Research Article

EASL EUROPEAN ASSOCIATION FOR THE STUDY OF THE LIVER | JOURNAL OF HEPATOLOGY

Sequential transcriptome analysis of human liver cancer indicates late stage acquisition of malignant traits

Jens U. Marquardt[1,2,†], Daekwan Seo[1,†], Jesper B. Andersen[1,†], Matthew C. Gillen[1], Myoung Soo Kim[4], Elizabeth A. Conner[1], Peter R. Galle[2], Valentina M. Factor[1], Young Nyun Park[3], Snorri S. Thorgeirsson[1,*]

[1]*Laboratory of Experimental Carcinogenesis (LEC), Center for Cancer Research, National Cancer Institute, NIH, United States;* [2]*Department of Medicine I, Johannes Gutenberg University, Mainz, Germany;* [3]*Department of Pathology, Yonsei University College of Medicine, Seoul, South Korea;* [4]*Department of Surgery, Yonsei University College of Medicine, Seoul, South Korea*

Background & Aims: Human hepatocarcinogenesis is as a multi-step process starting from dysplastic lesions to early carcinomas (eHCC) that ultimately progress to HCC (pHCC). However, the sequential molecular alterations driving malignant transformation of the pre-neoplastic lesions are not clearly defined. This lack of information represents a major challenge in the clinical management of patients at risk.

Methods: We applied next-generation transcriptome sequencing to tumor-free surrounding liver (n = 7), low- (n = 4) and high-grade (n = 9) dysplastic lesions, eHCC (n = 5) and pHCC (n = 3) from 8 HCC patients with hepatitis B infection. Integrative analyses of genetic and transcriptomic changes were performed to characterize the genomic alterations during hepatocarcinogenesis.

Results: We report that changes in transcriptomes of early lesions including eHCC were modest and surprisingly homogenous. Extensive genetic alterations and subsequent activation of prognostic adverse signaling pathways occurred only late during hepatocarcinogenesis and were centered on *TGFβ*, *WNT*, *NOTCH*, and *EMT*-related genes highlighting the molecular diversity of pHCC. We further identify IGFALS as a key genetic determinant preferentially down-regulated in pHCC.

Conclusions: Our results define new hallmarks in molecular stratification and therapy options for patients at risk for HCC, and merit larger prospective investigations to develop a modified clinical-decision making algorithm based on the individualized next-generation sequencing analyses.

Keywords: Hepatocarcinogenesis; RNA sequencing; Molecular pathogenesis.
Received 3 May 2013; received in revised form 30 September 2013; accepted 9 October 2013; available online 26 October 2013

* Corresponding author. Address: Laboratory of Experimental Carcinogenesis, Center for Cancer Research, National Cancer Institute, NIH, 37 Convent Drive, Room 4146, Bethesda, MD, United States. Tel.: +1 301 496 1935; fax: +1 301 496 0734.
E-mail address: snorri_thorgeirsson@nih.gov (S.S. Thorgeirsson).
† These authors contributed equally to the results.

Cancer

Introduction

Human hepatocarcinogenesis is a multi-step process starting from chronically altered hepatic microenvironment with dysplastic lesions to early carcinoma (eHCC) that ultimately progresses to HCC (pHCC) [1]. Four sequential morphological lesions characterized by distinctive clinicopathological features, i.e., dysplastic lesions, early well-differentiated HCC (eHCC) and progressed HCC (pHCC), are recognized during human hepatocarcinogenesis [2]. Dysplastic nodules (DN) are defined as hepatocyte-like lesions without unambiguous criteria of malignancy. Depending on the degree of dysplasia, they are divided into low (LGDN) and high grade (HGDN). However, diagnosis of pre-malignant lesions and discrimination from eHCC is often difficult, and, although dysplastic nodules are considered monoclonal in origin, a clear probability of malignant transformation could not be attributed to a specific pre-neoplastic lesion [3].

According to current clinical guidelines, advanced HCCs are diagnosed with the use of a single dynamic imaging technique with biopsies being optional and most often reserved for ambiguous cases [4]. Further, in clinical routine small lesions such as DN and potentially some of the more advanced HCCs are usually followed by ultrasound until further progression. These findings might account for recent observations that undiagnosed HCCs are found by histopathological examinations of explanted livers following liver transplantation. Therefore, preexisting lesions are likely contributing to the high rate of recurrent disease after putative curative treatment and limit the overall survival [4,5]. Therefore, definition of the sequential molecular events leading to HCC is urgently needed and represents a major challenge in the clinical management of patients at risk.

The main objective of this study was to apply next-generation mRNA sequencing (RNA-Seq) technology to comprehensively define changes in the transcriptome of liver cancer during the sequential evolution of pre-neoplastic lesions into HCC. We show here that changes in molecular profiles of dysplastic lesions and eHCC were small and quite uniform in contrast to a striking increase in both the extent and heterogeneity in pHCC at both mRNA and DNA levels. A massive deregulation of key oncogenic molecules, such as TGFβ1, MYC, PI3K/AKT, pro-metastatic/EMT

Table 1. Clinicopathological information and HCC marker expression.

Patient No.	Clinicopathological information						Marker expression				
	Sex	Age (yr)	Intrahepatic metastasis	Vascular invasion	Lesion	Size (cm)	GPC3	HSP70	GS	K19	AFP
1	M	68	No	No	LGDN	0.8 x 0.8	Neg	Neg	Neg	Neg	Neg
					HGDN	1.5 x 1.4	Neg	Neg	Neg	Neg	Neg
2	M	54	No	No	HGDN	0.8 x 0.8	Neg	Neg	Neg	Neg	Neg
					HGDN	0.8 x 0.7	Neg	Neg	Neg	Neg	Neg
					eHCC	1.2 x 1.0	Neg	Neg	Pos	Neg	Neg
					pHCC	3.0 x 2.0	Neg	Pos[a]	Neg	Neg	Neg
3	F	42	No	No	HGDN	1.0 x 0.9	Neg	Weak[a]	Neg	Neg	Neg
4	M	64	No	No	HGDN	1.5 x 1.3	Neg	Neg	Weak[a]	Neg	Neg
5	M	61	No	No	LGDN	1 x 0.8	Neg	Neg	Neg	Neg	Neg
					LGDN	0.8 x 0.8	Neg	Neg	Neg	Neg	Neg
					LGDN	1.2 x 1.1	Neg	Neg	Neg	Neg	Neg
					HGDN	1.0 x 1.0	Neg	Neg	Neg	Neg	Neg
					HGDN	1.3 x 0.9	Neg	Neg	Pos	Neg	Neg
					HGDN	1.0 x 1.1	Neg	Weak[a]	Pos	Neg	Neg
					eHCC	1.6 x 1.4	Weak[a]	Pos[a]	Pos	Neg	Neg
					pHCC	3.8 x 3.2	Weak[a]	Pos[a]	Pos	Neg	Neg
6	M	61	No	No	eHCC	2 x 1.9	Neg	Weak[a]	Neg	Neg	Neg
7	M	50	No	No	eHCC	1 x 1.3	Neg	Weak	Neg	Neg	Neg
8	F	60	No	No	HGDN	1.0 x 1.0	Neg	Neg	Neg	Neg	Neg
					eHCC	1.1 x 1.0	Pos	Pos[a]	Pos	Neg	Pos[a]
					pHCC	1.2 x 1.0	Pos	Pos[a]	Pos	Neg	Neg

[a]Focal positivity.
Neg, negative staining; Weak, weak staining; Pos, positive staining.

SNAIL and Twist, NOTCH, WNT/β-Catenin and MET, was observed in pHCC, suggesting that activation of prognostically adverse signaling pathways is a late event during hepatocarcinogenesis. Additionally, we identified a down-regulation of *IGFALS* as a common feature in a large proportion of HCC patients, indicating that *IGFALS* might be a useful diagnostic and/or therapeutic target.

Materials and methods

Samples

A total of 28 samples were collected, including 7 surrounding liver tissues, 4 low-grade dysplastic nodules (LGDN), 9 high-grade dysplastic nodules (HGDN), 5 early HCC (eHCC), and 3 progressed HCC (pHCC). The nodules were resected from explanted cirrhotic livers from 8 patients with chronic hepatitis B infection. All lesions were classified according to the criteria of "International Consensus Group for Hepatocellular Neoplasia" by two independent expert pathologists [2]. All procedures were approved by the local authorities and prior patient consent was obtained. Demographic and clinicopathological data of the patients can be found in Table 1.

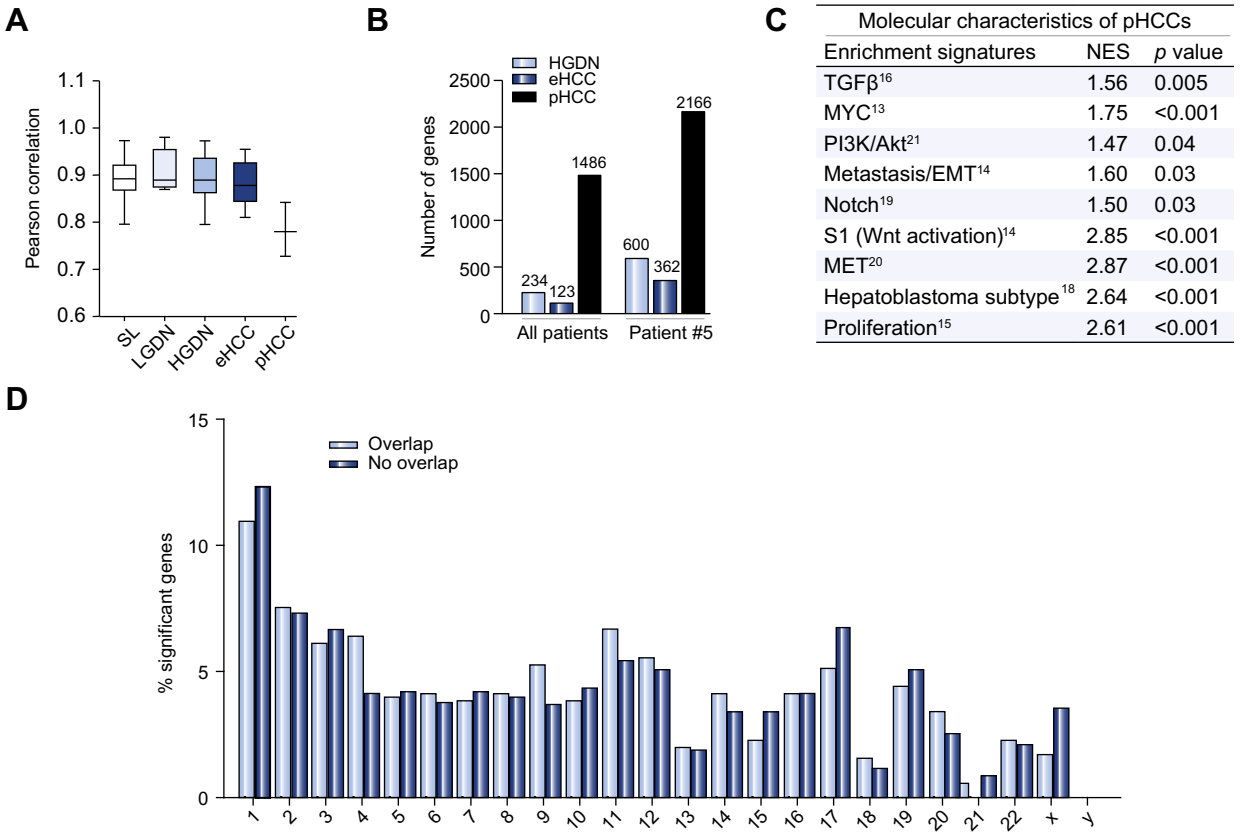
Next generation transcriptome sequencing

RNA isolation was performed using Qiagen RNEasy mini Kit (Qiagen GMBH, Hilden, Germany) according to the instructions provided by the manufacturer and quality assessment was performed using a Nanodrop ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE) and Agilent 2100 Bioanalyzer (Agilent, Palo Alto, CA). Each sample was sequenced in one sequencing lane with 107 bp paired-end sequencing in an Illumina GAII (Solexa) and 100 bp in an Illumina HiSeq2000 sequencer (SRA accession number: SRP030040). Illumina standard pipeline and software-CASAVA, GAII and HiSeq2000 were employed for processing of raw images, base calling and generating FASTQ sequence reads from paired-end RNA-sequencing data. The reads sequences were 107 bp (GAII) and 100 bp (HiSeq2000) and aligned to the GRCH37.p5 Primary Assembly and hg19 human reference genome in CLC Genomics Workbench-5.0.1 (CLC bio, Cambridge, MA) and bowtie-0.12.7, respectively [6]. Before aligning reads, we filtered low quality reads, reads containing adapter sequences, and duplicate mapping reads using Samtools [7], Picard (http://picard.sourceforge.net), and FASTX-Toolkit (FASTQ/A short-reads pre-processing tools, hannonlab.cshl.edu/fastx_toolkit). The abundance of the expression of a transcript was measured as the score of RPKM (Reads Per Kilobase of exon model per Million mapped reads) [8] in CLC Genomics Workbench. We applied Quartile Normalization to all 28 samples to reduce variations across samples. Among a total of 36,036 transcripts, transcripts (12,125) that had differently expressed with normal liver were used for further analysis by applying equivalence test. Analysis of differential expression was performed by an in house Bootstrap 1-way ANOVA script and Bootstrap *t* test. For the longitudinal analysis of PT5, we identified differently expression transcripts with DEGseq [9] and Chi-square test with Bonferroni adjusted *p* value. Cluster analysis was performed using Cluster and TreeView programs from the Michael Eisen Laboratory, Lawrence Berkeley National Laboratory and University of California, Berkeley; (http://rana.lbl.gov/eisen/). Functional classification and network analysis were performed using Ingenuity Pathway Analysis (Ingenuity Systems Inc.) and the GeneGo pathways analysis (Pathway Analysis MetaCore – GeneGo Inc., St. Joseph, MI) tools. Venn-Diagrams were generated using the VENNY software by JC. Oliveros (bioinfogp.cnb.csic.es/tools/venny/index.html). We identified known SNVs, novel SNVs, somatic SNVs, In/Del from Samtools [7], VarScan [10] and CLC Genomic Workbench. We also filtered and annotated genetic variants by using ANNOVAR software [11]. To minimize errors, we applied reads whose base quality score is greater than 20. We also filtered our SNVs against dbSNP135 and the 1000 genome project released April 2012. To establish a baseline reference, corresponding surrounding liver was used since no blood/skin tissue was available.

Results

RNA sequencing and baseline characteristics of the lesions

We used paired-end second-generation mRNA sequencing (RNA-Seq) to explore in detail the relationship and sequential evolution of pre-neoplastic lesions into HCC in 8 individuals with chronic hepatitis B (HBV) infection (Table 1; Supplementary Fig. 1).



Molecular characteristics of pHCCs		
Enrichment signatures	NES	*p* value
TGFβ[16]	1.56	0.005
MYC[13]	1.75	<0.001
PI3K/Akt[21]	1.47	0.04
Metastasis/EMT[14]	1.60	0.03
Notch[19]	1.50	0.03
S1 (Wnt activation)[14]	2.85	<0.001
MET[20]	2.87	<0.001
Hepatoblastoma subtype[18]	2.64	<0.001
Proliferation[15]	2.61	<0.001

Fig. 1. Molecular heterogeneity drives late stage activation of adverse signaling pathways. (A) Whole transcriptome correlations among the lesions for each morphological stage. *p* Values were calculated using a 1-way ANOVA. A significantly reduced correlation was demonstrated for pHCCs (*p* value <0.001). (B) Number of significant genes in the whole group comparison and in patient 5. (C) Activated gene sets in pHCC as determined by Gene Set Enrichment analyses (GSEA). Normalized Enrichment score (*NES*) reflects degree of over-representation for each group at the peak of the entire set. Statistical significance calculated by nominal *p* value of the *NES* by using an empirical phenotype-based permutation test. (D) Chromosomal distribution of SNVs overlapping and non-overlapping with the identified PT5 gene expression signature. Shown are the total numbers of affected genes for each chromosome.

RNA-Seq generated from 7,280,568 to 371,183,770 reads that were aligned to the human reference GRCh37.p5/hg19, representing an average genome-coverage of 105x (Supplementary Table 1).

To establish baseline morphological characteristics of every lesion, we assessed expression of well-known markers (*AFP*, *GPC3*, *GS*, and *HSP70*) associated with liver cancer by immunohistochemistry (Table 1; Supplementary Fig. 2). The majority of the pre-neoplastic samples were negative for all markers, whereas eHCC and pHCC showed heterogeneous expression patterns. None of the single markers successfully distinguished the pre-malignant from the more advanced lesions, supporting current clinical management guidelines, highlighting the need for predictive biomarkers in liver cancer [4].

Group comparison of the different stages in hepatocarcinogenesis

We next computed the whole transcriptome correlations for each lesion at the corresponding stage based on RPKM quantile normalized expression values. As anticipated, the gene expression profiles of surrounding livers (SL) were very homogeneous (Pearson correlation r = 0.89) (Fig. 1A). Unexpectedly, high homogeneity was also observed for LGDNs (r = 0.91), HGDNs (r = 0.89) and eHCCs (r = 0.88). However, it dramatically decreased upon progression to pHCC (r = 0.78; p <0.001) (Fig. 1A), reflecting a well-recognized phenotypic heterogeneity of advanced liver cancer.

The gene expression signatures were then defined for each type of lesions using a 1-way ANOVA and Bootstrapping *t* test approach. A total of 234 and 123 genes were specifically associated with progression from LGDN to HGDN and to eHCC, respectively, demonstrating limited transcriptomic changes and homogeneous nature of pre-neoplastic stages in hepatocarcinogenesis (Fig. 1B; Supplementary Table 2). Major functional networks in HGDNs were involved in oxidative stress, glutathione metabolism, and apoptosis with addition of cell cycle, immune response and stellate cell activation in eHCC (Supplementary Fig. 3A). Progression to pHCC was accompanied by a striking increase in the number of significant genes (1486 genes) (Fig. 1B) associated with malignancy and metastatic spread, including TGFβ, VEGF, and NOTCH/γ-secretase target genes involved in cell adhesion, cytoskeleton remodeling and epithelial-to-mesenchymal transition (EMT) (Supplementary Fig. 3A and B) [12].

Among the most significant molecular changes driving progression to pHCC was a clear enrichment of several published gene sets associated with poor outcome in HCC and dependent on activation of key oncogenic drivers including TGFβ1, MYC,

PI3K/AKT, pro-metastatic/EMT SNAIL and Twist, NOTCH, WNT/β-Catenin, and MET (Fig. 1C–F; Supplementary Figs. 4 and 5) [13–21].

To directly test the clinical significance of pHCC gene signature, we next applied it to a cohort of 53 human HCCs [22]. Hierarchical clustering analysis demonstrated a significant overlap with prognostic HCC subtypes (Supplementary Fig. 6A) [18]. Subsequent analyses of patient outcomes revealed a strong association with a shorter survival ($p = 0.009$) and increased tumor recurrence ($p = 0.003$) (Supplementary Fig. 6B–C). Further, a meta-analysis using publically available transcriptome data from more than 40 cancer entities confirmed a prognostic implication of the pHCC signature for cancers other than HCC (Supplementary Fig. 6D).

Horizontal analyses of malignant transformation in PT5

A clear advantage of this study was the acquisition of samples at the sequential stages of hepatocarcinogenesis. In particular, one patient (referred to as PT5) presented a complete sequence of lesions including SL, LGDN, HGDN, eHCC, and pHCC (Table 1; Supplementary Fig. 2). To gain a more in-depth understanding of the consecutive molecular stages, we performed an integrative analysis of whole transcriptomic changes and somatic mutations (i.e., single nucleotide variances (SNVs)) located within the exonic regions of significant transcripts. Supporting the group comparison data, the number of significant genes was relatively limited in HGDN (600 genes) and eHCC (362 genes) but greatly increased upon progression to pHCC (2166 genes) (Fig. 1B; Supplementary Table 3). The transcriptome analysis of LGDN to HGDN also revealed deregulation of signaling pathways involved in metabolism followed by activation of immune response in eHCC (Supplementary Fig. 7A). Consistently, the majority of networks activated during the conversion to pHCC conferred malignant and invasive properties, i.e., cell adhesion and EMT (Supplementary Fig. 7B–E).

Genetic landscape of mutations during liver cancer development

Next, we defined the landscape of somatic mutations in the lesions from PT5. First, the total number of exonic SNVs were assessed in the different lesions and integrated with our RNA seq data to identify SNVs with putative effect on gene expression (Supplementary Fig. 1). While the overlap of SNVs and significant genes in the early lesions was minimal (15 HGDN and 1 eHCC), among the 2166 significant genes in pHCC, 735 genes (29.5%) showed alterations both on mRNA and DNA levels (Supplementary Tables 3–5). The latter included key cancer-associated genes *NOTCH2, VCAM1, JAK1, ITGA6, ITGB1, IGF2, IGFBP5, XRCC5, DKK3*, and *MMP14*. Interestingly, while most of the overlapping and non-overlapping genes were equally distributed across chromosomes, an increased number of genes with overlapping SNVs were located on chromosomes 4, 9, 11, 14, and 20 (Fig. 1D). Conversely, genes located on chromosome 1, 17, 19, and X seem to be less affected by genetic alterations suggesting another mode of activation.

Our analysis also identified previously unrecognized SNVs. The number of unique and novel exonic SNVs was low in LGDN

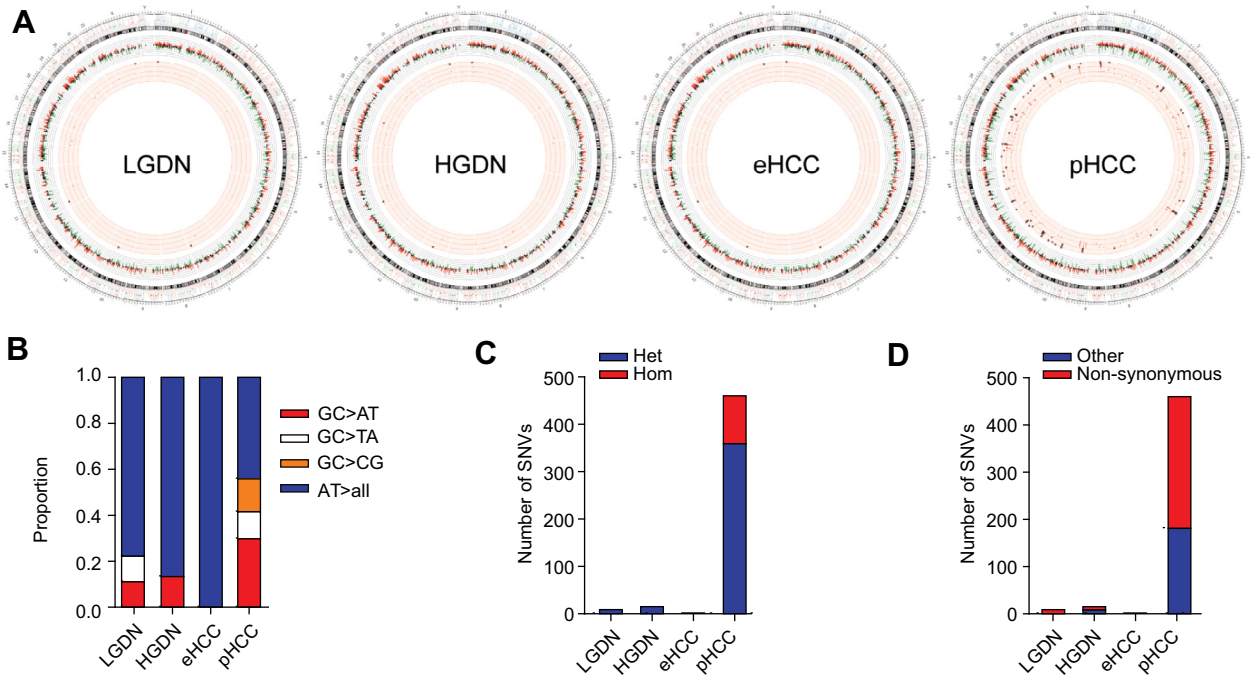


Fig. 2. Evolution of genetic landscape during hepatocarcinogenesis. (A) Figurative representation of the catalogue of somatic mutations and gene expression in different lesions by circus plots. The outside rings show the chromosome ideograms in a clockwise orientation from pter-qter. Centromer locations are indicated in red. The next tracks contain the $\log_2$ expression values with red and black colors indicating overexpression (>2-fold) and green color indicating downregulation (2-fold), respectively. The inner tracks contain the single nucleotide variants (SNV) in exonic regions. Only SNVs with a frequency of alternative allele >35% are displayed. The colors are represented as follows: red color (novel SNVs); red color with black circle (novel SNVs and frequency of alternative allele >0.75). (B) Frequency of nucleotide substitutions in LGDN, HGDN, eHCC and pHCC. (C) Distribution of homozygous *vs.* heterozygous and (D) non-synonymous *vs.* other genetic alterations in LGDN, HGDN, eHCC, and pHCC. All panels are from the PT5.

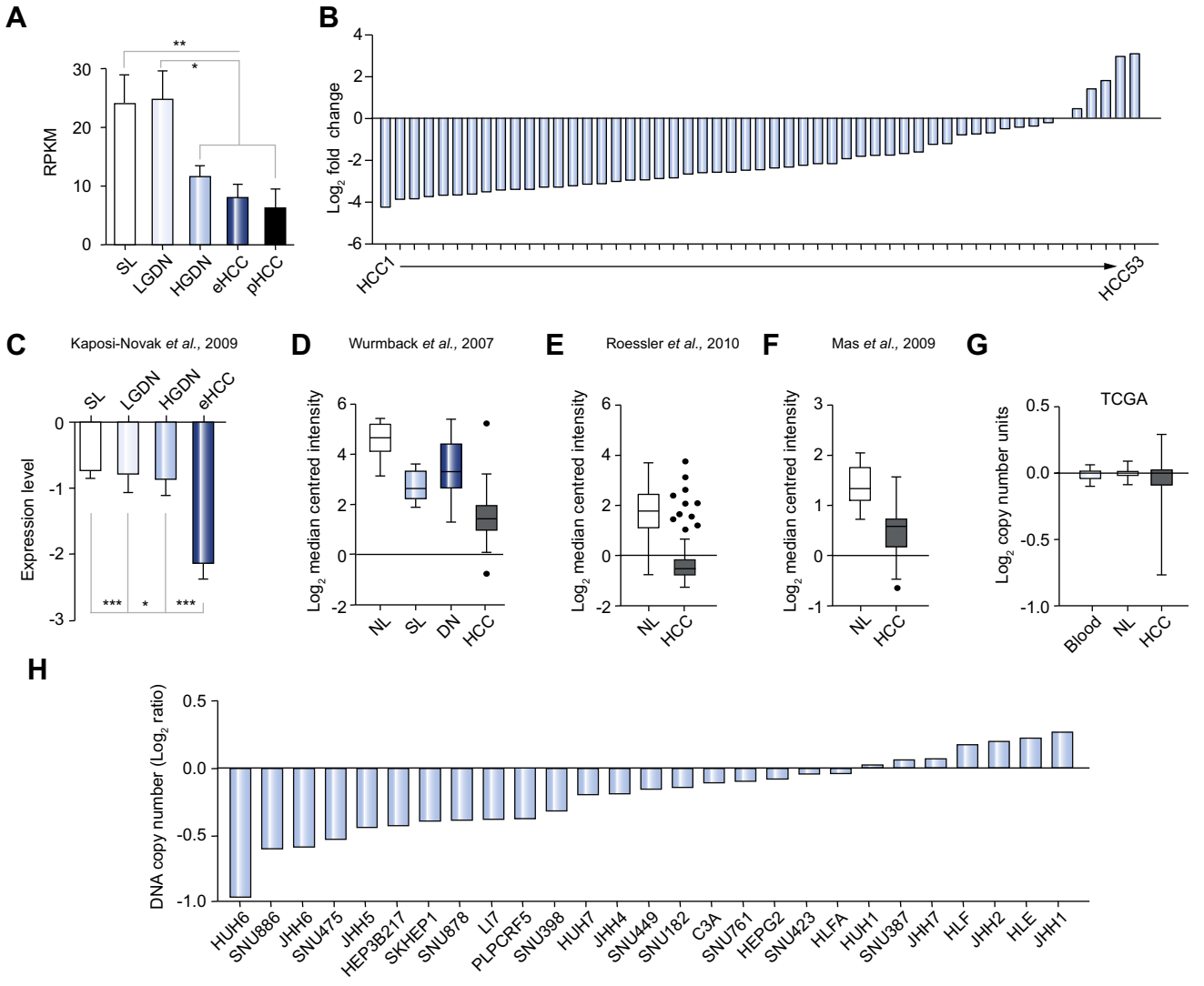


Fig. 3. IGFALS is a novel marker for malignant progression in liver cancer. (A) RPKM values for a new SNV variant in IGFALS show a significant downregulation from surrounding livers (SL) to pHCC (*p <0.05; **p <0.01). (B) Downregulation of IGFALS expression in our previously published cohort of primary human HCC (n = 53) [22]. Values are referenced to normal liver. (C) IGFALS expression in the cohort of patients with SL, LGDN, HGDN, and eHCC [25]. *p <0.05; ***p <0.001. (D–G) Validation of IGFALS repression in publically available HCC cohorts using the Oncomine database. For the selection of datasets, p <0.001 and fold change >2 were required. (D) Wurmbach *et al.* [44] (NL = 10; SL = 13; dysplastic nodules = 17; HCC = 35), (E) Roessler *et al.* [45] (NL = 220; HCC = 225), and (F) Mas *et al.* [46] (NL = 19; HCC = 38). (G) Alterations in *IGFALS* copy numbers in TCGA dataset (blood = 35; NL = 33; HCC = 53). (H) Copy-number variations in *IGFALS* in human hepatoma cell lines. Data were obtained from the Broad-Novartis Cancer Cell Line Encyclopedia (CCLE).

(9), HGDL (15), and eHCC (2) but drastically increased during conversion to pHCC (460 SNVs) (Fig. 2A). Despite the extensive differences in number, the most common proportional nucleotide exchange in the different lesions were AT>>all changes. In contrast to a recent study we also did not observe the over-representation of GC>>TA changes [23]. However, supporting a late stage acquisition of the genetic alterations, both GC>>TA and GC>>CG changes were predominantly seen in pHCC (Fig. 2B). SNVs in pHCC also showed a significantly higher proportion of homozygote and non-synonymous changes (Fig. 2C and D; Supplementary Table 4). In agreement with previous studies, the number of SNVs with a potentially damaging effect was reduced to 5 in LGDN, 4 in HGDN, 2 in eHCC and remained highest in pHCC (110) indicating that disruption of the transcriptome in pHCC could be partly explained by the observed 22–55-fold increase in deleterious genetic alterations (Supplementary Table 4) [23,24].

A small number of SNVs in exonic regions, i.e., with high probability of functional relevance, were present in all lesions and surrounding livers (*UQCRHL*, *GLUD2*, *EIF5AL1*, *PABC3*, and *ERGIC3*). The only exclusive exonic and non-synonymous SNV common for dysplastic and cancerous lesions was found in *IGFALS* gene (hs16_1840768) encoding a serum protein binding the insulin-like growth factors. The presence of this variant showed a predicted protein damaging effect (score 0.970; sensitivity: 0.77; specificity: 0.96) and potentially led to a progressive downregulation of *IGFALS* from HGDN to pHCC (Fig. 3A). The decrease in *IGFALS* expression was confirmed in >350 human HCCs

(Fig. 3B–F) including our previously published data set containing eHCC (Fig. 3C) and correlated with *IGFALS* gene copy number loss in human HCC (Fig. 3G) and human hepatoma cell lines (Fig. 3H) suggesting down-regulation of *IGFALS* as a genetic biomarker of hepatocarcinogenesis.

Discussion

We have applied an integrative transcriptome sequencing approach to address a central issue in human hepatocarcinogenesis, namely, the probability of malignant transformation of the pre-neoplastic lesions, starting from dysplastic lesions to early carcinoma (eHCC) and ultimately to fully progressed HCC (pHCC). We report that genome-wide molecular analysis does not accurately recapitulate the histological distinction of different lesions. We find that the transcriptomes of the early lesions including eHCC were surprisingly homogenous, despite a progressive increase in immune response and proliferation. Activation of prognostic adverse signaling pathways in pHCC, centered on *TGFβ, WNT, NOTCH, MYC,* and *EMT*-related genes, occurred only late during hepatocarcinogenesis. The prominent role of *MYC* target gene activation during the multistage malignant transformation is well recognized [25]. Although eHCC from Korean patients with hepatitis B infection in this study did not show early *c-MYC* activation, our results validate the role of c-MYC signaling in the malignant switch to pHCC (Fig. 1C and Supplementary Fig. 4C) [26]. Of note, the examined eHCC also lacked the expression of other established markers (i.e., *AFP*, *GPC3*, *GS*, and *HSP70*) (Table 1).

A unique feature of our study was a parallel analysis of the full spectrum of early and advanced liver lesions from a single patient. Besides a common deregulation of known pro-oncogenic signaling such as TGFβ, PI3K and WNT/β-Catenin pathways found in the pHCC group comparison (Supplementary Fig. 4), the pHCC in PT5 was characterized by a loss of *CDKN2A* as well as a deregulation of *EGFR* and activation of androgen receptor signaling (Supplementary Fig. 7B–E), both implicated in gender disparity characteristic of HCC development [27,28]. Of note, although the same cell origin of the eHCC and pHCC from PT5 cannot be excluded, the absence of vascular invasion and tumor metastasis and profound differences in tumor genetics (i.e., SNVs) and transcriptome are more consistent with a multicentric tumor origin rather than intrahepatic metastasis [29]. These results underscore the value of in-depth analyses of individual tumor samples to identify the unique patient-specific oncogenic pathways that are underappreciated by group comparisons. The concomitant analyses of individual as well as common characteristics of HCCs might open new windows for diagnostic and/or therapeutic interventions and help to improve the individual patient survival.

The massive disruption of the transcriptome in pHCC could be explained at least in part by a sharp rise in the spectrum of genetic alterations (Fig. 2A). The numbers of SNVs identified in early lesions and eHCC were consistently low but significantly increased upon progression to pHCC. Further, the SNVs with functional consequences on gene expression, i.e., detected in genes with differential expression in pHCC, affected key cancer-associated molecules, such as TGFβ1, NOTCH2, VCAM1, JAK1, ITGA6, ITGB1, IGF2, IGFBP5, XRCC5, DKK3, and MMP14 [30–37], suggesting that genetic variations in these genes may represent acquisition of additional driver mutations during late stages of hepatocarcinogenesis [38]. Notably, SNVs recently identified in the context of hepatocarcinogenesis were exclusively present in pHCC and associated with frequently deregulated genes such as *APC*, *IRF2*, *KRAS*, *MET*, *NFE2L1*, *RPS6KA1*, *RPS6KB2* in addition to p53 (*TP53*, *TP53BP2*, *TP53RK*), WNT/β-Catenin (*AXIN1*), ERB (*EGFR*, *ERBB2*, *ERBB2IP*) and genes involved in chromatin remodeling (*ARID3A*, *ARID3C*, *ARID4A*, *ARID5B*) [23,24], again emphasizing that acquisition of malignant traits is a relatively late event. We also detected previously recognized recurrent somatic homozygotic exonic SNV (rs2304347) for AZIN1 in pHCC as well as several exonic SNVs in ADAR that might result in AZIN1 changes. These results validate the importance of RNA editing for hepatocarcinogenesis [39]. However, the distribution of fusion events was comparable within the different stages, and we did not identify any apparent cancer driving fusion events (data not shown). Of note, there is a possibility that some stop or frameshift mutations (e.g., in tumor suppressor genes), which could potentially lead to mRNA decay, as well as mutations in non-transcribed regions (e.g., promoter regions) could have been missed by RNA-Seq [40]. Furthermore, no blood samples were available from the investigated patients. Therefore, surrounding liver was used as a reference to identify somatic mutations. Since the examined HCCs developed in the context of HBV-induced liver cirrhosis, we cannot exclude that some genetic/epigenetic alterations have already occurred within the altered liver microenvironment.

Another important finding of this study was the identification of IGFALS as a key genetic determinant preferentially down-regulated in pHCC (Fig. 3). Notably, alterations in IGFALS were previously recognized in advanced HCC by integrating information from different molecular layers [41]. In addition to a proposed tumor suppressor role, we demonstrate here that the disruption of IGFALS led to the enhanced IGF signaling [42,43] (Supplementary Fig. 8A–D) thereby providing a potential mechanism for its role in promoting hepatocarcinogenesis. Future clinical and translational efforts are warranted to address the potential utility of IGFALS for routine clinical applications.

Collectively, our results have two major implications. The homogenous expression patterns in preneoplastic lesions do not allow to predict, which lesions progress to liver cancer, highlighting the importance of intensified surveillance and early detection for successful therapeutic interventions. Secondly, the high molecular heterogeneity of pHCC characterized by activation of the prognostic adverse pathways, such as TGFβ, WNT, NOTCH, MYC, and EMT, may underscore the poor response to standard therapies in the current clinical trials and the need for individualized treatment at the progressed stages of HCC. According to the current guidelines, lesions <1 cm are closely monitored by ultrasound [4]. If the results of the present study are validated in larger patient cohorts with a more diverse etiological background, they could serve as a basis for a new intensified diagnostic algorithm that includes consequent biopsies of early lesions. In particular, it can be applied to non-cirrhotic patients with a well-preserved liver function, when early lesions are easier to target, and biopsies can be relatively safely performed. Future translational efforts should further evaluate if biopsies of suspicious lesions during early stages of hepatocarcinogenesis could be supported by implementation of next-generation sequence technologies. This approach would ultimately provide a detailed catalogue of genetic alterations during hepatocarcinogenesis that

can serve as guide for preventive treatment of the lesions (e.g., local ablation) and contribute to reduce HCC incidence and recurrence.

Financial support

This work was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. This work was partly supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MOST) (2011-0030707) to Y.N.P. J.U.M. is supported by a grant from the German Research Foundation (MA 4443/2-1).

Conflict of interest

The authors who have taken part in this study declared that they do not have anything to disclose regarding funding or conflict of interest with respect to this manuscript.

Acknowledgements

This work was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. This work was partly supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MOST) (2011-0030707) to Y.N.P. J.U.M. is supported by a grant from the German Research Foundation (MA 4443/2-1).

Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jhep.2013.10.014.

References

[1] Kojiro M, Roskams T. Early hepatocellular carcinoma and dysplastic nodules. Semin Liver Dis 2005;25:133–142.
[2] Pathologic diagnosis of early hepatocellular carcinoma. A report of the international consensus group for hepatocellular neoplasia. Hepatology 2009;49:658–664.
[3] Aihara T, Noguchi S, Sasaki Y, Nakano H, Monden M, Imaoka S. Clonal analysis of precancerous lesion of hepatocellular carcinoma. Gastroenterology 1996;111:455–461.
[4] European Association For The Study Of The L, European Organisation For R, Treatment Of C. EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma. J Hepatol 2012;56:908–943.
[5] Llovet JM, Di Bisceglie AM, Bruix J, Kramer BS, Lencioni R, Zhu AX, et al. Design and endpoints of clinical trials in hepatocellular carcinoma. J Natl Cancer Inst 2008;100:698–711.
[6] Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009;10:R25.
[7] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAM tools. Bioinformatics 2009;25:2078–2079.
[8] Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 2008;5:621–628.
[9] Wang L, Feng Z, Wang X, Wang X, Zhang X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics 2010;26:136–138.
[10] Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics 2009;25:2283–2285.
[11] Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.
[12] Marquardt JU, Galle PR, Teufel A. Molecular diagnosis and therapy of hepatocellular carcinoma (HCC): an emerging field for advanced technologies. J Hepatol 2012;56:267–275.
[13] Alfano D, Votta G, Schulze A, Downward J, Caputi M, Stoppelli MP, et al. Modulation of cellular migration and survival by c-Myc through the downregulation of urokinase (uPA) and uPA receptor. Mol Cell Biol 2010;30:1838–1851.
[14] Alonso SR, Tracey L, Ortiz P, Perez-Gomez B, Palacios J, Pollan M, et al. A high-throughput study in melanoma identifies epithelial-mesenchymal transition as a major determinant of metastasis. Cancer Res 2007;67:3450–3460.
[15] Chiang DY, Villanueva A, Hoshida Y, Peix J, Newell P, Minguez B, et al. Focal gains of VEGFA and molecular classification of hepatocellular carcinoma. Cancer Res 2008;68:6779–6788.
[16] Coulouarn C, Factor VM, Thorgeirsson SS. Transforming growth factor-beta gene expression signature in mouse hepatocytes predicts clinical outcome in human cancer. Hepatology 2008;47:2059–2067.
[17] Hoshida Y, Nijman SM, Kobayashi M, Chan JA, Brunet JP, Chiang DY, et al. Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. Cancer Res 2009;69:7385–7392.
[18] Lee JS, Heo J, Libbrecht L, Chu IS, Kaposi-Novak P, Calvisi DF, et al. A novel prognostic subtype of human hepatocellular carcinoma derived from hepatic progenitor cells. NatMed 2006;12:410–416.
[19] Nguyen BC, Lefort K, Mandinova A, Antonini D, Devgan V. Della Gatta G, et al. Cross-regulation between Notch and p63 in keratinocyte commitment to differentiation. Genes Dev 2006;20:1028–1042.
[20] Rutella S, Bonanno G, Procoli A, Mariotti A, de Ritis DG, Curti A, et al. Hepatocyte growth factor favors monocyte differentiation into regulatory interleukin (IL)-10++IL-12low/neg accessory cells with dendritic-cell features. Blood 2006;108:218–227.
[21] Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, et al. PID: the pathway interaction database. Nucleic Acids Res 2009;37:D674–D679.
[22] Andersen JB, Factor VM, Marquardt JU, Raggi C, Lee YH, Seo D, et al. An integrated genomic and epigenomic approach predicts therapeutic response to zebularine in human liver cancer. Sci Transl Med 2010;2:54ra77.
[23] Guichard C, Amaddeo G, Imbeaud S, Ladeiro Y, Pelletier L, Maad IB, et al. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. Nat Genet 2012;44:694–698.
[24] Fujimoto A, Totoki Y, Abe T, Boroevich KA, Hosoda F, Nguyen HH, et al. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. Nat Genet 2012;44:760–764.
[25] Kaposi-Novak P, Libbrecht L, Woo HG, Lee YH, Sears NC, Conner EA, et al. Central role of c-Myc during malignant conversion in human hepatocarcinogenesis. Cancer Res 2009;69:2775–2782.
[26] Thorgeirsson SS. The almighty MYC: orchestrating the micro-RNA universe to generate aggressive liver cancer. J Hepatol 2011;55:486–487.
[27] Keng VW, Sia D, Sarver AL, Tschida BR, Fan D, Alsinet C, et al. Sex bias occurrence of hepatocellular carcinoma in Poly7 molecular subclass is associated with EGFR. Hepatology 2013;57:120–130.
[28] Li Z, Tuteja G, Schug J, Kaestner KH. Foxa1 and Foxa2 are essential for sexual dimorphism in liver cancer. Cell 2012;148:72–83.
[29] Roncalli M, Park YN, Di Tommaso L. Histopathological classification of hepatocellular carcinoma. Dig Liver Dis 2010;42:S228–S234.
[30] Chen TY, Li YC, Liu YF, Tsai CM, Hsieh YH, Lin CW, et al. Role of MMP14 gene polymorphisms in susceptibility and pathological development to hepatocellular carcinoma. Ann Surg Oncol 2011;18:2348–2356.
[31] Li R, Yang Y, An Y, Zhou Y, Liu Y, Yu Q, et al. Genetic polymorphisms in DNA double-strand break repair genes XRCC5, XRCC6 and susceptibility to hepatocellular carcinoma. Carcinogenesis 2011;32:530–536.
[32] Luo JH, Ren B, Keryanov S, Tseng GC, Rao UN, Monga SP, et al. Transcriptomic and genomic analysis of human hepatocellular carcinomas and hepatoblastomas. Hepatology 2006;44:1012–1024.
[33] Migita K, Miyazoe S, Maeda Y, Daikoku M, Abiru S, Ueki T, et al. Cytokine gene polymorphisms in Japanese patients with hepatitis B virus infection–association between TGF-beta1 polymorphisms and hepatocellular carcinoma. J Hepatol 2005;42:505–510.

[34] Pei Y, Kano J, Iijima T, Morishita Y, Inadome Y, Noguchi M. Overexpression of Dickkopf 3 in hepatoblastomas and hepatocellular carcinomas. Virchows Arch 2009;454:639–646.
[35] Umemura A, Itoh Y, Itoh K, Yamaguchi K, Nakajima T, Higashitsuji H, et al. Association of gankyrin protein expression with early clinical stages and insulin-like growth factor-binding protein 5 expression in human hepatocellular carcinoma. Hepatology 2008;47:493–502.
[36] Villanueva A, Newell P, Chiang DY, Friedman SL, Llovet JM. Genomics and signaling pathways in hepatocellular carcinoma. Semin Liver Dis 2007;27:55–76.
[37] Xie HJ, Bae HJ, Noh JH, Eun JW, Kim JK, Jung KH, et al. Mutational analysis of JAK1 gene in human hepatocellular carcinoma. Neoplasma 2009;56:136–140.
[38] Woo HG, Park ES, Lee JS, Lee YH, Ishikawa T, Kim YJ, et al. Identification of potential driver genes in human liver carcinoma by genomewide screening. Cancer Res 2009;69:4059–4066.
[39] Chen L, Li Y, Lin CH, Chan TH, Chow RK, Song Y, et al. Recoding RNA editing of AZIN1 predisposes to hepatocellular carcinoma. Nat Med 2013;19:209–216.
[40] Nault JC, Mallet M, Pilati C, Calderaro J, Bioulac-Sage P, Laurent C, et al. High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions. Nat Commun 2013;4.
[41] Neumann O, Kesselmeier M, Geffers R, Pellegrino R, Radlwimmer B, Hoffmann K, et al. Methylome analysis and integrative profiling of human HCCs identify novel protumorigenic factors. Hepatology 2012;56:1817–1827.
[42] Boudoukha S, Cuvellier S, Polesskaya A. Role of the RNA-binding protein IMP-2 in muscle cell motility. Mol Cell Biol 2010;30:5710–5725.
[43] Kuninger D, Kuzmickas R, Peng B, Pintar JE, Rotwein P. Gene discovery by microarray: identification of novel genes induced during growth factor-mediated muscle cell survival and differentiation. Genomics 2004;84:876–889.
[44] Wurmbach E, Chen YB, Khitrov G, Zhang W, Roayaie S, Schwartz M, et al. Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma. Hepatology 2007;45:938–947.
[45] Roessler S, Jia HL, Budhu A, Forgues M, Ye QH, Lee JS, et al. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. Cancer Res 2010;70:10202–10212.
[46] Mas VR, Maluf DG, Archer KJ, Yanek K, Kong X, Kulik L, et al. Genes involved in viral carcinogenesis and tumor initiation in hepatitis C virus-induced hepatocellular carcinoma. Mol Med 2009;15:85–94.