 2012; Sa et al., 2016; Llovet et al., 2018). Within the bloodstream, peripheral blood mononuclear cells (PBMCs) represent a reservoir of inflammatory cells that contribute to disease progression by different means. It has been repeatedly demonstrated in recent years that genetic expression in PBMCs is altered in the context of malignancy (Baine et al., 2011; Jiang et al., 2016). Some studies have identified characteristic changes (e.g. expression or mutation) of disease-associated genes in PBMCs that could be used for HCC diagnosis, due to the fact that both chronic HBV infection and HCC progression can lead to dysfunctions of the immune system, and PBMCs, as the most common immune cell population throughout the whole body, may play an important role in reflecting these dysfunctions (Shi et al., 2014; Ding et al., 2015; Peng et al., 2016). Recent publications suggest that genes in tumor-educated circulating PBMCs are valuable surrogate markers with diagnostic potential and prognostic applications in different cancer localizations such as lung (Zhou et al., 2016), colorectal (Ganapathi et al., 2014), breast (Suzuki et al., 2019), and digestive cancers (Honda et al., 2010; Marshall et al., 2010). These findings have provided remarkable evidence for developing blood-based tests for HCC diagnosis. However, most of these studies focused on individual and static biomarkers (i.e. node-based or vertex-based biomarkers), which mainly contribute to HCC diagnosis or prognosis based on static and low-dimensional characteristics (e.g. by using single markers with corresponding thresholds according to their different concentrations). Considering the complicated and personalized pathogenesis of HBV-induced HCC, it is necessary to introduce a new way to integrate the interaction effects of individual molecules into clinical diagnosis and prognosis from the perspectives of network and dynamics—the so-called edge-based biomarker (Zeng et al., 2016). Different from the traditional node-based biomarker, the edge-based biomarker requires significant differences of correlations between two biomolecules (e.g. genes and proteins) rather than their individual concentrations to predict the specific stage during disease progression for individual patients (Zhang et al., 2014; Liu et al., 2016a; Yu et al., 2017). The high-dimensional information included in an edge-based biomarker can achieve accurate diagnosis and prognosis, even with individual heterogeneity. In addition, another advantage of an edge-based biomarker is that it can provide clues on pathological mechanisms during disease progression or functional activity (i.e. gain or loss) responses to therapeutic regimens (Zeng et al., 2016).

In this study, to identify novel edge-based biomarkers for specific and non-invasive diagnosis of HCC with HBV infection for individual patients, we collected blood samples from both healthy donors and patients with chronic HBV infection in different states (i.e. HBV carrier (HBVC), CHB, cirrhosis, and HCC). By analyzing the transcriptomic profiles of PBMCs with our developed algorithm at a network level (Figure 1), we identified 19 candidate genes whose properties satisfied dynamic network biomarkers (DNBs) (Chen et al., 2012). After functional analysis and further selection, COL5A1, HLA-DQB1, MMP2, and CDK4 were considered as key genes due to their robustness in an independent cohort. Then, we integrated the interaction terms of the individual genes to build prediction models. Actually, our edge-based biomarkers could be considered to specifically distinguish HCC stage and to further effectively predict outcomes for individual patients.

Results

Dynamic transcriptomic profiles of PBMCs from patients in different stages were measured during HBV-associated HCC and healthy donors

We recruited 306 clinical participants, including healthy donors and patients in different disease states according to the progression of HBV-associated HCC (i.e. HBVC, CHB, cirrhosis, and HCC), according to the corresponding criteria of inclusion and exclusion (Supplementary Table S1), while their clinical information (Supplementary Table S2) and blood samples were collected in a methodical and systemic way (details shown in Methods). Then, we randomly chose blood samples from 3 healthy donors as well as 3 HBVC, 10 CHB, 10 cirrhosis, and 10 HCC patients from the previous cohort to measure the transcriptomic profiles of their PBMCs.

A total of 1906 differentially expressed genes (DEGs) (Supplementary Table S4) were identified by a random variant model (RVM) F-test with a corresponding P-value of <0.05 after FDR correction. Then, we performed unsupervised hierarchical clustering (Figure 2A) based on these DEGs to characterize different stages during the progression of HBV-associated HCC from the normal. Unfortunately, samples at different stages could not be clearly and independently distinguished by hierarchical clustering, except for some of samples at cirrhosis. This result might imply that using molecular signatures at the individual expression level might make it difficult to represent complicated changes in a systematic and organized manner during the initiation and evolution of a complex disease.
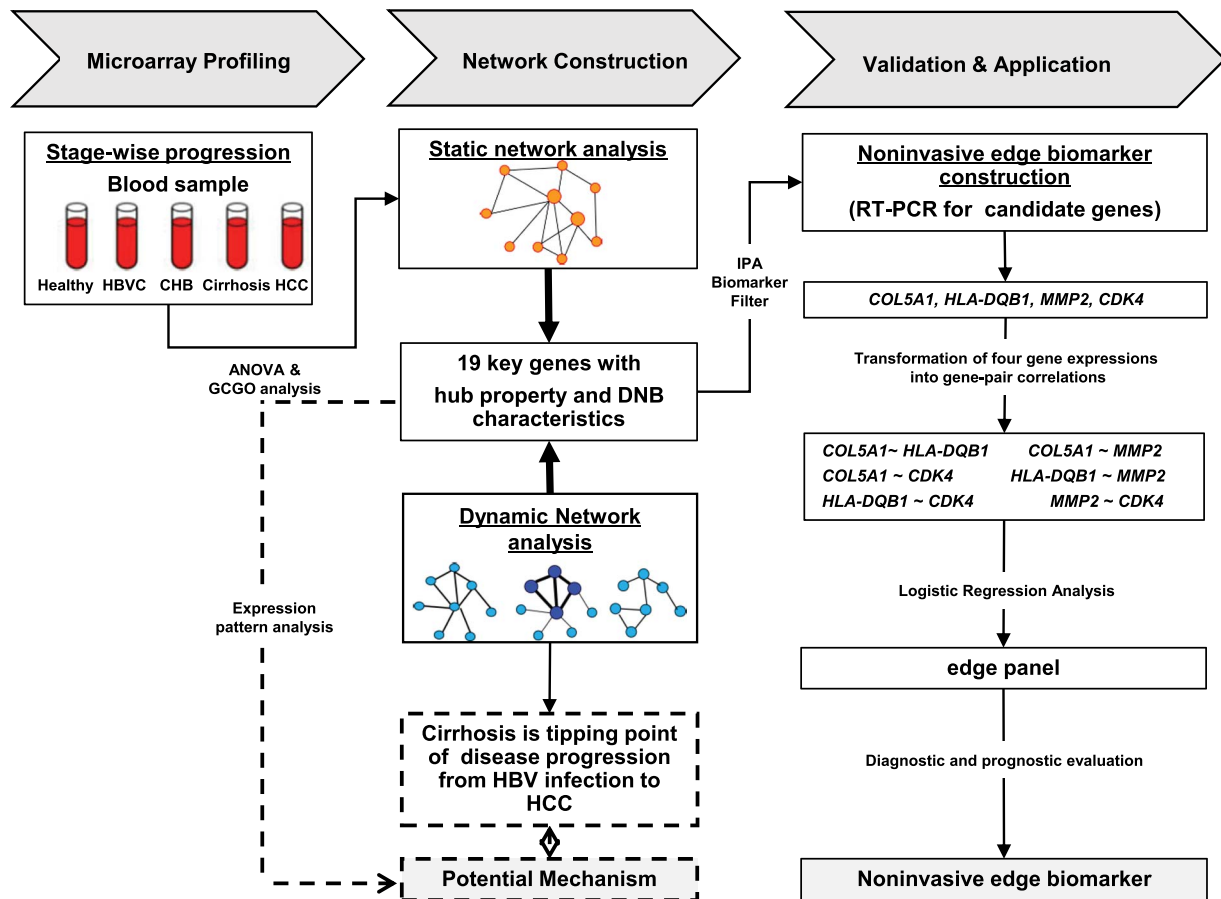
Figure 1 Study flowchart of identifying edge-based biomarkers. Candidate genes were roughly filtered by ANOVA and gene profile clustering with gene ontology analysis (GCGO). Next, we created static correlation networks across all disease stages to identify hub genes for further selecting candidate molecular signatures with network information and then tested these genes with dynamic characteristics based on the three criteria of DNBs, which can predict and signal the critical transition before hepatocellular carcinogenesis. To build the predictive edge-based model with the selected genes, we transformed their gene expression levels (vertex space in math) into gene-pair correlations (edge space in math) based on each single sample data point (Zeng et al., 2016). Finally, a multiple logistic regression model with forward selection techniques was used to build a baseline risk model. Three edges (gene pairs) were retained to form the edge panel, which was significantly associated with an HCC diagnosis in the model.

Stage-characterized candidate biomarkers were identified by static network analysis

To address the above challenge on the poor performance of distinguishing stages by molecular signatures at the individual expression level, we introduced biomarkers with network information. Generally, a complex disease results from dysfunctions, not in individual molecules but via a systematic interplay of multiple molecules and even biological functions (Barabási and Oltvai, 2004; Elkington, 2006). To obtain noninvasive biomarkers to precisely and specifically distinguish HCC stages from other disease stages during the progression of HBV-associated carcinogenesis, we developed an algorithm to identify stage-characterized biomarkers based on network information (Figure 1).

First, based on the changing trends in expression between two consecutive stages, we clustered the DEGs into corresponding profile patterns by cluster analysis of gene expression dynamics

(Ramoni et al., 2002; Hörmann, 2003) (Supplementary Table S5). A total of 7 out of 80 total clusters were selected with statistical significance ($P < 0.05$) (Figure 2B; members of each cluster are listed in Supplementary Table S6).

Then, we constructed a static network weighted by pair-wise correlations among members of the seven significant profile clusters to estimate their functional activity at a comprehensive interactome level during the progression of HBV-associated HCC (details shown in Methods). To investigate the mechanisms of HCC carcinogenesis, we performed functional analyses on active genes with the acquisition or loss of co-expression edges in a stimulus-dependent manner across stages of HCC development. As shown in Figure 2C, typical HCC-associated pathways (e.g. PI3K-Akt, p53, and Wnt signaling pathways) were identified (Tian et al., 2016; Debebe et al., 2017), and their dynamic activities implied the complicated regulations during disease progression. In particular, KEGG pathways associated with cell
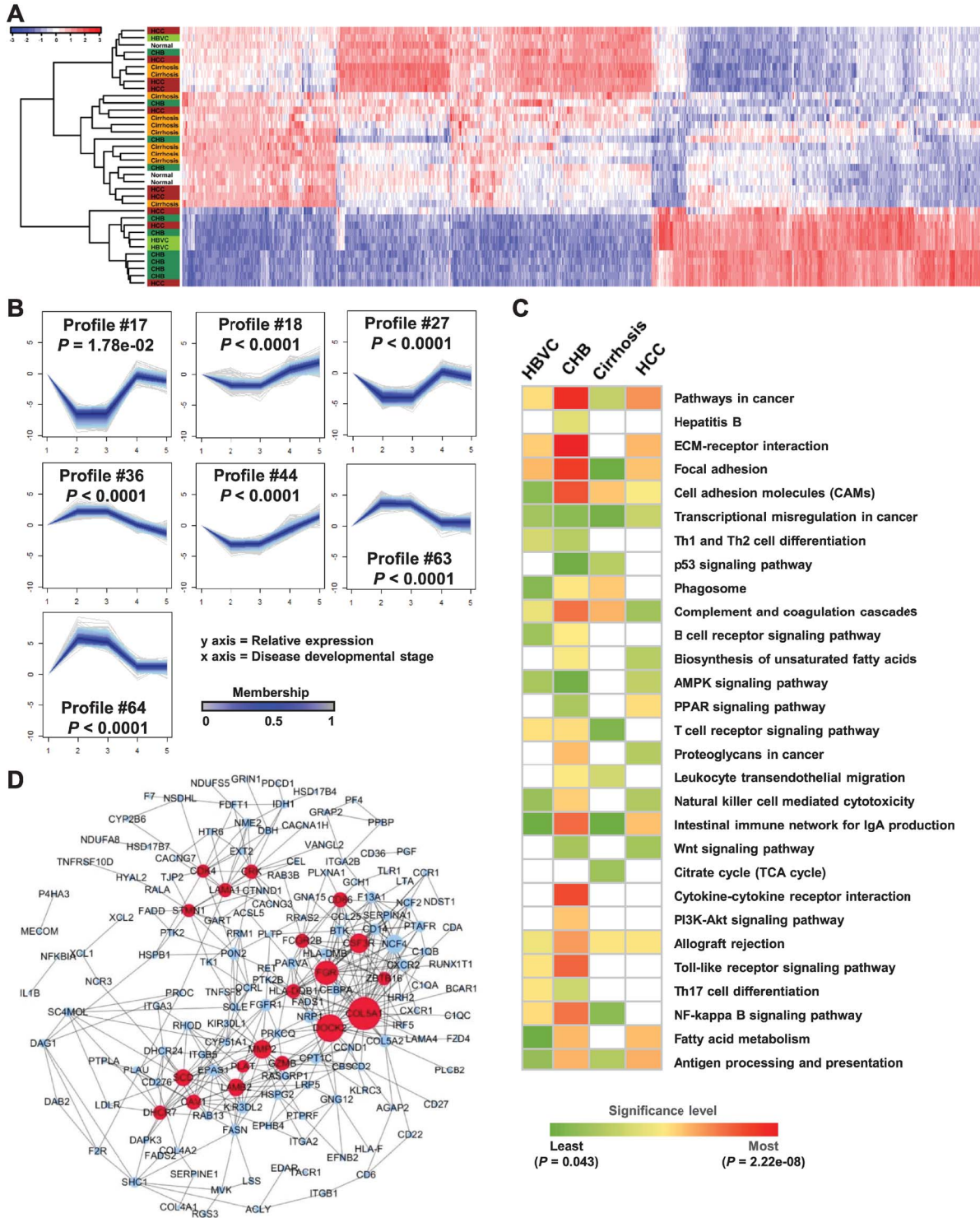
Figure 2 Gene profile cluster analysis and cross-stage network construction from HBV infection to HCC. (**A**) Unsupervised hierarchical clustering was performed to distinguish different stages based on 1906 DEGs. (**B**) Seven unique gene profile clusters (#17, #18, #27, #36, #44, #63, and #64) were identified using cluster analysis of gene expression dynamics according to the time-series gene expression data. The horizontal axis represents different disease stages from healthy to HCC, and the vertical axis represents the gene expression value after log-normalized transformation. The *P*-value indicates significance. (**C**) Functional phenotyping of active genes with the acquisition or loss of co-expression edges in the static network across stages of HCC development. (**D**) The static subnet shows 19 signature genes (red nodes) with the highest degrees, consisting of 159 nodes (genes) and 421 edges (gene pairs). The size of each node represents the degree value.
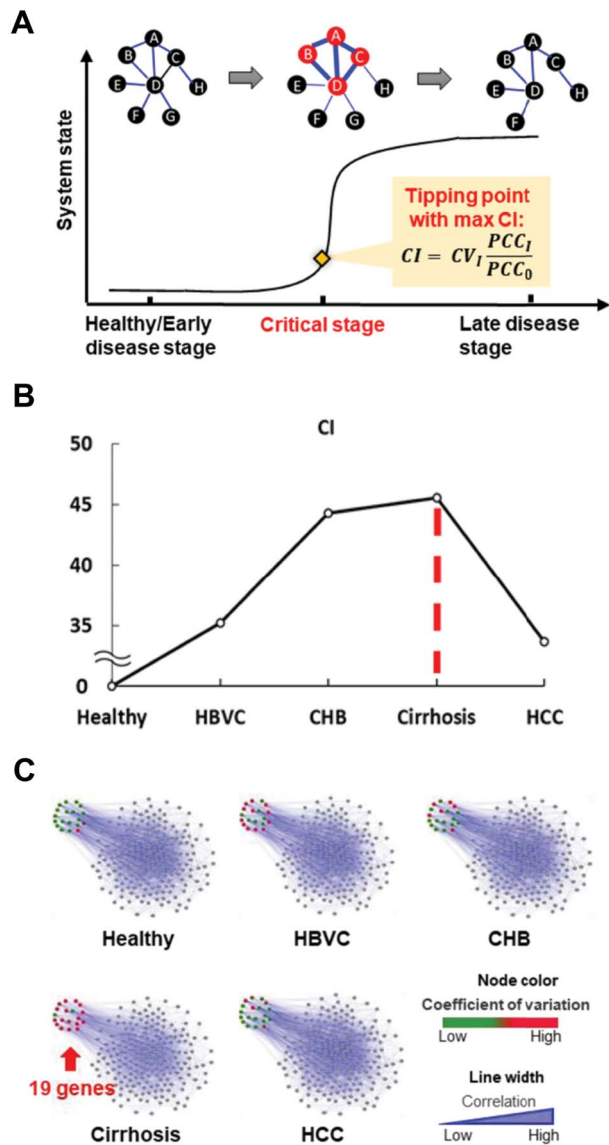
A

B

C

Figure 3 A total of 19 signature genes with dynamic characteristics satisfying the 3 criteria of DNB. (**A**) Brief description of the three criteria of DNB theory. Specifically, near the tipping point, DNB members have a high coefficient of variation (*CV*), the PCC among members of DNB (*PCC$_i$*) increases, and the average PCC between the DNB members and others (*PCC$_o$*) decreases. (**B**) DNB analysis on 19 signature genes from healthy to HCC disease states. The critical stage was identified as cirrhosis, with a max *CI* score. (**C**) Networks graphically illuminate the dynamic changes of the 19 signature genes in the network structure.

metabolism and communications (e.g. focal adhesion, cell adhesion molecules, complement and coagulation cascades, steroid biosynthesis, and transcriptional misregulation in cancer) were significantly and abnormally active across all stages of HBV-associated carcinogenesis; six related to the immune response (i.e. Th1 and Th2 cell differentiation, B cell receptor signaling pathway, cytokine–cytokine receptor interaction, toll-like recep-

tor signaling pathway, Th17 cell differentiation, and chemokine signaling pathway) were enriched only in stages of HBV infection and CHB (Lara-Pezzi et al., 2001). Similarly, results of GO analysis showed that dysfunctions in HBV infection and CHB are mainly involved in fatty acid/lipid oxidation and regulation of cell proliferation, while dysfunctions in HCC initiation and development are mainly involved in regulation of the Wnt signaling pathway (Supplementary Table S7 and Figure S1). These findings might suggest that complicated dysfunctions in immune responses from HBV infection to final HCC were involved in a failure of anti-tumor immunity and an altered tumor microenvironment (Marshall et al., 2013).

Generally, hub molecules have been considered to play important regulatory roles in conducting biological functions and maintaining stability of a network. Thus, we extracted 19 hub genes with the highest degrees as candidate biomarkers (Figure 2D). These signature molecules presented stage-characterized dynamics during the progression of HBV-associated HCC (Supplementary Table S8); furthermore, their involved biological functions and processes were significantly related to human hepatic dysfunctions and hepatocarcinogenesis (Supplementary Figure S2). These functional analysis results showed that these candidate biomarkers played different and complicated roles during the progression of HBV-associated HCC.

The candidate molecular signatures satisfied the three criteria of DNBs

Traditionally, molecular biomarkers for disease diagnosis or prognosis are mainly based on their static information (e.g. different concentrations), which cannot provide early-warning signals at the tipping point or critical period (i.e. pre-disease state) during disease progression. To address this issue for early diagnosis, DNBs were introduced based on the three criteria from the non-linear dynamic theory (e.g. fluctuation changes of concentrations of leading molecules, dynamic associations, or correlations among leading molecules) to predict malignant phase transition during disease initiation and deterioration (Chen et al., 2012). Here, we tested whether these 19 candidate molecules have dynamic characteristics according to the three criteria of DNBs. The DNB method has been applied successfully to studies in multiple fields (e.g. disease diagnosis and prognosis, therapeutic response, and cell differentiation) (Mojtahedi et al., 2016; Richard et al., 2016; Lesterhuis et al., 2017; Li et al., 2017b; Yang et al., 2018). Specifically, if a system approaches the critical phase transition, a leading group of indicative molecules (i.e. DNBs) appear and satisfy the following three conditions: (i) drastic increases in fluctuation for molecular concentrations, which can be evaluated by the coefficient of variation (*CV*); (ii) increases in correlations among molecules in this leading group, which can be evaluated by Pearson correlation coefficients (PCCs) (*PCC$_i$*); and (iii) decreases in correlations between molecules in this leading group and others, which can be evaluated by PCCs (*PCC$_o$*). A criterion index (*CI*) is provided as a numerical signal by comprehensively
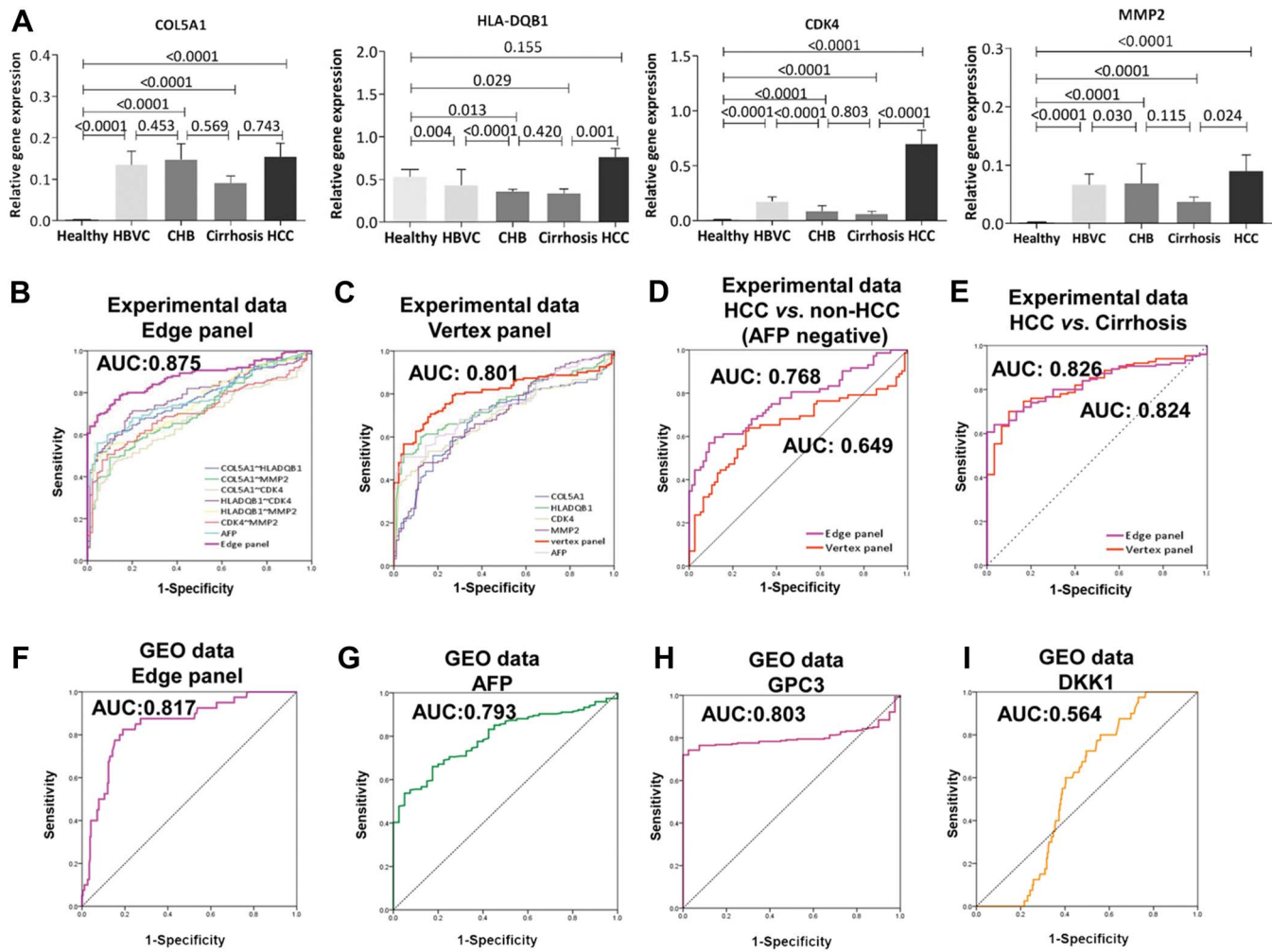
Figure 4 Diagnostic capacity assessments for the four core candidate genes. (**A**) RT-PCR validation of *COL5A1*, *HLA-DQB1*, *CDK4*, and *MMP2* in blood samples of other independent participants from our cohort. The vertical axis represents GAPDH-normalized relative expression values and their expression differences among stages during the progression of HBV-associated HCC measured by Mann–Whitney U tests. (**B–E**) ROC curve analysis of the logistic panel, individual genes, and panels in AFP-negative (AFP <20 ng/ml) patients (n = 72 out of 150 HCC samples; n = 77 non-HCC samples), HCC vs non-HCC patients and cirrhosis patients vs HCC patients. (**F–I**) Comparison of the edge panel, the traditional biomarker AFP, and the new proposed biomarkers GPC3 and DKK1 in the GEO data set GSE25097 by ROC curve analysis.

integrating the three parameters (CV, PCC_i, and PCC_o) of the DNB method. When CI reaches the peak across the measured stages, the corresponding stage is considered the critical period of the biological system (Figure 3A).

After calculating the CIs according to expression data of these candidate molecules in each stage during HCC progression, we found a strong signal of critical stage before hepatocarcinogenesis at cirrhosis (Figure 3B), which is considered as the tipping point. This result is consistent with the liver morphological alterations during the initiation and evolution of HBV-associated HCC (Wang et al., 2002; Sung et al., 2008). As shown in Figure 3C, the series of networks graphically demonstrated the dynamic changes in the network structure and expression variations of these candidate biomarkers. This result also confirmed that the 19 signature genes possessed both network features (hubs) and

dynamic features (DNBs) of HCC progression that not only characterized disease progression but also provided early-warning signals of disease deterioration.

The candidate signatures were validated in other independent samples of the clinical cohort

In clinical practice, biomarkers are expected to provide accurate predictions and to be convenient on operation. Therefore, it is necessary to further select core typical markers from the previous 19 candidate molecules. After literature mining and Ingenuity Pathway Analysis (IPA) biomarker searching, 8 out of 19 molecules were further selected and confirmed to have significant relationships to hepatic diseases (Supplementary Table S9). Then, to obtain the robust signatures for predicting the progression of HBV-associated HCC, we
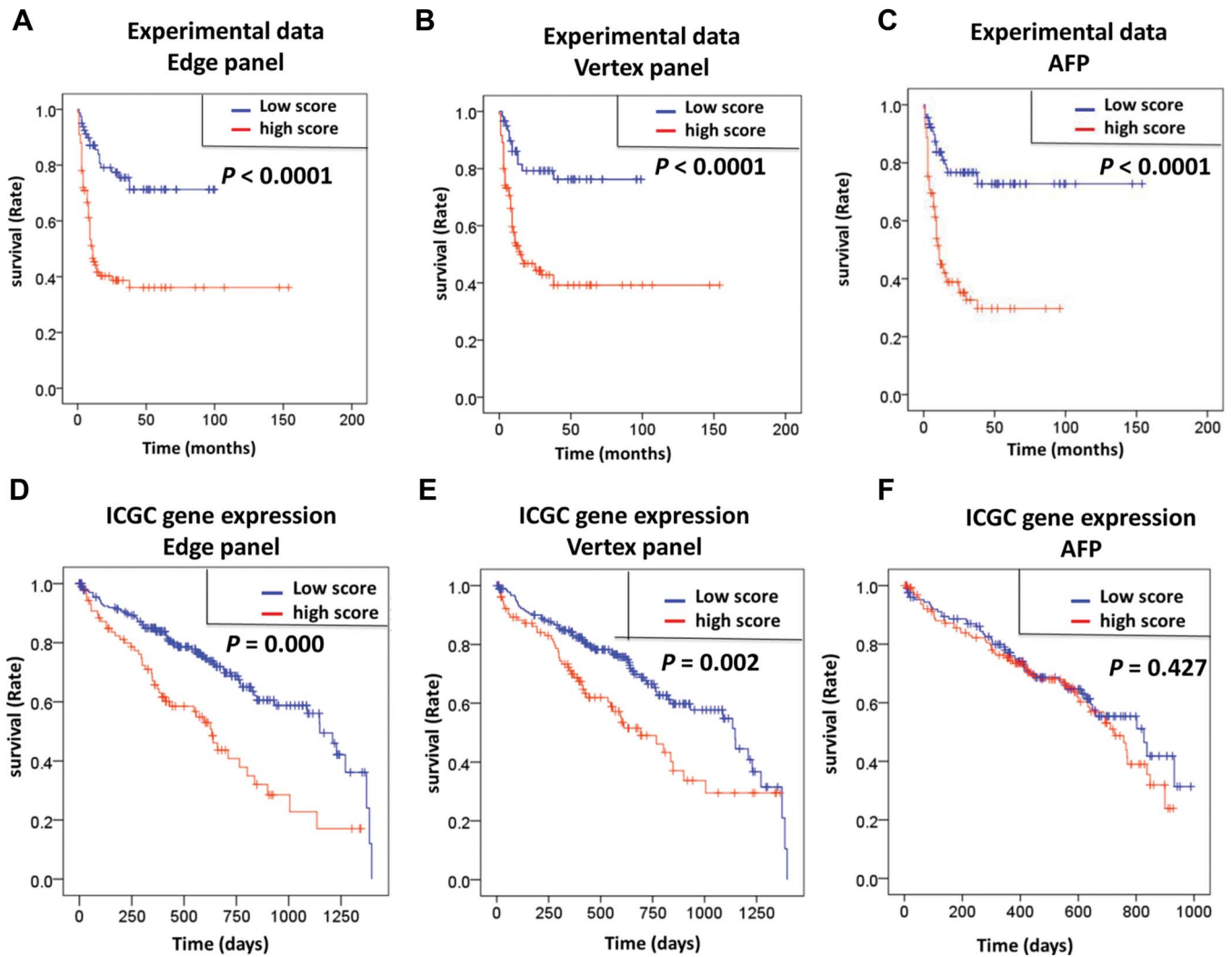
Figure 5 Kaplan–Meier plot analysis and log-rank test survival analysis. (**A–C**) Negative associations between the vertex panel, edge panel, and AFP with overall survival were noted for HCC patients (all $P < 0.001$, log-rank test) in an experimental data set. (**D–F**) Negative associations were found between the vertex panel and edge panel with overall survival (all $P < 0.001$, log-rank test) on our experimental data and independent ICGC data, but the AFP mRNA level did not show a prognostic capacity for HCC patients in the ICGC data set.

validated the expression profiles of these 8 candidate molecules across all stages of disease progression in other independent samples of the clinical cohort, which includes 30 healthy donors as well as 30 HBVC, 30 CHB, 30 cirrhosis, and 150 HCC samples. Finally, four out of the eight genes (*COL5A1*, *CDK4*, *MMP2*, and *HLA-DQB1*) were used for the further diagnosis and prognosis of HBV-associated HCC due to their stage-characterized signatures and remarkable consistency in expression levels in other independent samples (Figure 4A; Supplementary Figure S3).

The edge-based biomarkers achieved better performance in diagnosis

Traditionally, molecular biomarkers (i.e. node-based biomarkers) contribute to HCC diagnosis mainly according to individual and static information (e.g. differential expression). However, considering the complicated and personalized pathogenesis of

HBV-induced HCC, edge-based biomarkers with the interaction effects of individual molecules (e.g. pair-wise correlations of molecules) were expected to improve biomarker performance in terms of precision, sensitivity, and specificity. Here, we created predictive models via logistic regression based on individual gene expression values (vertex panel) and co-expressions of gene pairs (edge panel) (details shown in Methods).

According to the expression levels of the last four candidate biomarkers measured by quantitative RT-PCR in the independent samples, we found that the diagnostic performance of the edge panel was better than individual gene pairs, even for the typical AFP (Figure 4B; Supplementary Table S10) as with the vertex panel (Figure 4C; Supplementary Table S10). Obviously, the edge panel had its own advantages in HCC diagnosis in AFP-negative patients (Figure 4D; Supplementary Table S10). Meanwhile, these identified molecules with dynamic information

were expected to predict the critical transition to HCC for early diagnosis. Thus, we respectively used the vertex panel and the edge panel to distinguish cirrhosis from other stages during the progression of HBV-associated HCC and found that the edge panel showed better performances in indicating cirrhosis, which was considered as the critical stage of malignant transition to HCC (Figure 4E; Supplementary Table S10). To test the generality and superiority of our edge panel for HCC diagnosis, we collected another independent data set from the Gene Expression Omnibus (GEO) (GSE25097), which included 268 HCC patients, 40 cirrhosis patients, and 6 healthy samples, and we compared diagnostic performance of our edge panel (Figure 4F) with the traditional biomarker AFP (Figure 4G) and the new biomarkers GPC3 (Filmus and Capurro, 2013) (Figure 4H) and DKK1 (Shen et al., 2012) (Figure 4I) by receiver operating characteristic (ROC) curve analysis and found that our edge panel had higher discriminatory power (Figure 4F). These findings suggested that our edge panel might be useful in clinical practice based on blood testing, not only to non-invasively distinguish HBV-associated HCC and non-HCC but also to signal imminent carcinogenesis for early diagnosis.

The edge-based biomarkers can also be used for prognosis

Diagnostic biomarkers are also expected to be used for prognosis (e.g. AFP) (Trevisani et al., 2001; Qin and Tang, 2002). Here, we performed Kaplan–Meier and log-rank test survival analyses to assess the correlations of the vertex and edge panels with overall survival in HCC patients. According to the data from our clinical cohort, the results of the ROC analysis showed that values of the vertex and edge panels were significantly negatively associated with overall survival similar to AFP (all $P < 0.001$; Figure 5A–C), while the vertex and the edge panels had better performances in HCC prognosis than individual genes (Supplementary Figure S4) and gene pairs (Supplementary Figure S5). In addition, we validated our results by using other cohorts (including 298 HCC patients) from the International Cancer Genome Consortium (ICGC) and unexpectedly found that the vertex and edge panels could robustly contribute to HCC prognosis (Figure 5D and E); however, it seemed that AFP lost its power in HCC prognosis (Figure 5F). These results suggested that the vertex and edge panels consisting of the candidate molecules identified by our algorithm could also be a powerful and robust prognostic indicator for HBV-associated HCC.

Discussion

In this study, we broadly recruited 273 patients suffering from liver diseases (i.e. HBVC, CHB, cirrhosis, and HCC) and 33 healthy donors (Supplementary Table S1). Then, we profiled whole-genome expression levels of PBMCs from blood samples of the above cohort to identify molecular biomarkers for non-invasively diagnosing HBV-associated HCC. After GCGO analysis, 19 candidate genes were found to predict the critical transition stage (i.e. cirrhosis) before hepatocarcinogenesis based on the

3 criteria of DNBs (Figure 3). Then we narrowed the selection (Supplementary Figure S2 and Table S8) and introduced the edge-based biomarker, considering the complicated and personalized pathogenesis of HBV-induced HCC (Figure 1). We selected 4 core genes (*COL5A1*, *HLA-DQB1*, *MMP2*, and *CDK4*) from the 19 candidates to form the node/edge panel after performing literature mining and expression validation in independent cohorts (Figure 4A; Supplementary Figure S2). Interestingly, following our algorithm, most candidate signatures belonged to well-known cancer hallmarks (Hanahan and Weinberg, 2011).

Although the complex process of hepatocarcinogenesis is still not fully understood, several signal pathways have been identified as critical players in the pathophysiology of HCC, including the Wnt/β-catenin pathway, the p53 pathway, and so others (Aravalli et al., 2008). The alterations in these pathways may differ at the same time, which is probably the cause of the insufficient sensitivity of single biomarkers. The possibility to perform gene profiling in combination with systems biology approaches has led to a new era in biomarker development using multiple genes (Schütte et al., 2015). In our study, though *COL5A1*, *HLA-DQB1*, *CDK4*, and *MMP2* are related to different functions in complex pathways of hepatocarcinogenesis, they formed a promising biomarker panel with diagnostic and prognostic capacities.

For each gene in this panel, studies have indicated relationships with carcinogenesis. The *COL5A1* gene is a member of the clade B fibrillar collagen gene family and is involved in the regionalization of fibril-associated macromolecules that are necessary for tissue-specific regulation of the later fibril growth and matrix assembly stages (Wenstrup et al., 2004). Meanwhile, the type V collagen assembled by *COL5A1* may mediate interleukin products and further regulate immune responses, which play especially important roles in auto-immunity (Yamada et al., 2009; Sullivan et al., 2017). Another possible mechanism is *COL5A1*'s involvement in extracellular matrix (ECM) formation (An et al., 2015). *CLO5A1* has been verified to be correlated with several types of cancer (Chai et al., 2016; Zhao et al., 2016; Li et al., 2017a; Liu et al., 2018). *COL5A1* may play an oncogene role, as it can promote cell proliferation, invasion, and survival (Liu et al., 2018). In our study, *COL5A1* expression in cirrhosis was higher than in CHB, which indicated that *COL5A1* might play important roles in the initiation and activation of liver fibrosis as well as the regulation of immune responses (Figure 4). In most organs, the principle components of the ECM are collagens and numerous other proteins that make up the basement membrane. The MMP family has been implicated in the ECM degradation associated with tumor growth and angiogenesis, one of the earliest stages of tumor progression (Lamoreaux et al., 1998; Fang et al., 2000). *MMP2* has been studied as a potential diagnostic or prognostic biomarker of colorectal and ovarian cancers (Hilska et al., 2007; Périgny et al., 2008). Upregulated *MMP2* expression is often detected in solid tumor tissues and is associated with tumor metastasis in HCC (Tang et al., 2016). Interestingly, matrix metallopeptidase 2 encoded by *MMP2* can mediate type V collagen degradation (Veidal et al., 2012; Van Doren, 2015) and can also initiate an innate immune response

with activation of the pro-inflammatory pathways by regulating inflammatory cytokines and chemokines (Nissinen and Kahari, 2014). Overactive cyclin protein in cancer cells often leads to uncontrolled proliferation. Cyclin-dependent kinase 4, encoded by *CDK4*, might play key roles in the tumorigenesis (Wang et al., 2017) of a variety of cancers through deregulation of the CDK4/6−cyclin D−INK4−RB pathway (Graf et al., 2010). In addition, *CDK4* might control cell cycle progression by inducing and maintaining cytokine responsiveness in lymphocytes (Modiano et al., 2000). In our study, *CDK4* expression was significantly increased in the HCC stage (Figure 4). This result is consistent with previous studies, which also reported overexpression of *CDK4* in many tumor types (Lindberg et al., 2007; Poomsawat et al., 2010), suggesting that *CDK4* is a key factor in promoting the initiation and development of tumors. HLA class II molecules have a relevant role in the inflammatory response elicited by most infections. Human *HLA-DQB1* belongs to the HLA class II beta chain paralogs that are expressed in APCs and plays a central role in the immune system by presenting peptides derived from extracellular proteins, which has been recommended for the diagnosis and prognosis of HBV-associated liver diseases (Doganay et al., 2014; Liu et al., 2016b).

To summarize the discussion, the potential mechanisms by which *COL5A1*, *HLA-DQB1*, *CDK4*, and *MMP2* can predict disease stage could include the following: (i) these genes are closely connected to key mechanisms of hepatocarcinogenesis, including immunity (*HLA-DQB1*), ECM-induced tumor formation (*COL5A1* and *MMP2*), and cancer cell proliferation (*CDK4*); (ii) unlike the conventional method of biomarker hunting, these genes were selected from continuous disease stage data with both static and dynamic characteristics instead of comparing HCC vs healthy group only; and (iii) from the networks and dynamics perspectives, we integrated four genes as edge biomarkers that included high-dimensional information and could achieve accurate early diagnosis for HCC.

After determining the core biomarker genes that were significantly associated with the pathogenesis and progression of HBV-induced HCC, we introduced the edge-based method to build an edge panel by logistic regression with integrating interaction terms (e.g. co-expression) among signature molecules instead of a traditional model of diagnosis or prediction by using the vertex panel with a composite of expression levels or concentrations of individual molecules. As a remarkable advantage, edge-based biomarkers can predict phenotypes in an accurate manner, even though their individual genes may have no significantly different expression profiles. As expected, we found better performances of the edge panel in both diagnosis (Figure 4) and prognosis (Figure 5) for HCC than in the vertex panel and the typical biomarkers (e.g. AFP). In particular, the edge panel had a clear advantage for HCC diagnosis in patients without significant increases in AFP (Figure 4D), which implied that our edge-based biomarker could have wider applications in the clinic.

In summary, we recruited specific participants (including HBVC, CHB, cirrhosis, and HCC patients as well as healthy donors) to identify non-invasive biomarkers of HCC. The iden-

tified candidate genes at both the network and dynamic levels could not only characterize the progression of HBV-associated HCC but also identify cirrhosis as the critical transition stage before carcinogenesis according to the mathematical model of DNBs. Meanwhile, we introduced edge-based biomarkers with differential correlation/network information instead of traditional differential gene/protein expression information to build the prediction model, which demonstrated that the edge panel had better performances in both the diagnosis and prognosis in terms of both precision and specificity for HCC, especially for patients with AFP-negative HCC, than individual genes and gene pairs as well as the vertex panel. Thus, our study not only provided a novel edge-based biomarker for non-invasive and effective diagnosis of HBV-associated HCC for individual patients based on blood testing but also introduced a new way to integrate the interaction terms of individual molecules for clinical diagnosis and prognosis from the network and dynamics perspectives.

Materials and methods
Study design and participants

The study design is shown in Figure 1. First, we broadly recruited participants who satisfied the eligibility criteria (Supplementary Table S1) from the Shuguang Hospital, Shanghai University of Traditional Chinese Medicine (TCM), and the Qidong Liver Cancer Institute between August 2012 and December 2014. A total of 306 blood samples from these participants (including 33 healthy donors, 33 HBVC, 40 CHB, 40 cirrhosis, and 160 HCC patients) were collected (Supplementary Table S2). This study was approved by the Official Ethics Committee of the Shanghai University of TCM, and written informed consent was obtained from all participants. The healthy volunteers had no history of liver disease, no viral infections, and no other diseases. All patients were screened by the diagnostic standard as referred to the 'Chronic hepatitis B prevention and treatment guidelines (2010)' (Chinese Medical Association, 2011) (Supplementary Table S1). Tumor differentiation was graded using the Edmondson−Steiner grading system. Demographic and clinic-pathological characteristics were collected for all included participants. Trial registration—Identifier: NCT03189992. Registered on June 4, 2017. Retrospectively registered (www.clinicaltrials.gov).

In addition, we obtained a data set (GSE25097) from the GEO database (http://www.ncbi.nlm.nih.gov/geo/) and one HCC data set from the ICGC (https://icgc.org/) for our independent assessment of diagnostic or prognostic capacity. The HCC data set from ICGC provided by RIKEN (project code: LIRI-JP) includes 298 donors (Hudson et al., 2010).

RNA isolation

Leukocytes were isolated from the blood samples by Ficoll-optimized density gradient separation and were frozen at −80°C (Gertler et al., 2003). Total RNA was extracted using a 'two-step' protocol as described previously (Guo et al., 2012). Total RNA was extracted from leukocytes from whole blood using TRIzol®

Reagent (Invitrogen) and was frozen at −80°C. A quality control check was carried out using the NanoDrop ND-1000 (NanoDrop).

Microarray detection

We randomly selected blood samples from 3 healthy donors, 3 HBVC, 10 CHB, 10 cirrhosis, and 10 HCC patients from our clinical cohort to measure the transcriptomic profiles of PBMCs. Total RNA was extracted, and the biotinylated cDNAs were hybridized to the NimbleGen *Homo sapiens* 12x135K Array (Roche, CAT No. A6484-00-01). Raw data were extracted as pair files by the NimbleScan software (version 2.5); the data were considered robustly expressed if the signal-to-noise ratio was >2. The NimbleScan software uses a robust multi-array analysis algorithm that offers quantile normalization and background correction. Probe level and gene summary files were produced. All microarray data were deposited in the GEO (http://www.ncbi.nlm.nih.gov/geo/), accession number GSE114783. The gene summary files were imported into the Agilent Gene Spring Software (version 11.0) for further analysis. An expression signal cutoff level was set at 50.0 as the minimum number of falsely called probes. Only genes with a signal value above the cutoff level were used in the subsequent analyses.

Data processing

We selected DEGs according to a random variance model-corrected analysis of variance (ANOVA). An RVM *F*-test is commonly used to filter DEGs because it can effectively increase the degrees of freedom in cases with small sample sizes (Wright and Simon, 2003). These values were FDR-adjusted *P*-values and were considered significant if they were <0.05. All data analyses were performed using the statistical software R version 3.1.0 (http://www.R-project.org).

Gene profile clustering combining gene ontology analysis (GCGO)

In accordance with the different signal density change tendencies of genes under different situations, we identified a set of unique model expression tendencies. The raw expression values were converted into $\log_2$ ratios. We defined unique gene profile clusters by using cluster analysis of gene expression dynamics (Ramoni et al., 2002; Hömann, 2003) according to the time-series gene expression data. The significance of each profile pattern was estimated by Fisher's exact test and a multiple comparison test. Additionally, GO analysis was used for function detection for genes of each cluster with specific expression tendencies to uncover potential HCC-associated biological functions enriched by these identified genes.

Static network

To further select candidate signature genes with network information for distinguishing different stages during hepatocellular carcinogenesis, we constructed a static network with z-transformed PCCs of gene pairs to obtain the hub genes with high degrees (Barabási and Oltvai, 2004). We considered absolute PCCs > 0.9 as strong co-expressions (i.e. links).

Criteria of DNBs

DNBs were considered to predict the tipping points or critical transitions during disease progression, whose theoretical derivation was demonstrated in our previous work (Chen et al., 2012). Briefly, when a biological system approaches the tipping point or critical state just before the critical transition, the DNB (i.e. a dominant cluster of molecules) would appear and satisfy three criteria as a predictive signal:

(i) The molecules in the cluster are highly fluctuated in terms of their expression levels at the critical stage (i.e. the average coefficients of variation or standard deviations of DNB members are large).

(ii) The molecules in the cluster are highly correlated in terms of their expression levels (i.e. the average expression correlation among the DNB members in absolute values is high).

(iii) The molecules in the cluster are weakly correlated in terms of expression levels with other molecules (i.e. the average correlation between the molecules of the cluster and other molecule clusters in absolute values is decreased).

Thus, we have the following criterion index *CI* to quantify DNB molecules as well as the tipping point.

$$CI = CV_I \frac{PCC_I}{PCC_0},$$

where CV_I is the average standard deviation of the DNB members, PCC_I is the average PCC of the cluster of molecules in absolute values, and PCC_0 is the average PCC between the cluster of molecules and other molecules in absolute values.

Based on non-linear dynamic theory, when a biological system is near the critical stage or tipping point, a DNB cluster exists and satisfies the above three features. This system will undergo drastic deterioration after the critical stage and further develop to the late disease state (Figure 3A). Thus, the DNB method is able to provide early-warning signals of the critical transition by quantifying DNBs (Chen et al., 2012).

Predictive models by logistic regression

We created predictive models via logistic regression based on individual gene expression values (vertex panel) and co-expressions of gene pairs (edge panel) as follows.

(i) Vertex (gene) panel. In the logit model based on the four vertices (genes) combined with AFP,

logit (P = HCC) = −0.998 + 1.42 × *COL5A1* + 1.278 × *HLA-DQB1*−0.816 × *CDK4* + 0.051 × *MMP2* + 0.019 × AFP.

(ii) Edge (gene-pair) panel. After stepwise logistic regression selection, three edges (gene pairs), including all four genes, were enrolled in the edge panel. In the logit model based on the three edges combined with AFP, we have

logit (P = HCC) = −1.714-6.892 × (*COL5A1~CDK4*) + 2.088 × (*HLADQB1~CDK4*) +7.267 × (*CDK4~MMP2*) + 0.029 × AFP.

Real-time RT-PCR

RNA samples were qualified for further processing if the A260/A280 spectrophotometric ratio was between 1.8 and 2.1. A total of 1 mg of total RNA was transcribed into cDNA

in a 20 μl reaction volume. The primers used are listed in Supplementary Table S3. Each real-time RT-PCR in a final volume of 25 μl contained 2× SYBR Green Real-Time RT-PCR Master Mix, 0.4 μM primers, and 0.5 μl of template cDNA. The cycling conditions consisted of an initial single cycle of 5 min at 95°C, followed by 40 cycles of 30 sec at 95°C, 30 sec at 54°C, 15 sec at 72°C, and fluorescence acquisition at 83°C for 1 sec. The cDNA was synthesized using reverse transcriptase (TOYOBO), oligo (dT), and random primers with 5 μg of RNA from the same samples used in the microarray. The PCR amplifications were performed in duplicate for each sample. The gene expression levels were quantified relative to the expression of β-actin by employing an optimized comparative Ct (ΔCt) value method. The differences in gene expression levels between groups were compared using Student's *t*-test. A *P*-value of <0.05 was considered significant.

Statistical analyses

Differences in gene expression levels between groups were compared using the Mann–Whitney *U* test. A *P*-value of <0.05 was considered significant. A stepwise logistic regression model was used to combine diagnostic markers. The predicted probability of being diagnosed with HCC was used as a surrogate marker to construct the ROC curve. The area under the ROC curve (AUC) was used as an accuracy index to evaluate the diagnostic performance (Hanley and McNeil, 1983). We defined the cutoff value according to the best predictive values calculated by the ROC analysis, where the Youden's index (sensitivity + specificity − 1) was maximal. Survival probabilities were calculated using the Kaplan–Meier method, and differences in survival between two patient groups were determined by the log-rank test.

Functional analysis

Genes were subjected to GO analysis (www.geneontology.org/) to uncover their biological processes and molecular functions. Similarly, KEGG pathway analysis (www.genome.jp/kegg/) was used to identify significant pathways associated with the DEGs. We used Fisher's exact test and the χ^2 test to select significant GO terms or pathways with a threshold of significance defined by the *P*-value and FDR. Within the significant category of both the GO/pathway analyses, the enrichment Re was given as follows:

$$\mathrm{Re} = \left(n_f/n\right) / \left(N_f/N\right),$$

where n_f is the number of DEGs within the particular category, n is the total number of genes within the same category, N_f is the number of DEGs in the entire microarray, and N is the total number of genes in the microarray.

Conventional biomarker filtration analysis of the selected gene set was performed using IPA software (Ingenuity® Systems).

Supplementary material

Supplementary material is available at *Journal of Molecular Cell Biology* online.

References

An, F., Zhang, Z., Xia, M., et al. (2015). Subpath analysis of each subtype of head and neck cancer based on the regulatory relationship between miRNAs and biological pathways. Oncol. Rep. *34*, 1745–1754.

Aravalli, R.N., Steer, C.J., and Cressman, E.N. (2008). Molecular mechanisms of hepatocellular carcinoma. Hepatology *48*, 2047–2063.

Baine, M.J., Chakraborty, S., Smith, L.M., et al. (2011). Transcriptional profiling of peripheral blood mononuclear cells in pancreatic cancer patients identifies novel genes with potential diagnostic utility. PLoS One *6*, e17014.

Barabási, A.L., and Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. Nat. Rev. Genet. *5*, 101–113.

Bray, F., Ferlay, J., Soerjomataram, I., et al. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. *68*, 394–424.

Chai, F., Liang, Y., Zhang, F., et al. (2016). Systematically identify key genes in inflammatory and non-inflammatory breast cancer. Gene *575*, 600–614.

Chen, L., Liu, R., Liu, Z.P., et al. (2012). Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. Sci. Rep. *2*, 342.

Chinese Medical Association. (2011). Chronic hepatitis B prevention and treatment guidelines (2010). Chin. J. Integr. Tradit. West Med. Liver Dis. *21*,121–122.

Debebe, A., Medina, V., Chen, C.Y., et al. (2017). Wnt/β-catenin activation and macrophage induction during liver cancer development following steatosis. Oncogene *36*, 6020–6029.

Ding, W., Xin, J., Jiang, L., et al. (2015). Characterisation of peripheral blood mononuclear cell microRNA in hepatitis B-related acute-on-chronic liver failure. Sci. Rep. *5*, 13098.

Doganay, L., Fejzullahu, A., Katrinli, S., et al. (2014). Association of human leukocyte antigen DQB1 and DRB1 alleles with chronic hepatitis B. World J. Gastroenterol. *20*, 8179–8186.

Elkington, P.T.G. (2006). Matrix metalloproteinases in destructive pulmonary pathology. Thorax *61*, 259–266.

Fang, J., Shing, Y., Wiederschain, D., et al. (2000). Matrix metalloproteinase-2 is required for the switch to the angiogenic phenotype in a tumor model. Proc. Natl Acad. Sci. USA *97*, 3884–3889.

Filmus, J., and Capurro, M. (2013). Glypican-3: a marker and a therapeutic target in hepatocellular carcinoma. FEBS J. *280*, 2471–2476.

Ganapathi, S.K., Beggs, A.D., Hodgson, S.V., et al. (2014). Expression and DNA methylation of TNF, IFNG and FOXP3 in colorectal cancer and their prognostic significance. Br. J. Cancer *111*, 1581–1589.

Gertler, R., Rosenberg, R., Fuehrer, K., et al. (2003). Detection of circulating tumor cells in blood using an optimized density gradient centrifugation. Recent Results Cancer Res. *162*, 149–155.

Graf, F., Mosch, B., Koehler, L., et al. (2010). Cyclin-dependent kinase 4/6 (cdk4/6) inhibitors: perspectives in cancer therapy and imaging. Mini Rev. Med. Chem. *10*, 527.

Guo, Z., Yu, S., Guan, Y., et al. (2012). Molecular mechanisms of same TCM syndrome for different diseases and different TCM syndrome for same disease in chronic hepatitis B and liver cirrhosis. Evid. Based Complement. Alternat. Med. *2012*, 120350.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646.

Hanley, J.A., and McNeil, B.J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology *148*, 839–843.

He, D., Liu, Z.-P., Honda, M., et al. (2012). Coexpression network analysis in chronic hepatitis B and C hepatic lesions reveals distinct patterns of disease progression to hepatocellular carcinoma. J Mol. Cell Biol. *4*, 140–152.

Hilska, M., Roberts, P.J., Collan, Y.U., et al. (2007). Prognostic significance of matrix metalloproteinases-1,-2,-7 and-13 and tissue inhibitors of metalloproteinases-1,-2,-3 and-4 in colorectal cancer. Int J. Cancer *121*, 714–723.

Honda, M., Sakai, Y., Yamashita, T., et al. (2010). Differential gene expression profiling in blood from patients with digestive system cancers. Biochem. Biophys. Res. Commun. *400*, 7–15.

Hörmann, A. (2003). Software telegram: CAGED v1.0: cluster analysis of gene expression dynamics. Comput. Stat. Data Anal. *42*, 268–268.

Hudson, T.J., Anderson, W., Artez, A., et al. (2010). International network of cancer genome projects. Nature *464*, 993–998.

Jiang, J.X., Yu, C., Li, Z.P., et al. (2016). Insights into significant pathways and gene interaction networks in peripheral blood mononuclear cells for early diagnosis of hepatocellular carcinoma. J. Cancer Res. Ther. *12*, 981.

Lamoreaux, W.J., Fitzgerald, M.E., Reiner, A., et al. (1998). Vascular endothelial growth factor increases release of gelatinase A and decreases release of tissue inhibitor of metalloproteinases by microvascular endothelial cells in vitro. Microvasc. Res. *55*, 29–42.

Lara-Pezzi, E., Majano, P.L., Yáñez-Mó, M., et al. (2001). Effect of the hepatitis B virus HBx protein on integrin-mediated adhesion to and migration on extracellular matrix. J. Hepatol. *34*, 409–415.

Lesterhuis, W.J., Bosco, A., Millward, M.J., et al. (2017). Dynamic versus static biomarkers in cancer immune checkpoint blockade: unravelling complexity. Nat. Rev. Drug Discov. *16*, 264–272.

Li, G., Li, X., Yang, M., et al. (2017a). Prediction of biomarkers of oral squamous cell carcinoma using microarray technology. Sci. Rep. *7*, 42105.

Li, M., Li, C., Liu, W.X., et al. (2017b). Dysfunction of PLA2G6 and CYP2C44-associated network signals imminent carcinogenesis from chronic inflammation to hepatocellular carcinoma. J. Mol. Cell Biol. *9*, 489–503.

Lindberg, D., Hessman, O., Akerstrom, G., et al. (2007). Cyclin-dependent kinase 4 (CDK4) expression in pancreatic endocrine tumors. Neuroendocrinology *86*, 112–118.

Liu, W., Wei, H., Gao, Z., et al. (2018). COL5A1 may contribute the metastasis of lung adenocarcinoma. Gene *665*, 57–66.

Liu, X., Wang, Y., Ji, H., et al. (2016a). Personalized characterization of diseases using sample-specific networks. Nucleic Acids Res. *44*, e164.

Liu, X., Yu, L., Han, C., et al. (2016b). Polymorphisms of HLA-DQB1 predict survival of hepatitis B virus-related hepatocellular carcinoma patients receiving hepatic resection. Clin. Res. Hepatol. Gastroenterol. *40*, 739–747.

Llovet, J.M., Montal, R., Sia, D., et al. (2018). Molecular therapies and precision medicine for hepatocellular carcinoma. Nat. Rev. Clin. Oncol. *15*, 599–616.

Marshall, K.W., Mohr, S., Khettabi, F.E., et al. (2010). A blood-based biomarker panel for stratifying current risk for colorectal cancer. Int. J. Cancer *126*, 1177–1186.

Marshall, A., Lukk, M., Kutter, C., et al. (2013). Global gene expression profiling reveals SPINK1 as a potential hepatocellular carcinoma marker. PLoS One *8*, e59459.

Modiano, J.F., Mayor, J., Ball, C., et al. (2000). CDK4 expression and activity are required for cytokine responsiveness in T cells. J. Immunol. *165*, 6693–6702.

Mojtahedi, M., Skupin, A., Zhou, J., et al. (2016). Cell fate decision as high-dimensional critical state transition. PLoS Biol. *14*, e2000640.

Nissinen, L., and Kähäri, V.M. (2014). Matrix metalloproteinases in inflammation. Biochim. Biophys. Acta *1840*, 2571–2580.

Peng, Y., Xiong, X., Tong, X., et al. (2016). Correlation of the content of hepatitis B core antigen in peripheral blood mononuclear cells with HBV virus load. Diagn. Microbiol. Infect. Dis. *85*, 154–158.

Périgny, M., Bairati, I., Harvey, I., et al. (2008). Role of immunohistochemical overexpression of matrix metalloproteinases MMP-2 and MMP-11 in the prognosis of death by ovarian cancer. Am. J. Clin. Pathol. *129*, 226–231.

Poomsawat, S., Buajeeb, W., Khovidhunkit, S.O., et al. (2010). Alteration in the expression of cdk4 and cdk6 proteins in oral cancer and premalignant lesions. J. Oral Pathol. Med. *39*, 793–799.

Qin, L.X., and Tang, Z.Y. (2002). The prognostic significance of clinical and pathological features in hepatocellular carcinoma. World J. Gastroenterol. *8*, 193–199.

Ramoni, M.F., Sebastiani, P., and Kohane, I.S. (2002). Cluster analysis of gene expression dynamics. Proc. Natl Acad. Sci. USA *99*, 9121.

Richard, A., Boullu, L., Herbach, U., et al. (2016). Single-cell-based analysis highlights a surge in cell-to-cell molecular variability preceding irreversible commitment in a differentiation process. PLoS Biol. *14*, e1002585.

Sa, R., Zhang, W., Ge, J., et al. (2016). Discovering a critical transition state from nonalcoholic hepatosteatosis to nonalcoholic steatohepatitis by lipidomics and dynamical network biomarkers. J. Mol. Cell Biol. *8*, 195.

Schütte, K., Schulz, C., Link, A., et al. (2015). Current biomarkers for hepatocellular carcinoma: surveillance, diagnosis and prediction of prognosis. World J. Hepatol. *7*, 139–149.

Shen, Q., Fan, J., Yang, X.-R., et al. (2012). Serum DKK1 as a protein biomarker for the diagnosis of hepatocellular carcinoma: a large-scale, multicentre study. Lancet Oncol. *13*, 817–826.

Shi, M., Chen, M.S., Sekar, K., et al. (2014). A blood-based three-gene signature for the non-invasive detection of early human hepatocellular carcinoma. Eur. J. Cancer *50*, 928–936.

Singal, A., Volk, M.L., Waljee, A., et al. (2009). Meta-analysis: surveillance with ultrasound for early-stage hepatocellular carcinoma in patients with cirrhosis. Aliment. Pharmacol. Ther. *30*, 37–47.

Sullivan, J.A., Jankowska-Gan, E., Hegde, S., et al. (2017). Th17 responses to collagen type V, kα1-tubulin, and vimentin are present early in human development and persist throughout life. Am. J. Transplant. *17*, 944–956.

Sung, J.J., Tsoi, K.K., Wong, V.W., et al. (2008). Meta-analysis: treatment of hepatitis B infection reduces risk of hepatocellular carcinoma. Aliment. Pharmacol. Ther. *28*, 1067.

Suzuki, E., Sugimoto, M., Kawaguchi, K., et al. (2019). Gene expression profile of peripheral blood mononuclear cells may contribute to the identification and immunological classification of breast cancer patients. Breast Cancer. *26*, 282–289

Tang, Y., Lv, P., Sun, Z., et al. (2016). 14-3-3β promotes migration and invasion of human hepatocellular carcinoma cells by modulating expression of MMP2 and MMP9 through PI3K/Akt/NF-κB pathway. PLoS One *11*, e0146070.

Tian, X.F., Ji, F.J., Zang, H.L., et al. (2016). Activation of the miR-34a/SIRT1/p53 signaling pathway contributes to the progress of liver fibrosis via inducing apoptosis in hepatocytes but not in HSCs. PLoS One *11*, e0158657.

Trevisani, F., D'Intino, P.E., Morsellilabate, A.M., et al. (2001). Serum alpha-fetoprotein for diagnosis of hepatocellular carcinoma in patients with chronic liver disease: influence of HBsAg and anti-HCV status. J. Hepatol. *34*, 570–575.

Van Doren, S.R. (2015). Matrix metalloproteinase interactions with collagen and elastin. Matrix Biol. *44–46*, 224–231.

Veidal, S.S., Larsen, D.V., Chen, X., et al. (2012). MMP mediated type V collagen degradation (C5M) is elevated in ankylosing spondylitis. Clin. Biochem. *45*, 541–546.

Waidely, E., Alyuobi, A.O., Bashammakh, A.S., et al. (2015). Serum protein biomarkers relevant to hepatocellular carcinoma and their detection. Analyst *141*, 36–44.

Wang, L., Zhang, R., You, X., et al. (2017). The steady-state level of CDK4 protein is regulated by antagonistic actions between PAQR4 and SKP2 and involved in tumorigenesis. J. Mol. Cell Biol. *9*, 409–421.

Wang, X.W., Hussain, S.P., Huo, T.I., et al. (2002). Molecular pathogenesis of human hepatocellular carcinoma. Toxicology *181*, 43–47.

Wenstrup, R.J., Florer JBBrunskill, E.W., Bell, S.M., et al. (2004). Type V collagen controls the initiation of collagen fibril assembly. J. Biol. Chem. *279*, 53331.

Wright, G.W., and Simon, R.M. (2003). A random variance model for detection of differential gene expression in small microarray experiments. Bioinformatics *19*, 2448–2455.

Yamada, Y., Sekine, Y., Yoshida, S., et al. (2009). Type V collagen-induced oral tolerance plus low-dose cyclosporine prevents rejection of MHC class I and II incompatible lung allografts. J. Immunol. *183*, 237–245.

Yang, B., Li, M., Tang, W., et al. (2018). Dynamic network biomarker indicates pulmonary metastasis at the tipping point of hepatocellular carcinoma. Nat. Commun. *9*, 678.

Yu, X., Zhang, J., Sun, S., et al. (2017). Individual-specific edge-network analysis for disease prediction. Nucleic Acids Res. *45*, e170.

Zeng, T., Zhang, W., Yu, X., et al. (2016). Big-data-based edge biomarkers: study on dynamical drug sensitivity and resistance in individuals. Brief. Bioinform. *17*, 576–592.

Zhang, W., Tao, Z., and Chen, L. (2014). EdgeMarker: identifying differentially correlated molecule pairs as edge-biomarkers. J. Theor. Biol. *362*, 35–43.

Zhao, X., Cai, H., Wang, X., et al. (2016). Discovery of signature genes in gastric cancer associated with prognosis. Neoplasma *53*, 239.

Zhou, H.X., Yang, M.X., Wang, Y., et al. (2016). Plasma LUNX mRNA, a non-invasive specific biomarker for diagnosis and prognostic prediction of non-small cell lung cancer. Am. J. Cancer Res. *6*, 452.