phase III SHARP trial Presidential Plennary Session. *Hepatology* 2008; **48**: 372A, Abstract no. 149.
26 Zhang Z, Zhou X, Shen H *et al.* Phosphorylated ERK is a potential predictor of sensitivity to sorafenib when treating hepatocellular carcinoma: evidence from an *in vitro* study. *BMC Med* 2009; **7**: 41.
27 Shao YY, Lin ZZ, Hsu C *et al.* Early alpha-fetoprotein response predicts treatment efficacy of antiangiogenic systemic therapy in patients with advanced hepatocellular carcinoma. *Cancer* 2010; **116**: 4590–6.
28 Okusaka T, Kasugai H, Ishii H. A randomized phase II trial of intra-arterial chemotherapy using a novel lipophilic platinum derivative (SM-11355) in comparison with zinostatin sitmalamer in patients with hepatocellular carcinoma ASCO Annual Meeting 2009. 2009. #4583 Poster session.
29 Okusaka T, Okada S, Nakanishi T *et al.* Phase II trial of intra-arterial chemotherapy using a novel lipophilic platinum derivative (SM-11355) in patients with hepatocellular carcinoma. *Invest New Drugs* 2004; **22**: 169–76.
30 Imai N, Ikeda K, Seko Y. Transcatheter arterial chemotherapy with miriplatin for patients with hepatocellular carcinoma and Child-Pugh B liver cirrhosis. *Kanzo* 2010; **51**: 758–60.
31 Hanada M, Takasu H, Kitaura M. Acquired resistance to miriplatin in rat hepatoma AH109A/MP10 is associated with increased Bcl-2 expression, leading to defects in inducing apoptosis. *Oncol Rep* 2010; **24**: 1011–8.
32 Fujiyama S, Shibata J, Maeda S *et al.* Phase I clinical study of a novel lipophilic platinum complex (SM-11355) in patients with hepatocellular carcinoma refractory to cisplatin/lipiodol. *Br J Cancer* 2003; **89**: 1614–9.
33 Yamashita Y, Takahashi M, Fujimura N *et al.* Clinical evaluation of hepatic artery embolization: comparison between Gelfoam and Lipiodol with anticancer agent. *Radiat Med* 1987; **5**: 61–7.
34 Ikeda K, Okusaka T, Ikeda M *et al.* [Transcatheter arterial chemoembolization with a lipophilic platinum complex SM-11355(miriplatin hydrate) – safety and efficacy in combination with embolizing agents]. *Gan to Kagaku Ryoho* 2010; **37**: 271–5.
35 Maeda M, Uchida N, Sasaki T. Liposoluble platinum(II) complexes with antitumor activity.(Japanese) Japanese Journal of. *Cancer Res* 1986; **77**: 523–5.

doi:10.1093/jmcb/mjs011
Published online March 31, 2012

Article

Coexpression network analysis in chronic hepatitis B and C hepatic lesions reveals distinct patterns of disease progression to hepatocellular carcinoma

Danning He[1,2,†], Zhi-Ping Liu[1,†,*], Masao Honda[3], Shuichi Kaneko[3], and Luonan Chen[1,*]

[1] Key Laboratory of Systems Biology, SIBS-Novo Nordisk Translational Research Centre for PreDiabetes, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China
[2] Department of Health Sciences Informatics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
[3] Department of Gastroenterology, Graduate School of Medical Science, Kanazawa University, Kanazawa, Ishikawa 920-8641, Japan
[†] These authors contributed equally to this work.
[*] Correspondence to: Luonan Chen, E-mail: lnchen@sibs.ac.cn; Zhi-Ping Liu, E-mail: zpliu@sibs.ac.cn

Chronic infections with the hepatitis B virus (HBV) and hepatitis C virus (HCV) are the major risks of hepatocellular carcinoma (HCC), and great efforts have been made towards the understanding of the different mechanisms that link the viral infection of hepatic lesions to HCC development. In this work, we developed a novel framework to identify distinct patterns of gene coexpression networks and inflammation-related modules from genome-scale microarray data upon viral infection, and further classified them into oncogenic and dysfunctional ones. The core of our framework lies in the comparative study on viral infection modules across different disease stages and disease types—the module preservation during disease progression is evaluated according to the change of network connectivity in different stages, while the similarity and difference in HBV and HCV are evaluated by comparing the overlap of gene compositions and functional annotations in HBV and HCV modules. In particular, we revealed two types of driving modules related to infection for carcinogenesis in HBV and HCV, respectively, i.e. pro-apoptosis modules that are oncogenic in HBV, and anti-apoptosis and inflammation modules that are oncogenic in HCV, which are in concordance with the results of previous differential expression-based approaches. Moreover, we found that intracellular protein transmembrane transportation and the transmembrane receptor protein tyrosine kinase signaling pathway act as oncogenic factors in HBV-HCC. Our findings provide novel insights into viral hepatocarcinogenesis and disease progression, and also demonstrate the advantages of an integrative and comparative network analysis over the existing differential expression-based approach and virus–host interactome-based approach.

Keywords: gene coexpression network, hepatitis B and C virus, hepatocellular carcinoma, disease progression, systems biology

Introduction

It has been estimated that chronic infections with the hepatitis B virus (HBV) and hepatitis C virus (HCV) account for up to 80% of hepatocellular carcinoma (HCC; Perz et al., 2006). Although chronic hepatitis caused by HBV and HCV is hardly distinguished by histological examination or clinical manifestations, the virological features of HBV and HCV are obviously different. HBV is a DNA virus that can be transported into the nucleus and integrated into the host DNA, thus directly transforming hepatocytes. In contrast, HCV is an RNA virus that replicates in the cytoplasm and is unable to integrate into the host genome (Tsai and Chung, 2010; Bouchard and Navas-Martin, 2011). Ever since the discovery of these two viruses, great efforts have been made towards the understanding of the molecular events and cellular signal transduction pathways that are altered by HBV and HCV infections (Iizuka et al., 2002; Honda et al., 2006; Mas et al., 2009; Ura et al., 2009), as well as the mechanisms that link HBV or HCV infections and hepatic lesions to HCC development (Wurmbach et al., 2007; Mas et al., 2009). Studies in this area include comparisons of microarray gene/microRNAs expression in HBV-HCC and HCV-HCC, identification of significantly differentially expressed genes/microRNAs under the two types of HCC, and analysis of functional annotations represented by them. It was reported that inflammation, anti-apoptosis, immune response, cell cycle and lipid metabolism were predominant in HCV, but pro-apoptosis, DNA damage and DNA repair response were predominant in HBV (Iizuka et al., 2002; Honda et al., 2006; Ura et al., 2009). There is also research (Wurmbach et al., 2007; Mas et al., 2009) focusing on a stepwise carcinogenic process from normal liver to HCV cirrhosis to HCV-HCC, or from preneoplastic lesions (cirrhosis and dysplasia) to HCV-HCC, and a positive trend was found in MHC class-I receptor activity, DNA damage checkpoint cell division and ubiquitin cycle genes

Received December 1, 2011. Revised February 21, 2012. Accepted March 4, 2012.

during this process (Mas et al., 2009). Although these efforts have suggested different oncogenic factors in HBV and HCV, as well as marker pathways during HCV-HCC progression, an integrative and comparative study of gene expression profiles in both HBV-HCC and HCV-HCC progression has yet to be conducted.

Network-based systems biology approaches (Liu et al., 2012) typically involve identification of groups of genes or network modules by microarray data analysis, whose expression levels are highly correlated across samples (Stuart et al., 2003; Zhang and Horvath, 2005; Oldham et al., 2008; Dewey et al., 2011). Such holistic approaches have several advantages over standard methods such as differential expression analysis, whose result is usually a list of genes, each of which is deemed significant in isolation (Chen et al., 2009, 2012). Actually, quantitative assessment of module preservation in different phenotypes using both gene expression and network connectivity as summation (Miller et al., 2010; Dewey et al., 2011) provides a new avenue in understanding of molecular differences that distinguish functional processes in disease progression (Oldham et al., 2008; Miller et al., 2010).

In this work, we developed a new framework to study the differences and similarities in HBV-HCC and HCV-HCC at a network level by an integrative and comparative analysis of weighted gene coexpression modules or networks in HBV-infected and HCV-infected liver tissues. We hypothesized that viral infection is an important stage or factor in carcinogenic progression (Tsai and Chung, 2010; Bouchard and Navas-Martin, 2011), and thus focused on the analysis of viral infection modules, e.g. oncogenic modules and dysfunctional modules. Using this approach, we identified distinct network modules of coexpressed genes with clear functional interpretations in HBV and HCV, as well as their implications of HCC development. We found that pro-apoptosis modules are oncogenic in HBV, but anti-apoptosis and inflammation modules are oncogenic in HCV, which is in concordance with previous differential expression-based approaches. Clearly, these modules are the driving force of carcinogenesis in HBV and HCV, respectively, which cannot be revealed by viral target analysis. In addition, we observed that intracellular protein transmembrane transportation and the transmembrane receptor protein tyrosine kinase signaling pathway were top enriched in HBV oncogenic modules, while a similar process of endosome to lysosome transport was observed in HCV dysfunctional modules. Those results are consistent with the existing knowledge that HCV enters hepatocytes via endocytosis (Bouchard and Navas-Martin, 2011). Although the entry mechanism of uncoated HBV into hepatocytes, and the transport of the viral genome into the nucleus of the host remain unclear (Seeger et al., 2007), the oncogenic modules identified by our approach show their important dysfunctions for HBV-HCC, and this can be a promising topic of future experimental research. Besides comparing the functional annotations of the top-ranked modules, we further identified the module overlap in HBV and HCV and found that the modules of HBV and HCV shared a significant overlap with each other. It implies that these subsets of genes are consistently coexpressed upon both HBV and HCV infection, but they result in the different network topologies and wiring that lead to contrasting functional performances. Last but not least, curating HBV/HCV protein targets (de Chassey et al., 2008; Wu et al., 2010) from literature research and combining them with our analysis result, we provided different viral targets as a potential root cause of these distinctions between HBV-HCC and HCV-HCC. Clearly, these new findings not only demonstrate the effectiveness of our network-based approach on analyzing the complex diseases, but also provide biological insights into viral hepatocarcinogenesis and disease progression.

Results

Overview of our framework

Figures 1 and 2 show the overview of our framework. Coexpression network reconstruction from high-throughput data are illustrated in Figure 1A. Module identification and functional analysis are summarized in Figure 1B, and module analysis for four types of viral infection modules is summarized in Figure 2. This paper focuses on the analysis of viral infection modules in disease progression. After we built gene coexpression networks for HBV and HCV, we identified their coexpresssion modules individually. After we validated their reproducibility in the independent datasets, we filtered out inflammation-related modules upon

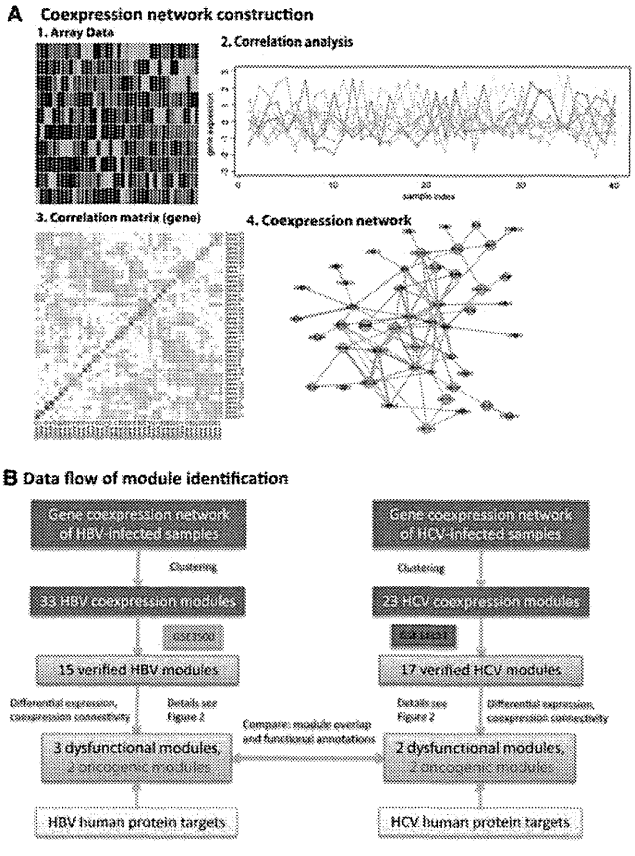
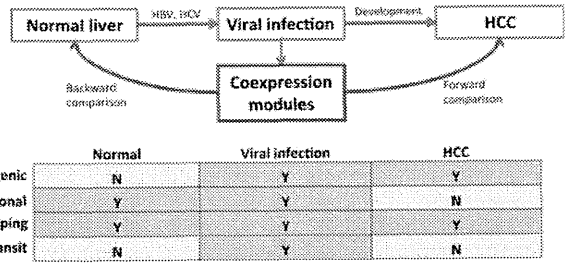


Figure 1 Overview of the framework. (**A**) Gene coexpression network reconstruction. (i) Microarray data filtering and prepossessing (rows correspond to samples and columns correspond to genes). (ii) Correlation analysis of individual genes expression across different samples. (iii) Construction of Pearson's correlation matrix and transformation into a matrix of connection strength. (iv) Coexpression network is established using hierarchical average linkage clustering (WGCNA). (**B**) Framework of module identification and analysis. The details of descriptions can be found in Materials and methods.



	Normal	Viral infection	HCC
Oncogenic	N	Y	Y
Dysfunctional	Y	Y	N
Housekeeping	Y	Y	Y
Transit	N	Y	N

Figure 2 Viral infection modules and their classification. This figure shows how to identify four types of viral infection modules (i.e. oncogenic, dysfunctional, housekeeping, and transit modules). The top subfigure shows the progression of HCC (i.e. from normal liver to viral infection and to HCC), and module comparison centered on viral infection or inflammation stage. The verified coexpression module of viral infection of HBV and HCV is classified into one type of 'oncogenic', 'dysfunctional', 'housekeeping', and 'transit' individually by backward and forward comparison for module preservation. 'Y' or 'N' represents its preservation 'yes' or 'no' in the three stages of disease progression, respectively. For example, one module ('Y' in viral infection) is identified to be 'oncogenic' when it is preserved in HCC ('Y'), but not in normal status ('N').

viral infections. The comparison of these modules in different disease stages for module preservation results in four types inflammation modules. And the comparison of oncogenic and dysfunctional modules in HBV and HCV provides evidence of the similarities and differences in the viral infections. We also tried to investigate their similarities and differences by analyzing the virus–host interactions of humans. The detailed descriptions of our framework are given in Materials and methods.

Constructing gene coexpression networks in HBV- and HCV-infected liver tissues

We set out to investigate the transcriptome upon viral infection and construct gene coexpression networks by applying weighted gene coexpression network analysis (WGCNA) (Zhang and Horvath, 2005). Our study was primarily based on Kanazawa data (Honda et al., 2006; Ivliev et al., 2010), which contains gene expression from 18 normal liver tissues (in normal stage), 36 HBV and 35 HCV-infected liver tissues (in viral-infected or inflammation stage), and different samples of 17 HBV-HCC and 17 HCV-HCC (in HCC stage). The other three datasets were mainly used for validation purposes. Two coexpression networks—one for HBV and the other for HCV—were constructed by calculating the pairwise Pearson's correlation coefficients of gene expressions in 36 HBV-infected samples and 35 HCV-infected samples, respectively. The information about datasets used in the study is shown in Supplementary Table S1. Briefly, the Pearson's correlation matrix for each coexpression network was transformed into a matrix of connection strengths using a power function (power = 6). These connection strengths were then used to calculate the topological overlap (TO), which considers not only the correlation of the two genes, but also the degree of their shared neighbors across the whole network.

Detecting gene coexpression modules in HBV- and HCV-infected liver tissues

Hierarchical average linkage clustering based on TO was used to group genes with highly similar coexpression patterns into modules (Ravasz et al., 2002). For computational reasons, we conducted the network module identification procedure in a blockwise manner with the same parameter setting for all networks. To summarize the scaled gene expression profiles for the identified modules, we used the first singular vector (module eigengene, ME), which is equivalent to the first principle component and explains the largest proportion of the variance of the module genes. We then used the MEs in a procedure to reassign genes to the modules which maximizes the module memberships (see Materials and methods for details). To this end, we identified 33 modules in HBV-infected liver tissues and 23 modules in HCV-infected liver tissues individually (Figure 3A and C), and each of them, containing coordinately expressed genes potentially participated in common cellular processes. The full list of module memberships is provided in Supplementary Table S2.

Identifying viral infection modules that are highly preserved across independent datasets

Because of the different number of gene expression samples and the wide range of coordinate gene regulations (Ivliev et al., 2010), we first validated the identified modules internally by a data-splitting technique in which 70% of the samples were used as a training set (see Materials and methods). After generating 100 such training sets, modules with significant co-clustering statistics (empirical $P < 0.05$) were retained for further validation (Figure 4).

Microarrays are inconsistent for differences in gene expression profiles across datasets and platforms (Wang et al., 2005). To gauge the consistency of our identified modules in independent datasets, two hepatitis virus-infected liver datasets, GSE3500 (Chen et al., 2002, 2004) and GSE14323 (Mas et al., 2009), were assembled. GSE3500 contains 10 samples of normal liver, 33 HBV-infected liver samples and 52 HBV-infected HCC. GSE14323 contains 19 samples of normal liver, 41 HCV-infected liver samples and 55 HCV-infected HCC. Detailed descriptions about these datasets are provided in Supplementary Table S1. We filtered and preprocessed the two datasets, and further identified gene coexpression modules from virus-infected status using the same procedure as described previously. Since the datasets contain different genes, we used the common genes shared by two datasets to compute the significance of the module overlap based on the hypergeometric test (Figure 3B and D). For HCV modules, 21 out of 23 of them have significant overlap ($P < 0.05$) with at least one module derived from GSE14323 providing confidence in the reproducibility of HCV gene coexpression modules. For HBV modules, however, 17 out of 33 of them have significant overlap with at least one module derived from GSE3500. Nevertheless, to ensure the reliability of our study, we identified interested modules that not only pass the internal validation, but also can be reproduced on independent datasets, which eventually resulted in 17 HCV modules and 15 HBV modules. We found that some most important modules—modules that will be classified as oncogenic and dysfunctional modules in the later sections—were not affected by such filtering. These modules represent sets of genes that are presented on and consistently coexpressed in diverse microarray platforms of viral infection.

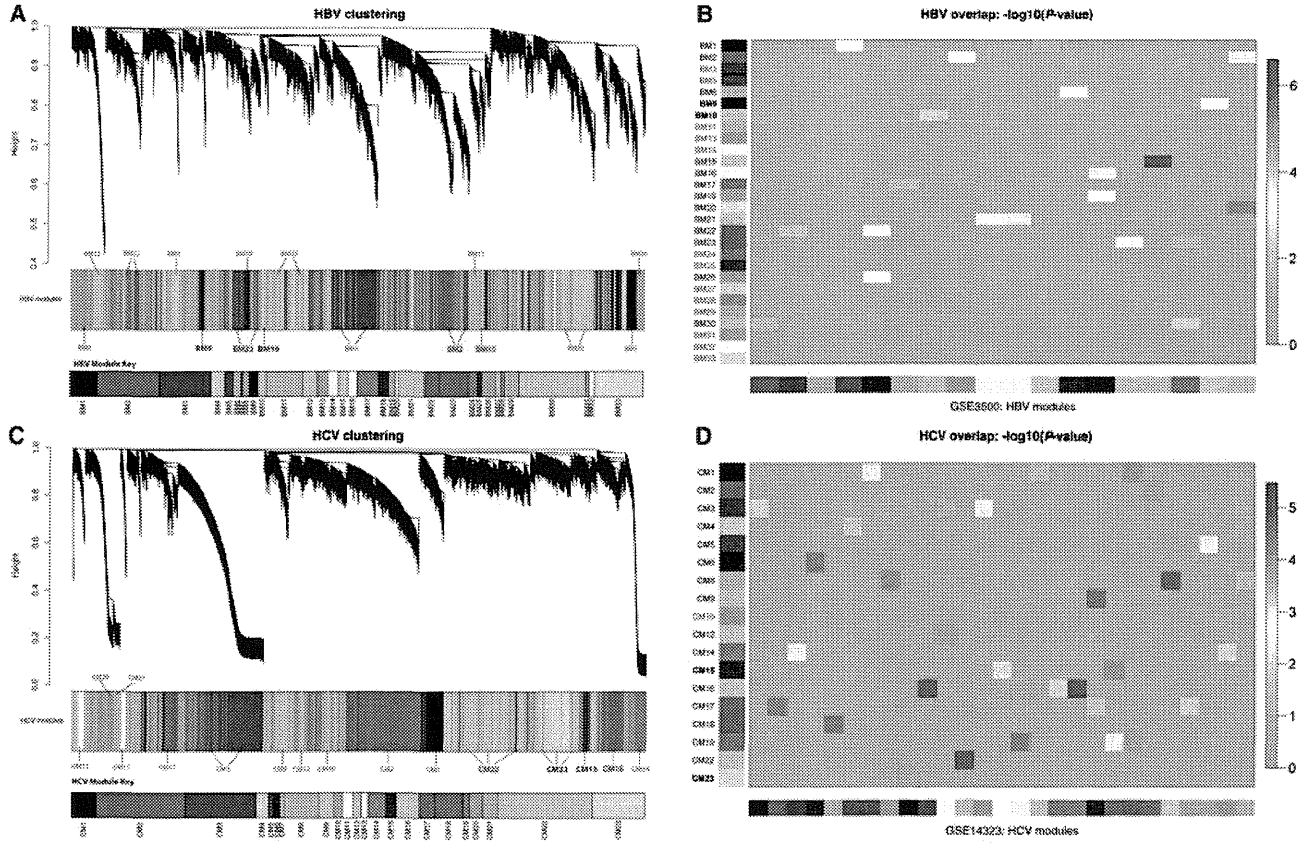


Figure 3 Identification of gene coexpression modules in 36 HBV- and 35 HCV-infected liver tissues and module reproducibility in independent datasets. Hierarchical average linkage clustering was applied to gene–gene adjacencies, which were defined on the basis of TO. Dynamic tree cut algorithm was applied to the dendrogram for module identification, and genes in the same branch can be assigned to different modules. The analysis identified 33 HBV modules (**A**) and 23 HCV modules (**C**) represented by different colors on the horizontal bar. Oncogenic modules (**A**: BM2, BM15, and BM23; **C**: CM18 and CM22) are marked in bold red font and dysfunctional modules (**A**: BM9 and BM10; **C**: CM15 and CM23) are marked in bold black font. In **B** and **D**, vertical modules were identified from our working dataset (Kanazawa data, corresponding to **A** and **C**, respectively), while horizontal modules were identified from independent dataset. Significance of pair-wise module-module overlap was based on Fisher's exact test *P*-values, using module assignment of the common genes shared by two datasets. (**B**) 21 out of 33 HBV modules have at least one significant ($P < 0.001$) overlapping modules in independent dataset (GSE3500). (**D**) 17 out of 23 HCV modules have at least one significant ($P < 0.001$) overlapping modules in independent dataset (GSE14323). Only these reproduced modules were kept for further analysis, and filtered module numbers are marked in grey.

We have validated the reproducibility of our identified gene coexpression modules in independent datasets, and further investigated whether these modules can be used to distinguish different stages of disease progression, reasoning that viral infection is an important transforming stage from normal to HCC (Tsai and Chung, 2010). MEs, i.e. the first singular vector of expressions in the module, were treated as the 'activity' and used to build classifiers for predicting the disease status given a test expression profile. For this purpose, MEs were used as feature values in a classifier based on svmRadial (Alexandros and David, 2006), and the technique of 5-fold cross validation was applied to select the optimal model that maximizes the area under the curve (AUC) of the receiver-operating characteristic. Once the optimal classifier was determined from one dataset, it was used to predict disease status for an independent dataset. Only the 15 HBV modules and 17 HCV modules that passed both internal and external validation were used for classification. Briefly, we trained classifiers on the working Kanazawa dataset and tested them on the validation one, and *vice versa*. To compute MEs on an independent dataset, we mapped gene compositions of each module to the independent dataset and calculated the first singular vector from the new gene expression profiles.

Our working Kanazawa dataset consists of various disease states in HCC progression: 18 normal, 36 HBV-infected, 35 HCV-infected, 17 HBV-HCC and 17 HCV-HCC (Supplementary Table S1). To examine the relationship among five categories of groups, i.e. normal, HBV-liver, HCV-liver, HBV-HCC, HCV-HCC, we built up five binary classifiers: normal and HBV-liver, HBV-liver and HBV-HCC, normal and HCV-liver, HCV-liver and HCV-HCC, HBV-HCC, and HCV-HCC. The final classification performance was defined as the AUC on one dataset using the classifier optimized from the other dataset (Figure 5 and Supplementary Figure S2). It was shown from Figure 5A and B

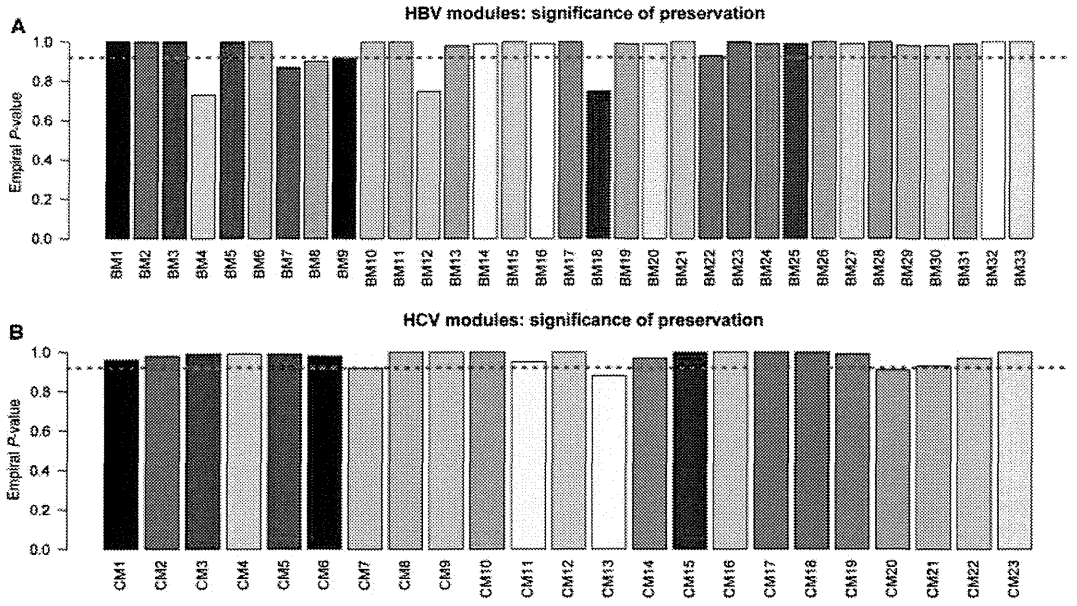


Figure 4 Internal validation of HBV (**A**) and HCV (**B**) modules. Each colored bar corresponds to a module. Red dash line indicates cutoff for statistical significance (empirical $P < 0.05$ or probability value > 0.95). Modules passing the cutoff line represent genes coexpressed in a wide range of samples while modules below the cutoff line represent genes coexpressed in only a subset of samples.

that gene coexpression modules identified from virus-infected status clearly distinguish expression profiles of normal and HCC. The results demonstrate the distinct module-gene expression profiles in different disease stages. However, the modules did not perform so well in classifying the two types of HCC, namely, HBV-HCC and HCV-HCC on the independent dataset GSE19665 (Deng et al., 2010; Supplementary Figure S2). One possibility is that the two types of HCC differ in the case of hepatocarcinogenesis, but they are rather similar at least in terms of the expression profile when cancer has already occurred. The other possibility is that the gene expression profile changes dramatically from viral infection to HCC, rendering it unsuitable to classify HCC types with these modules derived from the stage of viral infection.

Selecting oncogenic and dysfunctional modules related to inflammation

We have identified gene coexpression modules from HBV/HCV-infected liver tissues, i.e. in the viral infection or inflammation stage, validated their reproducibility in independent datasets, and we also discovered the distinct module expression profiles in the three stages of disease progression, i.e. normal, viral infection, and HCC, which could be used for phenotype classification in HBV and HCV, respectively. To focus on small subsets of modules which are most relevant to HCC, we investigated the dynamics of modules during disease progression and selected two types of modules, i.e. oncogenic and dysfunctional modules that are most likely to be related to HCC. As shown in Figure 2, we defined oncogenic and dysfunctional as follows. (i) 'Oncogenic': modules that are formed upon viral infection (i.e. they are disrupted in normal liver tissues) but are preserved in HCC, which represent inflammation-related oncogenic biological processes that are activated only upon viral infection. (ii) 'Dysfunctional': modules that are preserved in normal liver tissues but are disrupted in HCC, which represent tumor suppressive processes that remain effective upon viral infection. There are two more types of modules identified from viral-infected status. (iii) 'Housekeeping': those modules are preserved in both normal tissue and HCC. (iv) 'Transit': those modules are preserved in neither normal nor HCC. The housekeeping modules remain static during disease progression and are more likely to perform essential housekeeping functions, while the transit modules are more likely to be identified only in viral infection. They may be specifically responsive to the viral infection in this critical process and may indicate no disease progression characteristics of HCC. A graphical illustration of the four types of modules is shown in Figure 2. In order to determine which modules and their corresponding dysfunctional processes were activated upon viral infection, we defined two types of changes, i.e. the change in network topology which measures the gene–gene coexpression relationship and in the enrichment of differential expressed (DE) genes which measures the alternation of individual gene expression across phenotypes. We noticed that direct comparison of gene–gene correlation coexpression within modules between disease stages is unsuitable because the sample size in each stage varies. Therefore, we adopted a previously developed measure of the preservation density in intramodular connections between two networks (Dewey et al., 2011), and random permutation was run to assess their significance of preservation density (see Materials and methods). We defined modules with preservation density higher than 95% random permutations as significantly conserved and those with preservation density lower than 95% random permutations (empirical $P < 0.05$) as significantly disrupted (Figure 5C and D, and Table 1). To identify modules with significant differential

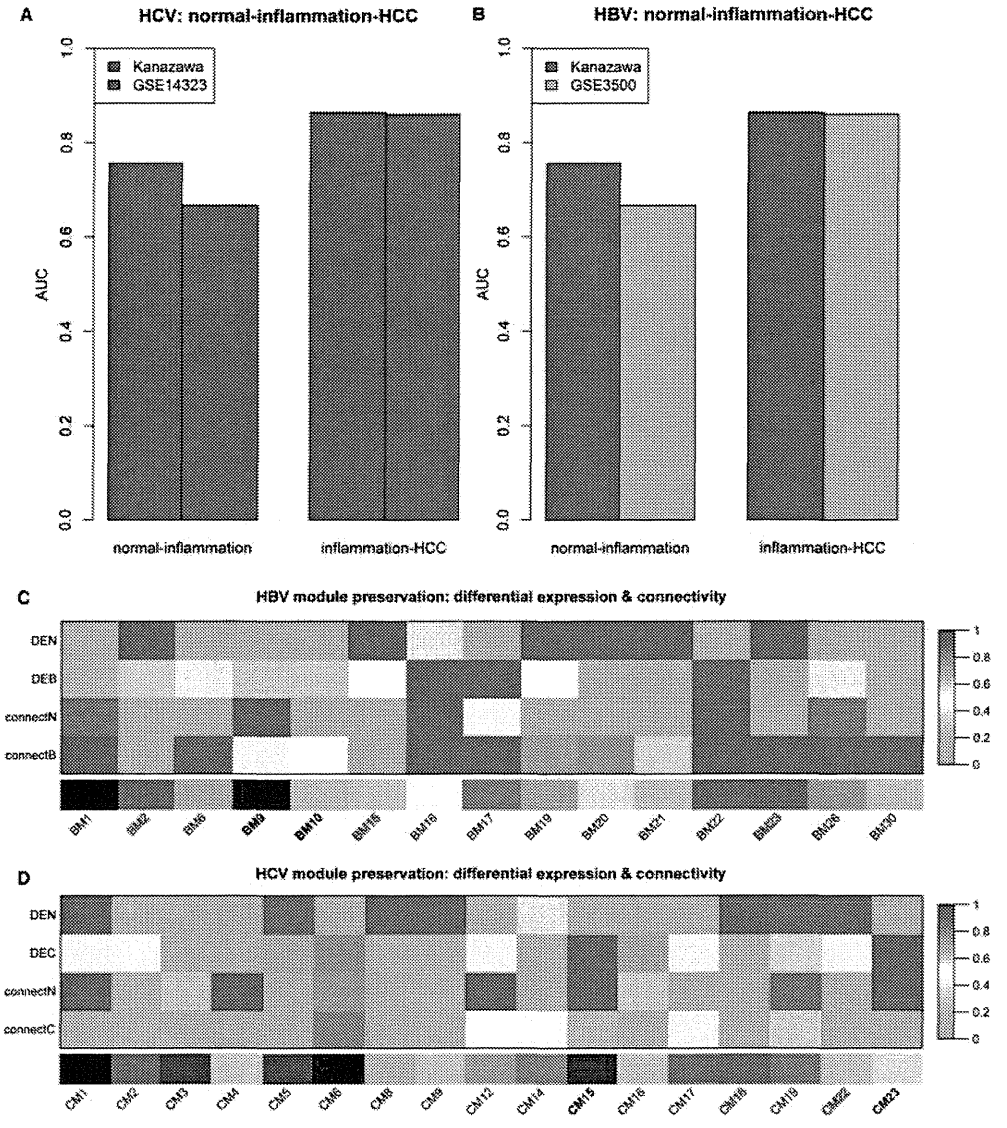


Figure 5 Phenotype classification results of the identified gene coexpression modules and preservation of viral infection modules in different disease stages. The coexpression modules identified from virus-infected inflammation status could distinguish status of normal and HCC (**A** and **B**), indicating the distinct expression profiles in three stages of disease progression, e.g. normal, virus-induced inflammation, and HCC. MEs of the reproduced modules were used as feature values, and svmRadial-based classifiers were trained in one dataset and evaluated in the other dataset, respectively. Preservation of viral infection modules in normal status and HCC (**C** and **D**) was evaluated in terms of differential expression (DEN: differential expression in normal vs HBV/HCV, DEB: differential expression in HBV vs HBV-HCC, DEC: differential expression in HCV vs HCV-HCC) and connectivity (connectN: correlation in normal vs HBV/HCV, connectB: correlation in HBV vs HBV-HCC, connectC: correlation in HCV vs HCV-HCC). The permutation-based score corresponds to the proportion of one thousand permutations in which random gene modules were more preserved (under-representation of differentially expressed genes or enrichment of conserved gene–gene coexpression relationship) than the derived modules. Therefore, red color (score > 0.95) corresponds to highly disrupted modules while green color (score $<$ 0.05) corresponds to highly conserved modules.

expression across phenotypes, we identified differentially expressed genes (adjusted $P < 0.05$), and measured the enrichment in the module using a permutation-based approach (see Materials and methods). The reported empirical *P*-value was equivalent to the proportion of random permutations in which random gene modules of the same size had a greater significance of DE than the module tested (Figure 5C and D, and Table 1). To this end, out of 15 HBV modules and 17 HCV modules, we identified 3 HBV modules and 2 HCV modules as oncogenic modules (italic type in Table 1), and 2 HBV modules and 2 HCV modules as dysfunctional modules (bold black in Table 1).

Comparison of selected HBV and HCV modules

Natural questions following module identification are (i) what are the similarities and differences between HBV and HCV modules? (ii) What are the dysfunctional implications for such similarities and differences for HCC? In this section, we analyzed the overlap between modules and enrichment of functional annotations to answer these questions.

Comparison of module overlap. First, comparisons of gene compositions of HBV and HCV modules based on the Fisher's

Table 1 Inflammation-related oncogenic (italic-type font) and dysfunctional (black font) modules, their top functional annotations and viral targets.

Virus	Cluster_index	Cluster_name	DE_normal_virus	DE_virus_HCC	Normal_virus	Virus_HCC	Category	Top functional annotations	Virus targets
HBV	BM2	***Blue***	**1**	0.324	0	0*	Oncogenic	Positive regulation of apoptosis	AIP,BHMT2,CHEK1,FETUB,HIF1A,MAPK9,MMP2,PTEN, PTGS2,RXRA,SDC4,SKP2,XBP1
	BM9	**darkred**	0*	0.314	0.436	**0.954**	Dysfunctional	Cell motion, positive regulation of apoptosis	PSMA7
	BM10	**darkturquoise**	0.04*	0.304	0.488	**0.96**	Dysfunctional	–	DNAJB1
	BM15	***lightgreen***	**1**	0.494	0.004	0.028*	Oncogenic	–	JAG1
	BM23	***Red***	**0.99**	0.014*	**0.974**	0.914	Oncogenic	Intracellular transport	–
HCV	CM15	**Midnight blue**	0.004*	**0.992**	0.216	0.192	Dysfunctional	Endosome to lysosome transport	H19,LZTS2,SRPX2
	CM18	***Red***	**1**	0.012*	0	0*	Oncogenic	Regulation of cell death	ANKRD12,FBN1,FXYD6,ITGB4,JAG2,JAK2, POU3F2,RUSC2,SSR4,TP53BP2,TP53BP2
	CM22	***Turquoise***	***1***	0.43	0	0*	Oncogenic	Positive regulation of transcription, negative regulation of apoptosis	C16orf7,C7,CANX,CANX,CTSB,FES,GRN,GSK3A,ITGAL,KRT18, LAMB2,NID2,NPM1,PFN1,PMVK,RAI14,SDC2,SERPINC1, SERPINF2,SFRP4,SLC31A2,SPOCK3,TAF1,VAPB,VAPB,VPS62,ZNF410
	CM23	**Yellow**	0*	**0.994**	0*	0.93	Dysfunctional	Positive regulation of cell proliferation, immune system development	ACP1,CENPC1,FKBP7,GPS2,HBXAP,LCK,LTBR,NCL,PIK3R1,SDCCAG8, SLC22A7,SRC,TAF11,UBE1C

Oncogenic modules are formed upon viral infection and preserved in HCC, dysfunctional modules are preserved in normal status but disrupted in HCC. Bold font corresponds to significant disruption (score > 0.95), and asterisk corresponds to significant preservation (score < 0.05). If a module has both significant disruption and preservation in the same stage of progression, only disruption is considered.

Table 2 Top enriched functional annotation clustering of HBV and HCV human protein targets

Virus	Cluster	Functional annotation	*P*-value	FDR
HBV	Cluster 1	hsa05200:Pathways in cancer	4.33E−27	4.98E−25
		hsa04115:p53 signaling pathway	6.91E−12	7.94E−10
		hsa04110:Cell cycle	5.09E−08	5.86E−06
	Cluster 2	hsa04920:Adipocytokine signaling pathway	1.12E−05	1.29E−03
		P00036:Interleukin signaling pathway	1.29E−05	7.35E−04
	Cluster 3	P00006:Apoptosis signaling pathway	6.19E−18	3.53E−16
		hsa04210:Apoptosis	2.01E−12	2.31E−10
HCV	Cluster 1	hsa04510:Focal adhesion	2.32E−08	2.74E−06
		REACT_13552:Integrin cell surface interactions	2.86E−07	1.52E−05
	Cluster 2	hsa04520:Adherens junction	1.64E−05	1.93E−03
	Cluster 3	hsa05200:Pathways in cancer	2.71E−08	3.19E−06
		P04398:p53 pathway feedback loops	3.99E−04	3.14E−02

exact test revealed several pairs of oncogenic and dysfunctional modules with a significant overlap ($P < 0.05$; Figure 6). Especially, we noticed that 3 HBV oncogenic modules (BM2, BM15, BM23) and 2 HCV oncogenic modules (CM18, CM22) have significant overlap with each other, e.g. BM2 with CM18 (Figure 7A and B), BM15 with CM18, and BM23 with CM22 (Figure 7C and D), representing the subsets of genes consistently coexpressed upon viral infection in both HCV- and HBV-infected status. We reasoned that it is these common subsets of genes that lead to carcinogenesis, and such genes can only be extracted by comparing the overlap between HBV and HCV modules. The documented HCC genes curated from literature (Wu et al., 2010) are marked as red in Figure 7. Although shared by overlapping modules, they occupy different network positions (intramodular connectivity, corresponding to the node size) and have different interacting partners (corresponding to their strongest first neighbors).

Comparison of functional enrichment. Secondly, common pathways of biological process were found in both HBV and HCV modules, which were associated with a wide range of functions that can be grouped into several categories: regulation of apoptosis, immune response, inflammation, cell cycle, cell migration, intracellular transport, signal transduction, and nitrogen compound catabolic process (Table 2). They represent general dysfunctional processes that are related to carcinogenesis, regardless of viral types. Distinct functional annotation clusters were also identified, which suggests the differences between HBV and HCV. A detailed functional enrichment of GO annotations in these modules is provided in the Supplementary Tables S3 (HBV modules) and S4 (HCV modules), and all GO terms mentioned in this section are highlighted in yellow background to facilitate search.

We are most interested in inflammation-related oncogenic modules, because they indicate the oncogenic processes that are directly activated by virus (these modules are recapitulated in HCC but not in normal liver tissues). The most contrasting distinction is that positive regulation of apoptosis (BM2, HBV, blue, 3.89E−6), programmed cell death (BM2, HBV, blue, 4.62E−5)

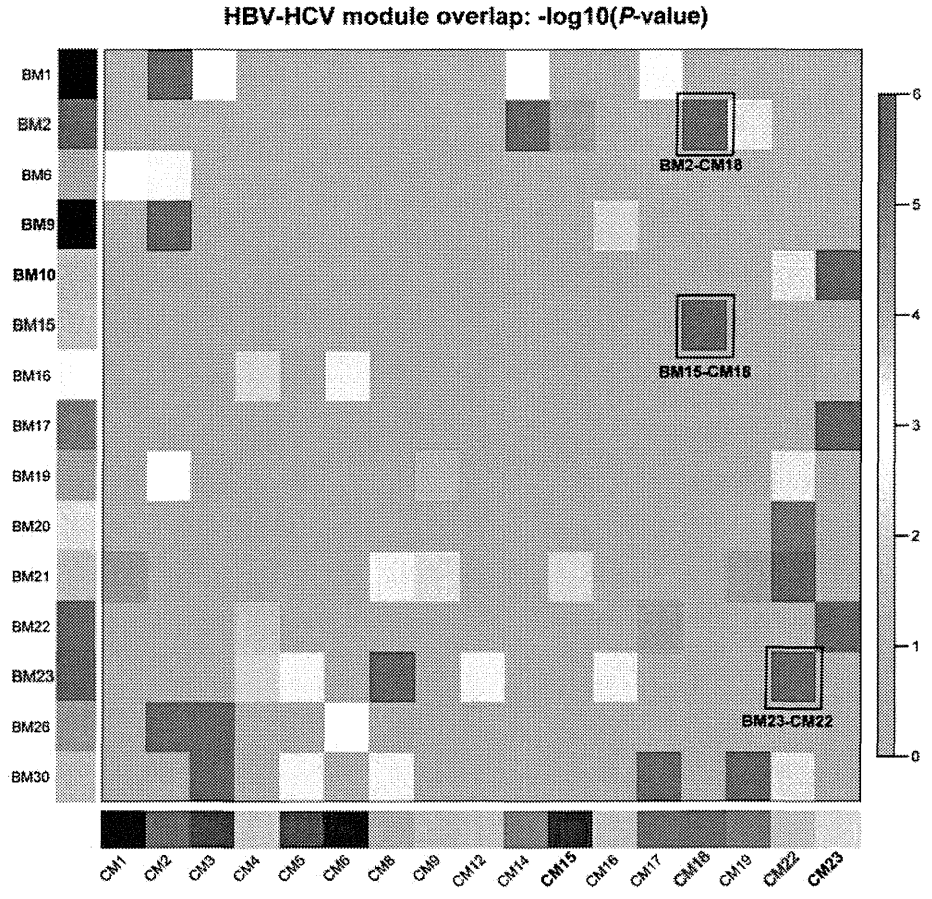


Figure 6 Overlap in gene compositions between HBV and HCV modules. Significance of pairwise module–module overlap was based on Fisher's exact test *P*-values. All HBV oncogenic modules (BM2, BM15, BM23) and HCV oncogenic modules (CM18, CM22) have significant overlap with each other, e.g. BM2 with CM18, BM15 with CM18, BM23 with CM22, representing smaller subsets of genes within modules that are consistently coexpressed in both HBV- and HCV-liver tissues. However, it is the different network properties and combinations of these subsets of genes that lead to the distinct functional annotations enriched in the corresponding HBV and HCV modules.

and cell death (BM2, HBV, blue, 5.05E−6) were top enriched in HBV infection-related modules whereas negative regulation of apoptosis (CM22, HCV, turquoise, 4.0E−6), programmed cell death (CM22, HCV, turquoise, 5.73E−6) and cell death (CM22, HCV, turquoise, 6.12E−6) were top enriched in HCV infection-related modules. The HCV oncogenic module was also top enriched in positive regulation of transcription (CM22, HCV, turquoise, 1.69E−6). This is in concordance with previous research findings that anti-apoptosis is predominant in HCV while pro-apoptosis is predominant in HBV, and that transcription regulation is activated in HCV (Honda et al., 2006). As is summarized previously (Bouchard and Navas-Martin, 2011), one of the mechanisms for HBV-induced HCC is the endless cycle of destruction of HBV-infected hepatocytes by immune cells and concomitant liver regeneration, during which a mutagenic environment is generated. In HCV-induced HCC, however, chronic inflammation that changes the microenvironment but does not lead to immediate death of infected hepatocytes plays the leading role. In fact, HCV core protein targets several tumor suppressor proteins (such as P53, P73, and pRb; Zhang and Horvath, 2005), and HCV nonstructural NS5A protein can block the cell death activity while promoting cell survival pathways by interacting with various cellular regulators (Lan et al., 2002; Chung et al., 2003).

We observed that intracellular transport (BM23, HBV, red, 5.37E−4), intracellular protein transmembrane transport (BM23, HBV, red, 9.02E−3), and transmembrane receptor protein tyrosine kinase signaling pathway (BM23, HBV, red, 1.20E−2) were top enriched in a HBV oncogenic module. Since cell surface receptor and intracellular signaling factors define the host range of HBV (Seeger et al., 2007), these processes can be related to the entry of uncoated HBV into hepatocytes. Interestingly, nucleocytoplasmic transport (BM23, HBV, red, 0.044) and nuclear transport (BM23, HBV, red, 0.047) are uniquely, although marginally, enriched in the HBV oncogenic module, which is consistent with the fact that HBV is able to transport its DNA genome into the nucleus (Rabe et al., 2009). For HCV, endosome to lysosome transport (CM15, midnightblue, 3.46E−3) and endosome transport (CM15, midnightblue, 5.95E−3) were top enriched in a dysfunctional module. Since endosome and lysosome are compartments of the endocytic membrane transport pathway, this is consistent with our existing knowledge that the whole body of HCV enters hepatocytes via endocytosis (Ashfaq et al., 2011). Compared with HCV, intracellular transport can play more important roles in carcinogenesis in HBV.

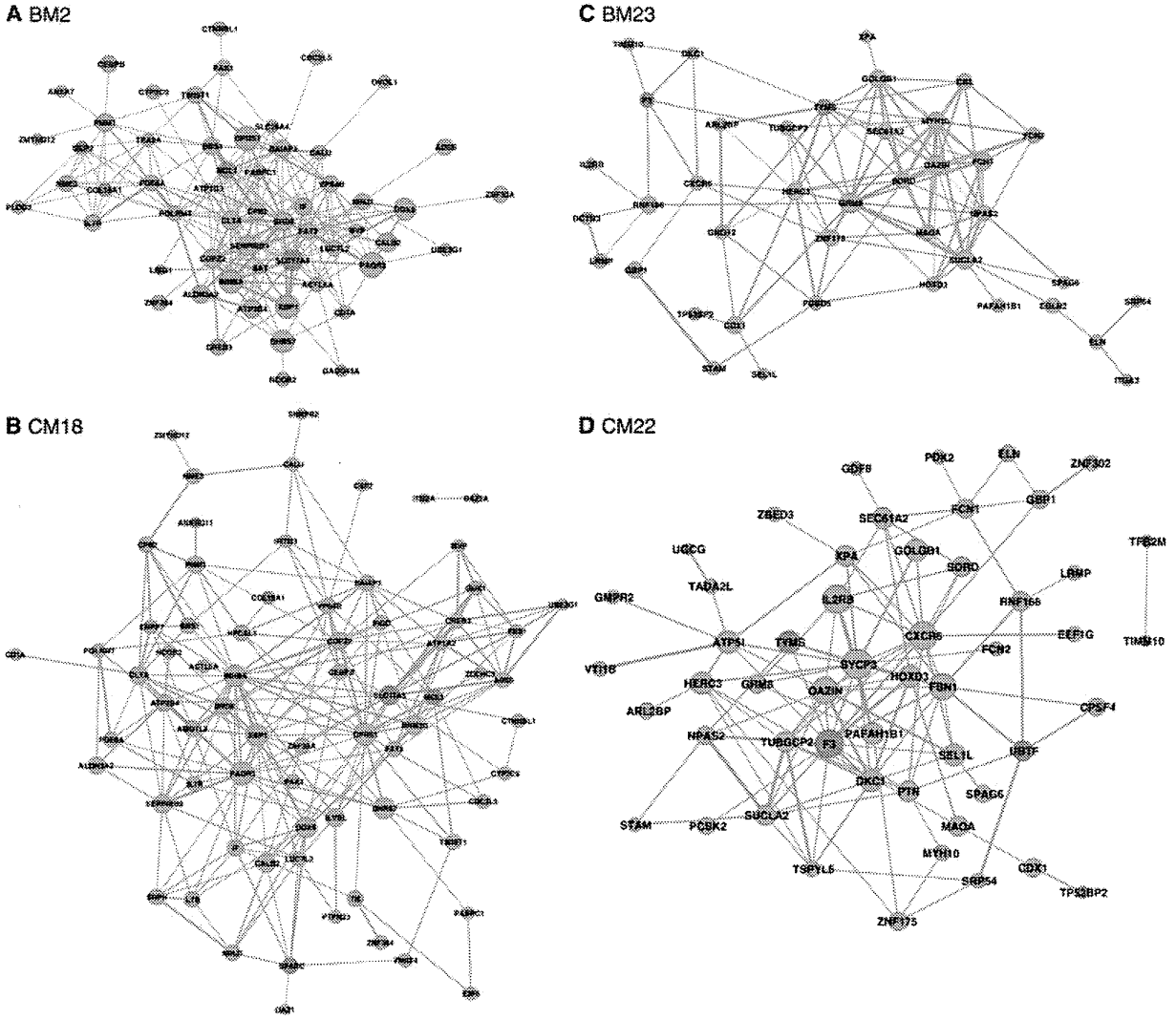


Figure 7 Topology of two pairs of overlapping HBV and HCV modules. For clarity, only the edges corresponding to top 5% correlations are shown. The node size corresponds to within-module connectivity, and the edge width corresponds to absolute value of correlation. Candidate HCC genes curated from literature are marked as red. For the BM2-CM18 pair, corresponds to the overlapping part in BM2 (**A**) and CM18 (**B**). For the BM23-CM22 pair, corresponds to the overlapping part in BM23 (**C**) and CM22 (**D**).

Another of our discoveries is that in both of the oncogenic HCV modules (namely, CM18, red, and CM22, turquoise), immune response (CM18, HCV, red, 9.55E−4) and inflammatory response (CM22, HCV, turquoise, 1.61E−5) were top enriched. Previous research reported that immune response and inflammatory phenotypes are predominant in HCV compared with HBV (Iizuka et al., 2002; Honda et al., 2006), and our result further suggested that compared with HBV-HCC, these two processes are more likely to be oncogenic for HCV-HCC. The HCV oncogenic module was uniquely enriched in lipid storage (CM22, HCV, turquoise, 0.0026) and previous findings also reported that lipid metabolism (Ura et al., 2009) is activated in HCV but not in HBV.

We also investigated the inflammation-related dysfunctional modules, because they represent tumor-suppressive processes that are disrupted upon cancer transformation. We observed that DNA damage response and signal transduction were uniquely enriched in the HBV dysfunctional module (BM9, HBV, darkred, 6.68E−3), and this is in concordance with previous research findings that DNA damage and signal transduction pathways are activated in HBV but not in HCV (Honda et al., 2006; Ura et al., 2009). In HCV dysfunctional modules, epithelial cell proliferation (CM23, HCV, yellow, 5.19E−5) was top enriched, which may be related to response to chronic inflammation upon HCV infection.

Different cellular processes and pathways represented by human protein targets of HBV and HCV

Beyond investigating the similarities and differences in these inflammation-related HBV and HCV modules in terms of gene compositions and functional annotations as well as dysfunctional implications, we attempted to provide a root cause analysis by exploring the human protein targets of HBV and HCV. Given that infection with HBV or HCV is one of the major risk factors contributing to HCC (Tsai and Chung, 2010), we considered whether it is the similarities and differences in viral targets of human proteins that explain the observed results. We constructed two interactome networks for human proteins interacting with HBV or HCV proteins (Supplementary Figure S1). The HCV interactome,

consisting of 11 HCV proteins and 481 human proteins, was generated from both Y2H assay and literature text-mining (de Chassey et al., 2008), and the HBV interactome, consisting of 5 HBV proteins and 250 human proteins, was generated from text-mining (Wu et al., 2010). We analyzed the pathway enrichment for each interactome to check whether HBV and HCV human protein targets correspond to distinct cellular pathways. A full list of enriched pathways and their gene compositions for HBV and HCV human protein targets is provided in Supplementary Tables S5 and S6, respectively. To analyze common and distinct cellular pathways represented by the two interactomes in a clear manner, we grouped annotations into clusters according to their semantic similarity (Kappa similarity threshold = 0.4) and ranked these functional annotation clusters (see Materials and methods for details). HBV human protein targets were found to be enriched in cancer pathways (rank 1, score = 5.08), inflammatory/immune pathways (rank 2, score = 3.1), and apoptosis signaling pathways (rank 3, score = 2.85). The HCV human protein targets were found to be enriched in cell surface interactions (rank 1, score = 5.03), and cancer pathways (rank 3, score = 1.51; Table 2). A detailed characterization of the functional annotation clusters is also provided in Supplementary Tables S5 and S6. Thus, we found that the cancer pathway is shared by two interactomes, but the HBV interactome is most enriched in apoptosis and the inflammatory/immune pathway while the HCV interactome is most enriched in cell surface interactions and the cell cycle.

The difference in annotated clusters between HBV and HCV interactome was confirmed by the distinct life cycles of HBV and HCV. HBV is non-cytopathic, and only its encapsulated DNA genome can be transported into the cell (Seeger et al., 2007). The virus-induced liver injury is associated with the influx of immune cells into the liver and the destruction of HBV-infected hepatocytes (Guidotti et al., 1999). Integration of viral DNA into the host genome can induce DNA recombination and damage (Bonilla Guerrero and Roberts, 2005). In contrast, HCV interacts with the host cell surface, and the whole virus is transported into the cell via receptor-mediated endocytosis (Blanchard et al., 2006). HCV is unable to reverse transcribe its RNA genome and thus unable to integrate into the host genome (Ashfaq et al., 2011). Our module-based approach not only re-addressed these aspects, but also identified pro-apoptosis and anti-apoptosis as the driving force of carcinogenesis in HBV and HCV, respectively, which cannot be revealed by viral target analysis.

Relating viral targets to the coexpression network, we are interested in protein targets, which belong to oncogenic modules. Although HBV and HCV viral targets have overlap, we found that none of the overlapping proteins belong to both HBV and HCV oncogenic modules (Table 1). In other words, HBV and HCV oncogenic modules each contain a disjoint set of target proteins. Supplementary Table S7 provides detailed information about human proteins targeted by HBV and HCV, their differential expression during disease progression and module memberships.

The KEGG database contains a pathway for Hepatitis C. Of the 134 genes contained in the Hepatitis C pathway, only 24 of them are direct targets of HCV. Functional annotation clusters showed that Hepatitis C is most enriched in the inflammatory/immune pathway (Table 2), which is different from HCV. One reason is that the upstream direct virus targets and their downstream response elements have different cellular functions. Another is that the Hepatitis C pathway is incomplete, and functions have not yet been attributed to all proteins. Besides comparing functional annotations of virus targets which represent initial perturbations, another powerful way to understand the different effects of HBV and HCV infections is to identify the response elements upon viral perturbations by analyzing gene expression profiles in our framework.

Discussion

We have conducted, to the best of our knowledge, the first comprehensive and comparative study of gene coexpression analysis at a network level to reveal the similarities and differences in HBV-HCC and HCV-HCC, in particular focusing on the inflammation-related analyses of viral infection. Our results demonstrate the advantages of a network-based systems biology approach over the previous differential expression approach and viral protein target-based approach. After validation by independent datasets, we identified 3 HBV and 2 HCV oncogenic modules, as well as 2 HBV and 2 HCV dysfunctional modules according to module preservation in normal livers and HCC. Those modules act as driving forces of carcinogenesis in HBV and HCV, respectively. The top enriched functional annotations of these modules are also in concordance with previous research and consistent with our existing knowledge of the distinct lifecycles of HBV and HCV in hepatocytes. In addition, the top enriched transmembrane transport and transmembrane receptor signaling pathway in one HBV oncogenic module suggested their potentially important roles in HBV-HCC.

Notably, our discoveries in distinct functional annotations represented by HBV and HCV modules could not have been revealed by existing standard methods such as differential expression and viral targets. First, we found no gene significantly differential expressed between HBV-infected and HCV-infected liver samples (Supplementary Table S8), rendering direct comparison of gene expression profiles in this status impossible. Second, we could not have selected those interesting inflammation-related modules and further classified them into four types without the use of module preservation in normal and HCC livers, which is also the advantage of our approach over other coexpression-based analysis of gene expression only in disease status, or disease vs control status (Ivliev et al., 2010). It should also be noted that if starting from modules in normal status, the viral infection oncogenic modules would be missed; and if starting from modules in HCC, the viral infection dysfunctional modules would be missed as well. Our theme is to identify these modules upon viral infection which we regarded as a process critical to HCC. Moreover, we narrowed down our analyses of these oncogenic and dysfunctional modules by considering all the combinatorial cases of module preservation in the three stages. The transit modules might particularly indicate the dysfunctional responses of virus infection, while we mainly focused on these repetitive modules in multiple disease progression stages. Third, we fully utilized the inherent variability in gene expression that exists in the same phenotype samples, and

further incorporated both the change of gene expression levels and the change of gene–gene coexpression relationships (i.e. connectivity) on the module level. By using a permutation-based approach, we eliminated the effect of different sample size between groups in the identification and comparison.

Materials and methods

Microarray data and workflow

Figure 1 shows the overview of our framework. A toy example of constructing gene coexpression network is illustrated in Figure 1A, our computational procedure is summarized in Figure 1B, and the module identification and classification of four types of modules are summarized in Figure 2. We analyzed four microarray datasets from independent studies, and a summary of the four datasets is described in Supplementary Table S1. The primary results were based on Kanazawa data (Honda et al., 2006; Ivliev et al., 2010). The other three datasets were mainly used for validation purposes. Both GSE14323 (Mas et al., 2009) and GSE19665 (Deng et al., 2010) were analyzed using the Affymetrix HG-U133A platform, and therefore, a probe set summary for each dataset was obtained using the RMA method in the affy package in R (Gautier et al., 2004). GSE3500 (Chen et al., 2002, 2004) was retrieved from the Stanford Microarray Database, using regression correlation. For GSE3500, samples and probe sets with >20% missing values were filtered, and the remaining missing values were imputed using impute package in R (Troyanskaya et al., 2001). When multiple probe sets were mapped to the same gene Entrez ID, the average expression vector was computed and used. Gene coexpression modules were identified from HBV- and HCV-infected liver samples, and validated on respective independent datasets. Only verified modules were used to analyze the dynamic change of modules during three stages of disease progression in HBV and HCV, respectively.

Weighted gene coexpression network construction and module identification

We built the weighted gene coexpression networks (Zhang and Horvath, 2005) for HBV and HCV by computing the gene correlation coexpression and inferring the coexpression networks in 36 HBV-infected samples and 35 HCV-infected samples, respectively. In a weighted gene coexpression network, the nodes represent genes and the edges represent the connection strength (adjacency), $a_{ij} = |\mathrm{cor}(x_i, x_j)|^{\beta}$, between the two gene expression profiles x_i and x_j. A major advantage of weighted networks is that the results are highly robust with regard to the choice of parameter β. Zhang and Horvath (2005) proposed a scale-free topology criterion for choosing β, and here we chose it to be six so that this yields approximately the same number of modules for HBV- and HCV-infected liver samples. The final adjacency was further transformed into a TO (Yip and Horvath, 2007). Then the modules were detected using the Dynamic Tree Cut algorithm (Langfelder et al., 2007; deep split = 2, cut height = 0.995, other parameters are defaulted).

As previously proposed, the module membership, k_{ME}, for each gene is defined as the Pearson's correlation between the expression level of the gene and the ME to which the gene belongs (Dong and Horvath, 2007). The k_{ME} for each gene was measured and the gene was assigned to the module which maximizes its k_{ME}. To avoid capturing weak associations, genes with $k_{ME} < 0.3$ for all of the MEs were assigned to none of them.

Functional annotation of gene sets and viral infection modules

Functional annotations of the gene sets and modules were performed on the basis of their gene composition using DAVID (http://david.abcc.ncifcrf.gov/). In DAVID, the reported *P*-values were derived from the EASE score probability, and a modified Fisher's exact test that is more conservative than the standard Fisher's exact test. 'BBID', 'BIOCARTA', 'KEGG_PATHWAY', 'PANTHER_PATHWAY', and 'REACTOME_PATHWAY' were selected for pathway enrichment analysis of viral protein targets. For characterization of modules, 'GO_BP_FAT' was selected. Due to the redundancy of annotations, similar or relevant annotations often appear repeatedly. We also adopted the functional annotation clustering provided in DAVID to help focus on biology in our study. We set the classification stringency to Medium, and clusters were ranked according to their *P*-values, which have exactly the same meaning as *P*-values for individual terms, and a false discovery rate (FDR) accompanying with each term was also reported.

Module internal validation and external validation

The purpose of internal validation is to rule out the possibility that some modules are based on gene coexpression across the full set of samples whereas others are the result of coordinate gene regulation in a subset of samples. Co-clustering statistics (Langfelder et al., 2011) is a cross-tabulation-based statistics for determining whether modules in the reference dataset are preserved in a test dataset. Reference modules are labeled $q = 1, 2, \ldots, Q^{[\mathrm{ref}]}$, test modules are labeled $q' = 1, 2, \ldots, Q^{[\mathrm{test}]}$, and the number of genes in module q or q' is denoted by $n^{(q)}$ or $n^{(q')}$. For HBV and HCV, respectively, we randomly chose 70% samples and identified modules, using the same procedure as described above, and iterated the random sampling and module identification process 100 times and generated 100 sets of test modules, $\{q'_i, i = 1, \ldots, 100\}$. For each set of modules, we computed its co-clustering statistics with reference modules, q (modules identified from full set of samples). The proportion of pairs of genes in both module q and module q' is given by $propCoClustering(q, q') = \binom{n_{qq'}}{2} / \binom{n^{(q)}}{2}$, where $n_{qq'}$ is defined as the number of genes that are both in the reference module q and in the test module q', the co-clustering statistics for module q is defined as the sum of the above proportions over all clusters q' in the test clustering, $coClustering(q) = \sum_{q'=1}^{Q^{[\mathrm{test}]}} propCoClustering(q, q')$. Then, a permutation test for 100 times was conducted for each test set of modules to determine whether the observed co-clustering statistics are significantly different from those expected by chance. Finally, we selected the reference modules with significant co-clustering statistics in 95% of the test sets.

To validate the reproducibility of modules in independent datasets, we identified coexpression modules in independent datasets (GSE14323 for HCV and GSE3500 for HBV individually) using the same procedures as described above and

extracting the common genes shared by two datasets, then we computed the significance of module overlap based on the Fisher's exact test using the common genes and their module memberships.

Module preservation in different disease stages

To capture both the dynamics of individual gene expression and the dynamics of gene–gene correlation coexpression relationship (Miller et al., 2010) between disease stages as shown in Figure 5, for each module we analyzed the enrichment of differentially expressed genes and the preservation of coexpression network topology using the permutation-based approach.

First, DE genes, $\{g_j\}$ (adjusted P-value < 0.05), were identified using the lmFit function provided in the R limma package (Smyth, 2004), and a t-score was assigned to each gene that quantified the significance of DE between phenotypes. A full list recording the significance of differential expression for each gene, normal vs viral infection and viral infection vs HCC, is provided in Supplementary Table S8. For each module, an average t-score was computed by dividing the sum of individual t-score by module size, $t\text{-score}(M_i) = \left(\sum_{g_j \in M_i} |t\text{-score}(g_j)|\right)/\text{size}(M_i)$. The significance of DE enrichment is given by the proportion of 1000 permutations in which random modules of the same size associated with a larger t-score than the reference module.

To evaluate the preservation of modules between two gene coexpression networks, N_l and N_m, constructed from samples of different size, we adopted a previous measure of intramodular connectivity preservation (Dewey et al., 2011). We first computed the intramodular connectivity (Dong and Horvath, 2007) vectors k^l and k^m, where $k = \left\{k_i : k_i = \sum_{j \neq i}^{i,j \in M} a_{ij}\right\}$, k_i is the intramodular connectivity of node i, a_{ij} is the adjacency, and nodes i and j belong to the same module M. Then for each module M_j, $M_j \in M$, we computed its intramodular connectivity preservation $\text{Pres}^{l,m}_{M_j} = \text{cor}(k^l_{i \in M_j}, k^m_{i \in M_j})$. Under the null hypothesis that the derived module, M_j, is preserved between N_l and N_m no better than modules derived from random clustering, we randomly permuted gene labels so that modules of the same size but random gene composition were generated. 1000 such permutations were performed, and the proportion of permutations in which $\text{Pres}^{l,m}_{M_{\text{rand}}} > \text{Pres}^{l,m}_{M_j}$ was used to evaluate the significance of test. Such a test was used to evaluate the preservation of modules in normal and HCC, for HBV- and HCV-liver samples, respectively. Since samples in viral-infected liver tissue consist of four stages of fibrosis, we also computed the enrichment of genes significantly correlated with fibrosis for each module using similar statistics as described above, where $\text{cor}(M_i) = \left(\sum_{g_j \in M_i} |\text{cor}(g_j)|\right)/\text{size}(M_i)$.

Supplementary material

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

Acknowledgements

We thank Dr Katsuhisa Horimoto of National Institute of Advanced Industrial Science and Technology, Japan for his kindly help.

Funding

This work was supported by the NSFC (Nos. 31100949, 91029301, 61134013 and 61072149), the Chief Scientist Program of Shanghai Institutes for Biological Sciences (SIBS), Chinese Academy of Sciences (CAS) (No. 2009CSP002), Shanghai NSF (No. 11ZR1443100) and the Knowledge Innovation Program of SIBS of CAS (No. 2011KIP203) and the SA-SIBS Scholarship Program. This research was also partially supported by the National Center for Mathematics and Interdisciplinary Sciences of CAS, Shanghai Pujiang Program, and the FIRST program from JSPS initiated by CSTP.

Conflict of interest: none declared.

References

Alexandros, K., and David, M. (2006). Support vector machines in R. J Stat. Softw. *15*, 9.

Ashfaq, U.A., Javed, T., Rehman, S., et al. (2011). An overview of HCV molecular biology, replication and immune responses. Virol. J *8*, 161.

Blanchard, E., Belouzard, S., Goueslain, L., et al. (2006). Hepatitis C virus entry depends on clathrin-mediated endocytosis. Virol. J. *80*, 6964–6972.

Bonilla Guerrero, R., and Roberts, L.R. (2005). The role of hepatitis B virus integrations in the pathogenesis of human hepatocellular carcinoma. J. Hepatol. *42*, 760–777.

Bouchard, M.J., and Navas-Martin, S. (2011). Hepatitis B and C virus hepatocarcinogenesis: lessons learned and future challenges. Cancer Lett. *305*, 123–143.

Chen, X., Cheung, S.T., So, S., et al. (2002). Gene expression patterns in human liver cancers. Mol. Biol. Cell *13*, 1929–1939.

Chen, X., Higgins, J., Cheung, S.T., et al. (2004). Novel endothelial cell markers in hepatocellular carcinoma. Mod. Pathol. *17*, 1198–1210.

Chen, L., Wang, R.S., and Zhang, X.S. (2009). Biomolecular Networks: Methods and Applications in Systems Biology. New Jersey: John Wiley & Sons, Inc.

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.

Chung, Y.L., Sheu, M.L., and Yen, S.H. (2003). Hepatitis C virus NS5A as a potential viral Bcl-2 homologue interacts with Bax and inhibits apoptosis in hepatocellular carcinoma. Int. J. Cancer *107*, 65–73.

de Chassey, B., Navratil, V., Tafforeau, L., et al. (2008). Hepatitis C virus infection protein network. Mol. Syst. Biol. *4*, 230.

Deng, Y.B., Nagae, G., Midorikawa, Y., et al. (2010). Identification of genes preferentially methylated in hepatitis C virus-related hepatocellular carcinoma. Cancer Sci. *101*, 1501–1510.

Dewey, F.E., Perez, M.V., Wheeler, M.T., et al. (2011). Gene coexpression network topology of cardiac development, hypertrophy, and failure. Circ. Cardiovasc. Genet. *4*, 26–35.

Dong, J., and Horvath, S. (2007). Understanding network concepts in modules. BMC Syst. Biol. *1*, 24.

Gautier, L., Cope, L., Bolstad, B.M., et al. (2004). affy—analysis of Affymetrix GeneChip data at the probe level. Bioinformatics *20*, 307–315.

Guidotti, L.G., Rochford, R., Chung, J., et al. (1999). Viral clearance without destruction of infected cells during acute HBV infection. Science *284*, 825–829.

Honda, M., Yamashita, T., Ueda, T., et al. (2006). Different signaling pathways in the livers of patients with chronic hepatitis B or chronic hepatitis C. Hepatology *44*, 1122–1138.

Iizuka, N., Oka, M., Mori, N., et al. (2002). Comparison of gene expression profiles between hepatitis B virus- and hepatitis C virus-infected hepatocellular carcinoma by oligonucleotide microarray data on the basis of a supervised learning method. Cancer Res. *62*, 3939–3944.

Ivliev, A.E., 't Hoen, P.A., and Sergeeva, M.G. (2010). Coexpression network analysis identifies transcriptional modules related to proastrocytic

differentiation and sprouty signaling in glioma. Cancer Res. *70*, 10060–10070.

Lan, K.H., Sheu, M.L., Hwang, S.J., et al. (2002). HCV NS6A interacts with p53 and inhibits p53-mediated apoptosis. Oncogene *21*, 4801–4811.

Langfelder, P., Zhang, B., and Horvath, S. (2007). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut library for R. Bioinformatics *24*, 719–720.

Langfelder, P., Luo, R., Oldham, M.C., et al. (2011). Is my network module preserved and reproducible? PLoS Comput. Biol. *7*, e1001057.

Liu, Z.P., Wang, Y., Zhang, X.S., et al. (2012). Network-based analysis of complex diseases. IET Syst. Biol. *6*, 22–33.

Mas, V.R., Maluf, D.G., Archer, K.J., et al. (2009). Genes involved in viral carcinogenesis and tumor initiation in hepatitis C virus-induced hepatocellular carcinoma. Mol. Med. *15*, 85–94.

Miller, J.A., Horvath, S., and Geschwind, D.H. (2010). Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. Proc. Natl Acad. Sci. USA *107*, 12698–12703.

Oldham, M.C., Konopka, G., Iwamoto, K., et al. (2008). Functional organization of the transcriptome in human brain. Nat. Neurosci. *11*, 1271–1282.

Perz, J.F., Armstrong, G.L., Farrington, L.A., et al. (2006). The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. J. Hepatol. *45*, 529–538.

Rabe, B., Delaleau, M., Bischof, A., et al. (2009). Nuclear entry of hepatitis B virus capsids involves disintegration to protein dimers followed by nuclear reassociation to capsids. PLoS Pathog. *5*, e1000563.

Ravasz, E., Somera, A.L., Mongru, D.A., et al. (2002). Hierarchical organization of modularity in metabolic networks. Science *297*, 1551–1555.

Seeger, C., Zoulim, F., and Mason, W.S. (2007). Fields Virology. Philadelphia: Lippincott Williams & Wilkins.

Smyth, G.K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. *3*, Article 3.

Stuart, J.M., Segal, E., Koller, D., et al. (2003). A gene-coexpression network for global discovery of conserved genetic modules. Science *202*, 249–255.

Troyanskaya, O., Cantor, M., Sherlock, G., et al. (2001). Missing value estimation methods for DNA microarrays. Bioinformatics *17*, 520–525.

Tsai, W.L., and Chung, R.T. (2010). Viral hepatocarcinogenesis. Oncogene *29*, 2309–2324.

Ura, S., Honda, M., Yamashita, T., et al. (2009). Differential microRNA expression between hepatitis B and hepatitis C leading disease progression to hepatocellular carcinoma. Hepatology *49*, 1098–1112.

Wang, H., He, X., Band, M., et al. (2005). A study of inter-lab and inter-platform agreement of DNA microarray data. BMC Genomics *6*, 71.

Wu, Z.J., Zhu, Y., Huang, D.R., et al. (2010). Constructing the HBV-human protein interaction network to understand the relationship between HBV and hepatocellular carcinoma. J. Exp. Clin. Cancer Res. *29*, 146.

Wurmbach, E., Chen, Y.B., Khitrov, G., et al. (2007). Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma. Hepatology *45*, 938–947.

Yip, A., and Horvath, S. (2007). Gene network interconnectedness and the generalized topological overlap measure. BMC Bioinf. *8*, 22.

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. *4*, Article 17.