<https://helda.helsinki.fi>

Gene expression profiling, pathway analysis and subtype classification reveal molecular heterogeneity in hepatocellular carcinoma and suggest subtype specific therapeutic targets

Agarwal, Rahul

2017-10

Agarwal , R , Narayan , J , Bhattacharyya , A , Saraswat , M & Tomar , A K 2017 , ' Gene expression profiling, pathway analysis and subtype classification reveal molecular heterogeneity in hepatocellular carcinoma and suggest subtype specific therapeutic targets ' , Cancer Genetics , vol. 216 , pp. 37-51 . <https://doi.org/10.1016/j.cancergen.2017.06.002>

<http://hdl.handle.net/10138/298096>

<https://doi.org/10.1016/j.cancergen.2017.06.002>

publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

ORIGINAL ARTICLE

Gene expression profiling, pathway analysis and subtype classification reveal molecular heterogeneity in hepatocellular carcinoma and suggest subtype specific therapeutic targets

Rahul Agarwal ^a, Jitendra Narayan ^b, Amitava Bhattacharyya ^c,
Mayank Saraswat ^d, Anil Kumar Tomar ^{e,*}

^a Department of Reproductive Biology, All India Institute of Medical Sciences, New Delhi, 110029 India; ^b Unité de recherche en biologie environnementale et évolutive (URBE), University of Namur, Belgium; ^c Excelra Knowledge Solutions Pvt Ltd, Hyderabad, Telangana 500039, India; ^d Transplantation Laboratory, Haartmaninkatu 3, University of Helsinki, Helsinki, Finland; ^e Kusuma School of Biological Sciences, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India

A very low 5-year survival rate among hepatocellular carcinoma (HCC) patients is mainly due to lack of early stage diagnosis, distant metastasis and high risk of postoperative recurrence. Hence ascertaining novel biomarkers for early diagnosis and patient specific therapeutics is crucial and urgent. Here, we have performed a comprehensive analysis of the expression data of 423 HCC patients (373 tumors and 50 controls) downloaded from The Cancer Genome Atlas (TCGA) followed by pathway enrichment by gene ontology annotations, subtype classification and overall survival analysis. The differential gene expression analysis using non-parametric Wilcoxon test revealed a total of 479 up-regulated and 91 down-regulated genes in HCC compared to controls. The list of top differentially expressed genes mainly consists of tumor/cancer associated genes, such as AFP, THBS4, LCN2, GPC3, NUF2, etc. The genes over-expressed in HCC were mainly associated with cell cycle pathways. In total, 59 kinases associated genes were found over-expressed in HCC, including TTK, MELK, BUB1, NEK2, BUB1B, AURKB, PLK1, CDK1, PKMYT1, PBK, etc. Overall four distinct HCC subtypes were predicted using consensus clustering method. Each subtype was unique in terms of gene expression, pathway enrichment and median survival. Conclusively, this study has exposed a number of interesting genes which can be exploited in future as potential markers of HCC, diagnostic as well as prognostic and subtype classification may guide for improved and specific therapy.

Keywords Hepatocellular carcinoma, kinases, mRNA expression, subtype classification, survival analysis

© 2017 Elsevier Inc. All rights reserved.

Introduction

One of the most common cancers worldwide, hepatocellular carcinoma (HCC) is the most frequently and recurrently occurring primary liver tumor. HCC is responsible for 500,000 deaths annually, mostly shared among Asian and African countries; however, the incidence rate is also continuously

increasing across all the continents (1–3). Early stage diagnosis of HCC is an insurmountable task due to lack of any specific symptoms. As a result most of the patients are diagnosed at very late stage, a cause of comparatively lower 5-year survival rate (~10%) (2,4,5). Patients with advanced stage HCC are likely to have higher chances of distant metastasis (6), which adversely affects survival of a patient. The surgical resection and liver transplantation are the main resorts for treating HCC patients, but the prognosis is still very difficult due to lack of early stage diagnosis, distant metastasis and high risk of postoperative recurrence. All these issues highlight a need to discover novel and potential molecular biomarkers

Received March 23, 2017; received in revised form June 12, 2017; accepted June 30, 2017.

* Corresponding author.

E-mail address: anilbioinfo@gmail.com

that can help in early stage diagnosis and also can be targeted to tailor the specialized treatment for the high-risk patients.

Recent technological advances and method automation of RNA-sequencing have enabled the scientific community to analyze the whole transcriptome comprehensively, that too in large number of patients and comparatively in very short span of time. The Cancer Genome Atlas (TCGA) (<http://cancergenome.nih.gov>), a collaborative unit of two National Institutes of Health (NIH) centers – National Cancer Institute and the National Human Genome Research Institute, is an enriched database of enormous amount of cancer genomics data for more than 30 different types of cancers, including HCC. Researchers worldwide routinely explore TCGA expression data (cancer vs. control) to gain crucial information to reveal plausible candidate molecular markers for better diagnostics, separation of cancer subtypes and alternative therapeutic options.

Here, we have investigated the gene expression data of HCC solid tumors vs. controls to unravel the list of the differentially expressed genes, followed by pathway analysis. The top twenty differentially expressed genes were correlated with overall survival and disease-free survival to estimate the prognostic impact of testing in HCC. In addition, we have performed analysis of dysregulated kinases, determination of possible HCC subtypes and their overall characterization based on median survival, gene expression and biological pathway analysis associated with differentially expressed genes assigned to each subtype.

Materials and methods

Retrieval of patients' data

Level 3 gene expression matrices of 423 HCC patients were retrieved from TCGA data portal (The Cancer Genome Atlas Liver Hepatocellular Carcinoma (TCGA-LIHC) data; <https://cancergenome.nih.gov/>). Out of 423 samples, 373 were solid tumor tissue samples (HCC) and 50 were the adjacent non-tumor tissue samples (controls). The transcriptomics expression data were prepared using reads obtained from the Illumina HiSeq 2000 sequencing machine (Illumina, Inc., San Diego, CA) and expression matrix of 20,498 genes for all 423 samples was already quantile normalized using RSEM (7).

Bioinformatics analysis

Combined gene expression data of 423 samples (HCC and controls) was preprocessed. Genes with more than 50% missing data were excluded and genes with expression >5 in more than 80% of the samples were used for further analysis. Differential gene expression analysis in HCC samples compared to controls was performed for 13,571 genes by calculating log2foldchange and non-parametric Wilcoxon test (8). To account for multiple testing, adjusted p-value was estimated using the Benjamin–Hochberg method (9). Transcripts having absolute log2foldchange values ≥ 2 (for up-regulation) and < -2 (for down-regulation) were considered as significantly differentially expressed. Aberrant expressed kinases were separately filtered out based on absolute $\log_2$ foldchange > 1 . Hierarchical clustering of 25 most variable kinase genes in overall 423 hepatocellular carcinoma samples (tumors and controls) was done

using Bioconductor R package *pvclust* (method.dist = “correlation”, method.hclust = “complete”, nboot = 1000). *Pvclust* is used to assess an uncertainty in hierarchical cluster analysis. *Pvclust* provides two types of p-values: AU (Approximately Unbiased) and BP (Bootstrap Probability). AU p-value is computed by multiscale bootstrap resampling and is considered a better approximation of unbiased p-value than BP p-value which is computed by normal bootstrap resampling.

Enrichment analysis and gene–gene interaction

Gene ontology of differentially expressed genes (up- and down-regulated) was performed using Database for Annotation, Visualization and Integrated Discovery (DAVID) gene enrichment tool (10). Reactome pathway database was used to report statistically enriched biological pathways in case of differentially expressed genes (11). The STRING: functional protein association networks database version 10.5 (12) was used to detect interaction networks between gene products of differentially expressed genes, first with default settings (active interaction sources: text mining, experiment, database, co-expression, neighborhood, gene fusion and co-occurrence; and minimum required interaction score: medium confidence, 0.4) and then with following settings: active interaction source: experiment; and minimum required interaction score: high confidence, 0.7. The disconnected nodes were removed from the network.

Classifiers for HCC

New expression matrix of 570 differentially expressed genes for 423 samples (HCC and controls) was created for preparing the list of top predictors. Oversampling of minority samples in new expression matrix was performed using SMOTE (13). The gene selection was performed by randomly splitting oversampled new expression matrix into n number of learning sets (nfolds) for x number of iterations (14). In this way, gene selection procedure was repeated $n \times x$ times to get the final list of the predictor genes (≤ 30) using Random Forest Variable Importance Measure (15). Final list of most important genes was utilized for training the classifiers to estimate classification error rate and/or prediction of classes (*i.e.*, HCC and normal) on uncategorized quantile normalized and log transformed expression data using learning sets generated earlier.

Subtype prediction

To predict HCC subtypes associated with TCGA-LIHC data, genes were filtered out based on the coefficient of variance. Overall, 2714 most variable genes were sorted out for predicting the subtypes. Unsupervised hierarchical clustering was done on these genes across all the 373 HCC tumor samples using Bioconductor R based package *ConsensusClusterPlus* (16). The following command was used:

```
ConsensusClusterPlus(data,maxK = 8, reps = 2000, pItem = 0.8, pFeature = 1, innerLinkage = "ward.D2", finalLinkage = "ward.D2", title = "LIHC_subtype", distance = "pearson", clusterAlg = "hc", seed = 1262118388.71279, plot = "png", writeTable = TRUE)
```

Final cluster attained the consensus after 2000 reiterations. The number of clusters that represent the HCC expression

data most significantly was selected by silhouette method of k-Means clustering (17). This method is used to calculate the separation distance between the resulting clusters and to estimate how close each point in one of the clusters is to points in the neighboring clusters. Silhouette coefficient values are in the range of $[-1, 1]$. Silhouette coefficients close to +1 indicate that the sample is far away from the neighboring clusters, 0 indicates that the sample is on or very close to the decision boundary between two neighboring clusters and negative values indicate that samples might have been assigned to the wrong cluster. Silhouette coefficients were estimated using Bioconductor R based package *Cluster* (18). HCC samples with positive silhouette coefficient values were selected for further analysis. Top differentially expressed genes were obtained for each $k = 1$ to n subtypes by employing *sam* function from Bioconductor package *siggenes* (19) using the following code-sam (data1, temp, gene.names = rownames (data1), rand = 123456).

To further distinguish each HCC subtype, associated biological pathways were analyzed for each subtype using Reactome pathway database.

Survival analysis

Median overall survival and progression-free survival (in months) were computed for a set of genes in HCC samples in which at least one of the query genes was differentially expressed vs. HCC samples in which none of the query genes was differentially expressed. A Kaplan–Meier (KM) curve with p-values from a log rank test was used for presenting the results of survival analysis (20). Overall median KM based survival analysis of selected HCC subtypes was performed using HCC clinical data retrieved from TCGA using the coxph model (21).

Results and discussion

TCGA is a cancer genomic database containing data sets for over 30 malignancies. The availability of rich data source enables cancer researchers to explore clinically relevant information for comprehensive understanding of biological mechanisms/pathways underlying different types of cancers. Early diagnosis of HCC is very rare due to lack of specific symptoms, while diagnosis of advanced stage HCC is hindered due to high tumor heterogeneity. Almost non-existence of precise diagnostic tools and the urge for early-stage diagnosis incited us to explore publicly available TCGA genomic data in quest of HCC biomarkers.

Gene expression analysis

We processed mRNA gene expression data from 423 samples (HCC tumor samples = 373; controls = 50) downloaded from TCGA. Principal component analysis (PCA) of strongly expressed genes showed an almost complete separation of tumor and control cohorts (Figure 1). A substantial variability was observed among the HCC tumor samples indicating the presence of tumor heterogeneity (Figures 2 and 3). Differential gene expression analysis revealed a large number of differentially expressed genes ($n = 570$) in tumor samples compared with controls. In HCC, 479 genes were up-regulated while 91 genes were down-regulated (Supplementary file 1).

Top 20 differentially expressed genes are listed in Table 1. Surprisingly, all top 20 differentially expressed genes were over-expressed in HCC. Out of these, one gene (NCRNA00176) was a non-protein coding RNA 176 located at Chromosome 20. The list of top 20 genes includes some of the important tumor/cancer associated genes such as AFP, THBS4, LCN2, GPC3, NUF2, etc. Overall survival analysis of top 20 differentially expressed genes indicated significant improvement in survival duration and disease-free survival of patients if none of these 20 genes was altered (Figures 4 and 5).

AFP encodes for alpha-fetoprotein, a major plasma protein produced by the liver and yolk sac during fetal development and a well-known biomarker for HCC evaluation. Serum AFP protein usually increases up to a concentration of approx. 3 g/L during 12–16 weeks, then rapidly decreases to traces and its abnormally high concentration in serum is correlated with the development of several malignant diseases, most notably HCC (22). THBS4 encodes for thrombospondin 4 protein, a secreted multidomain glycoprotein of the extracellular matrix. It is considered as the most potent marker for diffuse-type gastric adenocarcinoma with higher expression at both mRNA and protein levels (23). As shown by immune-histochemical colocalization studies, cancer-associated fibroblasts (CAFs) of the myofibroblast phenotype are the main sources of secretion of this gene (23).

LCN2 encodes for a protein lipocalin 2. This protein has been reported to be over-expressed in HCC tissue samples in a few of previous studies and its higher expression was correlated with overall shorter survival in HCC patients (24). Thus, it was proposed as a promising diagnostic protein marker for HCC progression. GPC3 encodes for glypican 3 protein, a cell-surface oncofetal heparan sulfate proteoglycan. Its over-expression in HCC compared to controls has been associated with poor prognosis in HCC patients. Recently GPC3 was reported to be involved in HCC progression (25). Immunotherapeutic potential of GPC3 in HCC treatment is under investigation and one of the recent studies has demonstrated significant regression of tumor xenografts in two of the human liver cancer cell lines, Hep3B and HepG2, treated with anti-GPC3 immunotoxin (25). NUF2 gene encodes for Kinetochore protein Nuf2, a component of a conserved protein complex NDC80 associated with the centromere. It is also known as cell division cycle-associated protein 1. In mitosis, Nuf2, as well as other NDC80 complex components (primarily NDC80, CDCA1, SPBC24 and SPBC25), takes part in kinetochore–microtubule attachment (26) and is considered as a tumor-associated antigen valuable for both cancer diagnosis and immunotherapy (27). It was strongly correlated with 5 kinase genes—NEK2, CDK1, TTK, MELK and PLK1 (>0.9 Spearman correlation between NUF2 and these kinases). NUF2 gene was found altered in 70 of 373 HCC tumor samples. It was significantly associated (p-value: 1.012×10^{-4}) with overall median survival. Average survival in HCC patients with altered NUF2 gene was 21.68 months in comparison to 69.51 months in HCC patients with non-altered gene. NUF2 knockdown studies have shown blocked cell proliferation and stimulated apoptosis in HCC, colorectal and gastric cancers (28). Also, a very recently proposed multivariate logistic regression model stated that joint expression patterns of CEP55, NUF2, PBK and TTK could differentiate between metastatic and localized prostate cancer (29).

The top three down-regulated genes found in HCC were FCN3, HAMP and OIT3 that encode for proteins ficolin 3,

Figure 3 Heat map of 100 randomly selected hepatocellular carcinoma samples. It indicates presence of disproportion between samples using distance method.

Table 1 Top differentially expressed genes in hepatocellular carcinoma compared to controls using log2foldchange and p-value estimation using Wilcoxon test

Genes	Tumor	Normal	log2fc	p-Values	Fdr
AFP	19681.19	116.2878	7.402974073	0.007667	1
THBS4	582.8846	3.81122	7.256813662	7.53E-26	1.01E-21
TGM3	576.9567	8.014694	6.169671709	3.79E-17	4.61E-13
S100P	2022.551	30.28527	6.061415762	2.04E-05	0.143583
GPC3	15735.71	237.9249	6.047391997	1.92E-19	2.42E-15
NXPH4	190.8503	3.066072	5.959906324	4.31E-24	5.72E-20
ALDH3A1	4838.89	119.7535	5.336535763	0.011256	1
DUSP9	743.4116	18.85853	5.300872191	4.14E-20	5.27E-16
MUC13	4183.897	106.8973	5.290549915	1.10E-10	1.12E-06
CDC25C	130.8541	3.833028	5.093331234	3.20E-27	4.33E-23
CYP17A1	2192.415	64.24748	5.092737137	0.001251	1
NUF2	189.9189	5.809948	5.030714194	1.03E-27	1.40E-23
GABRD	110.9463	3.600396	4.945562757	1.23E-29	1.67E-25
MYBL2	671.8886	21.97286	4.93442771	6.07E-26	8.17E-22
NCRNA00176	130.3388	4.293288	4.924039399	1.21E-19	1.53E-15
BIRC5	333.7033	11.51355	4.857161212	1.13E-27	1.53E-23
NQO1	5845.998	202.6475	4.850405146	3.34E-11	3.46E-07
LCN2	6440.081	230.745	4.802707738	8.41E-08	0.000729
COCH	281.4672	10.27386	4.775916977	2.25E-10	2.25E-06
TTK	145.9646	5.56638	4.712734896	2.36E-26	3.17E-22

Figure 4 Kaplan–Meier curves of overall survival for total hepatocellular carcinoma patient population (N = 373) using top 20 differentially expressed genes (as listed in Table 1).

Figure 5 Kaplan–Meier curves of disease-free survival for total hepatocellular carcinoma patient population (N = 373) using top 20 differentially expressed genes (as listed in Table 1).

hepcidin antimicrobial peptide and oncoprotein induced transcript 3 respectively. FCN3 (30), HAMP (31,32) and OIT3 have also been previously reported as less expressed or repressed in HCC (33). Also, a kinase gene PCK1 that encodes for phosphoenolpyruvate carboxykinase 1 was found down-regulated. PCK1 participates in gluconeogenesis and has been previously reported as down-regulated in tumor samples (33).

Pathway enrichment

Pathway enrichment analysis of differentially expressed genes was performed to reveal biological pathways dominant in HCC (Supplementary files 2 and 3). In the case of up-regulated genes

in HCC, cell cycle associated pathways were enriched dominantly (Table 2). One of the interesting and distinctive pathways was “Polo-like kinase mediated events” that was related to up-regulation of a mitotic kinase, Polo-like kinase-1 (PLK1). It is well known that PLK1 helps in tumor growth via negative regulation of p53 functioning. It has recently been shown to participate in induction of TLR2- and TLR4-induced inflammation by stimulating MDM2 gene activity through its phosphorylation and G2/M transition (34,35). Targeting p53 through PLK1 has been shown as an attractive chemotherapy strategy in adrenocortical cancer (36) and an inhibitor of PLK1. Volasertib is currently in clinical trials for treatment of acute myeloid leukemia (AML) (<https://clinicaltrials.gov/ct2/show/NCT01721876>). Volasertib blocks PLK1 activity, thus causing termination of cell cycle via G2/M cell cycle

Figure 6 Interaction network of up-regulated genes in hepatocellular carcinoma, generated by STRING database using the following parameters: source – experiment and interaction score ≥ 0.7 .

Table 2 Top 10 significantly enriched pathways derived from Reactome based enrichment analysis of up-regulated genes in hepatocellular carcinoma

Pathway name	#Genes mapped	#Total genes	P-Values
Mitotic prometaphase	36	136	1.11E-16
Cell cycle, mitotic	83	526	1.11E-16
Cell cycle	95	638	1.11E-16
Resolution of sister chromatid cohesion	34	128	3.33E-16
RHO GTPases activate formins	31	141	1.02E-12
Polo-like kinase mediated events	14	23	5.77E-12
Separation of sister chromatids	33	188	6.62E-11
Mitotic anaphase	33	202	3.99E-10
Mitotic metaphase and anaphase	33	203	4.50E-10
Mitotic G1-G1/S phases	27	143	7.97E-10

arrest followed by programmed cell death in leukemia cells (37). In case of down-regulated genes, a total of 35 Reactome pathway terms were enriched. Noticeably, metallothioneins bind metals pathway was the most enriched term and associated down-regulated genes were MT2A, MT1M, MT1F, MT1G, MT1X and MT1E. Several studies in the past have reported down-regulation of these genes and suggested them as plausible prognostic markers of HCC (38–42).

For network analysis of differentially expressed genes, STRING database was used. The interaction networks of down-regulated and up-regulated genes in HCC, created using default settings as described in *Materials and Methods* section, are shown in [Supplementary Figures S1 and S2](#) respectively. The network analysis reveals a strong association between up-regulated genes in comparison to down-regulated genes. CXCL12, CYP2B6, FOS, IGF1, MT1G, NR4A1 and SOCS3 genes form the major nodes of interaction network of down-regulated genes. No interaction was observed between down-regulated genes under more stringent settings (source: experiment and interaction score ≥ 0.7). Interaction network of up-regulated genes with these settings is shown in [Figure 6](#). Interestingly, the major nodes of the network primarily include

genes associated with cell cycle regulation and specifically with assembly of different regulatory components of cell cycle, such as ORC1, CDK1, CDC20, BUB1 and ZWINT. ORC1 is one of the components of origin recognition complex, which binds to origin of replication and is essential for assembly of pre-replication complex to start DNA replication (43). CDK1 or cyclin dependent kinase 1 is known to play a crucial role in controlling eukaryotic cell cycle through regulation of G2-M and G1-S transitions and G1 progress. Recently, it has been suggested as a potential prognostic biomarker of HCC (44). The mitotic spindle assembly checkpoint function is controlled by many genes, including CDC20 and BUB1 (45). In several human cancers, CDC20 protein is found over-expressed and has been associated with poor prognosis in cancers of pancreas, lung, bladder and colon (46). BUB1 is a serine/threonine–protein kinase, which is essential for the assembly of checkpoint proteins at the kinetochore and proper chromosome alignment during mitosis. A recent study has reported that ZWINT is required for kinetochore formation and spindle assembly checkpoint function (47).

Over-expressed kinases

Overall 59 kinase genes were over-expressed in HCC compared to controls, out of which 23 genes exhibited log2 fold value more than 2 (Supplementary file 4). STRING database based protein–protein interaction network of significantly over-expressed (log2 fold ≥ 2) kinases in HCC is presented in [Supplementary Figure S3](#). The top 10 kinases associated genes over-expressed in HCC included TTK, MELK, BUB1, NEK2, BUB1B, AURKB, PLK1, CDK1, PKMYT1, and PBK ($p < 0.05$) ([Table 3](#)). TTK, which participates in the regulation of the DNA damage checkpoint, was previously reported as over-expressed in HCC (48,49). A series of *in vitro* and *in vivo* functional experiments linked TTK over-expression with resistance to sorafenib in HCC cells (48). MELK that encodes for Maternal Embryonic Leucine Zipper Kinase was found highly over-expressed in HCC and correlated with early recurrence and poor survival of patients (50). Role of MELK in HCC was found in same line with the cell cycle- and mitosis-related genes as it binds and triggers transcription factors c-JUN and FOXM1. MELK was proposed as an oncogenic kinase that participates in the pathogenesis and recurrence of HCC and as a potential molecular target for treating advanced HCC cases (50).

Table 3 Top 10 up-regulated kinases associated genes in hepatocellular carcinoma using log2foldchange and p-value estimation by Wilcoxon test

Kinase genes	Tumor	Normal	log2fc	p-Values	Fdr
TTK	145.9645579	5.56638	4.712735	2.36E-26	3.17E-22
MELK	198.5217823	9.522538	4.381807	1.37E-27	1.85E-23
BUB1	217.4096863	11.33522	4.261531	3.97E-26	5.34E-22
NEK2	238.7958292	12.59146	4.24526	3.55E-26	4.78E-22
BUB1B	191.227892	11.44933	4.061958	2.61E-25	3.49E-21
AURKB	226.0252185	14.24596	3.987859	4.74E-26	6.37E-22
PLK1	339.7681271	21.64501	3.972444	4.61E-26	6.21E-22
CDK1	451.5283086	33.09692	3.770048	1.15E-26	1.55E-22
PKMYT1	202.6302445	15.54742	3.704102	2.87E-27	3.89E-23
PBK	179.9522434	13.94364	3.689935	7.68E-25	1.02E-20

The hierarchical clustering of the 25 most variable kinase genes based on *pvclust* (R package) identified two most significant clusters (Figure 7). The first cluster consisted of only two genes, MAPK11 and MAPK12, while the second one had six: BUB1, CDK1, CHEK1, MELK, NEK2 and PLK1. Based on this, we believe that targeting these two clusters would be valuable option for novel cancer therapy.

Reactome pathway analysis revealed that major pathways that enriched among kinase genes over-expressed in HCC included cell cycle, p38MAPK events, AURKA activation by TPX2 and mitotic cell cycle. An important pathway associated with kinase gene SRC CTLA4 inhibitory signaling was found up-regulated. CTLA-4 was proposed to restrict T-cell proliferation, as an early immune response, by blocking both IL-2 production and cell cycle progression (51). It is known that regulatory T cells support activation of CTLA-4 to maintain immune tolerance and an inhibitor of CTLA-4, ipilimumab, has recently acquired FDA approval for the treatment of advanced or unresectable melanomas (52). In addition, six other immune-related pathways were also found significantly enriched ($p < 0.05$), viz. myd88-independent TLR3/TLR4

cascade, TRIF-mediated TLR3/TLR4 signaling, toll like receptor 3 (TLR3) cascade, activated TLR4 signaling, toll like receptor 4 (TLR4) cascade and MAP kinase activation in TLR cascade. MAPK11, IRAK1 and IKBKE genes were associated with any of these six significant immune-related pathways. MAPK11 encodes for mitogen-activated protein kinase 11, a protein which is involved in the integration of biochemical signals for various cellular processes, specifically cell proliferation and transcriptional regulation. It plays a crucial role in the cascades of cellular responses evoked by inflammatory cytokines or physical stress which lead to direct activation of transcription factors. IRAK1 gene encodes for interleukin-1 receptor-associated kinase 1. It regulates expression of several inflammatory genes in immune cells, generating signals for elimination of viruses, bacteria and cancer cells (53). Pharmacologic inhibition of IRAK1 plays a vital role in the treatment of myelodysplastic syndromes and acute lymphoblastic leukemia (54). When cultured IKBKE-driven breast cancer cells were treated with CYT387 (a potent inhibitor of TBK1/IKBKE and JAK signaling), a significant reduction in cell proliferation was observed (55,56). IKBKE along with TBK1 plays a central

Figure 7 Hierarchical clustering of 25 most variable kinase genes in overall 423 hepatocellular carcinoma samples (tumors and controls). AU and BP p-values are shown in red and green colors respectively. Clusters with AU larger than 95% are highlighted by red thin rectangle lines, which are strongly supported by gene expression data of 25 most variable kinase genes. Two significant clusters passed this criterion: the cluster on the right side consists of MAPK11 and MAPK12; and the cluster on the left side consists of BUB1, CDK1, CHEK1, MELK, NEK2 and PLK1. Numbers of genes shown in the figure are those with log2 fold change >1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4 List of up-regulated TGF-beta pathway associated genes in hepatocellular carcinoma. Out of more than 300 TGF-beta pathway associated genes, only a fraction of genes were significantly up-regulated indicating that TCGA data may not represent TGF-beta induced HCC. Also, TGF- β 1 was up-regulated in only 23 of 373 tumor samples (6.16%)

TGF-beta genes	Tumor	Normal	Log2fc	p-Value	Fdr
SPP1	21129.38	999.0499	4.40255	2.74E-05	0.190622
CCNB2	295.7111	15.63288	4.241533	2.39E-26	3.22E-22
CDK1	451.5283	33.09692	3.770048	1.15E-26	1.55E-22
LEF1	189.3522	19.88161	3.251565	1.27E-14	1.46E-10
ITGA2	249.4739	53.30912	2.226435	1.62E-10	1.62E-06
MSX1	67.85667	14.679	2.208737	6.62E-18	8.17E-14
TUBB4	537.2559	118.3876	2.182092	7.80E-08	0.000678
FGD1	178.2042	47.81777	1.897913	1.20E-14	1.38E-10
LAMA4	740.8107	214.4201	1.788665	6.18E-22	8.03E-18
BMP4	332.0811	96.77892	1.778771	0.003654	1
ITGB4	568.1493	166.8375	1.767827	4.52E-05	0.305855
COL4A2	6890.349	2026.717	1.765433	7.17E-17	8.67E-13
COL1A2	7463.762	2611.095	1.515248	0.008552	1

role in synchronizing the activation of transcription factors, including interferon regulatory factor 3 (IRF3) and NF-kappaB in the innate immune response (57). TLR4 is well associated with cancer in many ways and its increased expression has been shown in various cancer cell lines and tissue samples derived from patients with head and neck, esophageal, gastric, colorectal, liver, pancreatic, skin, breast, ovarian and cervical cancers (58). TLR4 expression has been reported significantly elevated in HCC (59,60). LPS induced TLR4 signaling is involved in the invasion and metastasis of HCC and activates NF- κ B (61). MyD88-independent pathway, a pathway accountable for anti-viral responses, facilitates induction of the anti-viral type I IFNs (IFN $\alpha/\beta/\gamma$) and IFN-inducible genes through IRF3 (57). In one of the recent studies, TLR4 activity was found significantly associated with tumor differentiation grade and lymph node metastasis in esophageal squamous cell carcinoma (62). Thus, TLR-related pathways are definitely attractive therapeutic targets which we need to investigate profoundly in relation to effective HCC treatment.

Other important gene families

Up-regulated expression of transforming growth factor (TGF)- β 1 is associated with the risk of HCC. To identify if TCGA data represent the HCC subgroup triggered by up-regulation of TGF- β 1, a list of more than 300 genes that participate in TGF- β signaling pathway was prepared from different sources including wikipathway. Then, expression of these genes was evaluated from TCGA gene expression data (Supplementary file 6). Only few genes were found differentially expressed significantly in HCC compared to controls (Table 4). Out of 267 genes, 7 were over-expressed and 6 were under-expressed in case of HCC, suggesting that the TCGA data may not represent TGF- β 1 induced HCC cohort significantly. Further, genes associated with drug metabolism and cancer metabolism and those that belong to kinetochore and centromere families were sorted (Supplementary file 4). Out of 286 drug metabolism associated genes, only 21 were differentially expressed. Ten genes were significantly up-regulated (log2fc >2) and 11 genes were down-regulated (log2fc <-2) in HCC compared to control samples. Similarly, 5 of 17 cancer metabolism associated

genes were differentially expressed (4 up- and 1 down-regulated) in HCC. On the other hand, 11 of 18 genes of kinetochore and centromere families were over-expressed in HCC (Supplementary file 4).

Table 5 Set of top predictor genes for separation of hepatocellular carcinoma and controls. These predictors were identified from differentially expressed genes by employing supervised machine learning method based on estimation of random forest variable importance. Estimated AUC is almost 1

Gene symbol	Full name [HGNC Database*]
ANGPTL6	angiopoietin like 6
GABRD	gamma-aminobutyric acid type A receptor delta subunit
ADAMTS13	ADAM metalloproteinase with thrombospondin type 1 motif 13
VIPR1	vasoactive intestinal peptide receptor 1
ECM1	extracellular matrix protein 1
PTH1R	parathyroid hormone 1 receptor
SLC26A6	solute carrier family 26 member 6
OIT3	oncoprotein induced transcript 3
COL15A1	collagen type XV alpha 1 chain
CXCL12	C-X-C motif chemokine ligand 12
CDH13	cadherin 13
PLVAP	plasmalemma vesicle associated protein
FCN3	ficolin 3
LIFR	leukemia inhibitory factor receptor alpha
CFP	complement factor properdin
LOC222699(TOB2P1)	transducer of ERBB2, 2 pseudogene 1
AMIGO3	adhesion molecule with Ig-like domain 3
CPEB3	cytoplasmic polyadenylation element binding protein 3
ASPG	asparaginase
DBH	dopamine beta-hydroxylase
C9	complement component 9
AZI1 (CEP131)	centrosomal protein 131
APOF	apolipoprotein F

* <http://www.genenames.org/>.

Identification of HCC classifiers

All the differentially expressed genes were used to train the supervised model. The model was based on random forest which can efficiently separate two different populations (in our case, HCC and controls) using expression values of top predictor genes (Table 5). The model was trained after rigorous gene selection procedure based on random splitting of expression matrix in which minority samples were inflated according to SMOTE approach (13). Selection of top predictor genes was based on random forest variable importance estimation. The genes that appeared frequently in 100 iterations and scored maximum importance values were chosen as top predictors. A receiver operating characteristic (ROC) estimated the AUC value at almost 1. The top 4 predictors (ANGPTL6, GABRD, ADAMTS13 and VIPR1) appeared in most of the iterations. ANGPTL6 gene encodes for a secreted protein, angiopoietin like 6 and has been correlated with obesity, energy expenditure and diabetes (63,64); however, no studies in relation to any type of cancers have been reported. GABRD gene encodes for gamma-aminobutyric acid type A receptor delta subunit and was found over-expressed in 89% of HCC samples in this study. It was associated with functional changes in the GABAA receptor which might be essential for tumor cell differentiation (65). ADAMTS13 gene encodes for ADAM metalloproteinase with thrombospondin type 1 motif 13 protein. It controls von Willebrand factor (VWF) fiber formation that plays a prominent role in the cancer development (66,67). VIPR2

gene encodes for vasoactive intestinal peptide receptor 1 which is a cytokine and neuropeptide and has been positively correlated with immune responses (68). An important identifier was PLVAP gene. An expression study in paired tumor and adjacent non-tumor tissues has shown that PLVAP protein was expressed only in vascular endothelial cells of HCC but not in non-tumor liver tissues and suggested it as a potential target for treating HCC (69). Though these classifiers were predicted with rigorous dataset training and worked aptly in our case, there is a need to apply them cautiously on data other than TCGA by considering the fact that HCC data used to predict classifiers were highly heterogeneous and influenced by several varying factors including HCV, HBV, alcoholism, etc.

Identification of HCC subtypes and their features

The expression studies usually compare disease data with corresponding controls to recognize signaling pathways common to all disease samples, without considering the fact that there might be a robust heterogeneity among the disease samples and thus associated with different unique pathways. Classification of samples into disease subtypes can reduce that issue. To achieve the goal of precision medicine, it is quite important to classify patients into highly similar sub-groups to ascertain that patients would respond to a specific therapy in a similar way. Deciphering the similar genetic patterns among patient pools is thus crucial for potential and specific personalized

Figure 8 Silhouette width plot. Here, estimated silhouette widths for clusters 0 to 8 are 0.000, 0.374, 0.366, 0.414, 0.363, 0.352, 0.360 and 0.368. Silhouette width is maximum for $k = 4$. Thus, all of the hepatocellular carcinoma samples can likely be classified into 4 subtypes.

Figure 9 Heat map of four HCC subtypes with minimum overlap using unsupervised hierarchical consensus clustering (reps = 2000, distance = "Pearson") on 2714 most variable genes of 373 HCC patient samples. The four HCC subtypes designated as A, B, C and D represent 54, 193, 67 and 59 samples respectively.

therapeutics. There are several clustering methods available to predict subtypes. We have applied consensus clustering method based on unsupervised hierarchical clustering with Pearson correlation. Each subtype predicted attained a final cluster after going through bootstrapping ($n = 2000$). This method generates output for given number of subtypes ($k = 1$ to n) and final specific k subtypes were selected based on various statistical parameters, such as silhouette width, cumulative distribution function and the heat map plot showing minimum overlap between square blocks and coloring of the blocks (the deeper is the blue color of square block, the better the clustering is considered). Silhouette width analysis of HCC expression data estimated highest silhouette width for $k = 4$ (Figure 8). The heat map generated through stratification of HCC patients based on gene expression data using consensus unsupervised clustering method predicted four distinct subtypes, represented as HCC subtypes A, B, C, and D (Figure 9). The four perfect squared blue blocks of different sizes, representing four different HCC subtypes, accommodated all of the 373 HCC patient samples (Supplementary file 7).

The top genes belonging to each HCC subtype were extracted and pathway enrichment and survival analysis was performed (Supplementary file 9). The common gene sets were also listed and sorted in descending order of F-statistics (Supplementary file 8). T-box transcription factor TBX 3 (TBX3), which was the top common candidate, has been previously reported to be associated with TGF- β signaling, angiogenesis and cancer growth (70,71). Enrichment analysis of top genes of all four predicted HCC subtypes showed enrichment of different and specific pathway terms for each subtypes

Figure 10 Venn diagram showing the number of unique Reactome pathways for each type of HCC subtype and common pathways shared by them. Subtype A does not share any pathway with other subtypes, making it a completely distinct HCC subtype. Contrary to this, subtypes C and D shared maximum numbers of pathways ($n = 22$).

(Figure 10; Supplementary file 9). In particular, HCC subtype A did not share any biological pathway with any of the other subtypes. Survival analysis was performed over the clinical data linked with these subtypes using coxph model and alive/dead events were associated with each HCC subtype. Overall median survival was predicted in days for each subtype and it was found that patients categorized to subtype D have higher chance of survival (Figure 11).

As mentioned above, HCC subtype A was completely distinct from other subtypes. Pathway enrichment analysis of the top genes of subtype A showed a complete separation from others. Overall, 102 pathways were enriched and dominated by cancer- and immune-related pathways. More specifically, 11 kinases related genes (PDGFRA, MAP2K3, WEE1, PLK3, MAP2K1, LRRK2, PRKCB, TEK, PDK4, AXL and SGK1) were associated. Investigation of these pathways and associated genes may possibly help us to target and find specific treatment strategies for the HCC subtype A patients.

In the case of HCC subtype B, 20 pathways were enriched, predominantly Wnt and β -catenin related pathways. Degradation of β -catenin by the destruction complex, binding of TCF/LEF: CTNNB1 to target gene promoters and repression of Wnt target genes are among the pathway terms found enriched in this subtype. Hence, targeting Wnt/ β -catenin signaling pathway might be the most attractive target for finding potent treatment of HCC subtype B patients.

In HCC subtype C, 52 pathways were enriched. It is important to note that these included mainly drug metabolism and oxidation pathways associated with a number of genes, including UGT2B10, GSTM2, UGT1A1, GSTO2, ADH1C, ADH1B, ACSM1, GLYAT, CYP3A4, CYP8B1, CYP39A1, GSTZ1, ADH4, CYP2A6, CYP2C8, NAT2, CYP11A1, UGT1A4, UGT1A3 and UGT2B7. Any disturbance in expression of these genes may affect the clearance of the drug and/or may be responsible for undesirable side effects as well as drug toxicity. Another important category of pathways enriched was

Figure 11 Overall survival analysis of hepatocellular carcinoma patients. KM survival analysis of four HCC subtypes (A, B, C and D) was performed based on clinical data of 373 HCC patient samples. Vertical axis drawn in the plot shows the overall median survival (in days).

PI3K pathways, including PI3K cascade, PI3K/AKT signaling in cancer and negative regulation of PI3K/AKT network. Associated genes were FGF2, FGFR2 and FGFR3. A recent study has suggested that PI3K and RAS signaling pathways are the two most important pathways in cancers of the head and neck, lung and kidney (72). In many cancers, PI3K pathway associated genes and proteins have been widely investigated as therapeutic targets (73). Thus, achieving proper regulation of these genes would be essential for efficient treatment of HCC subtype C patients.

A total of 31 pathways were enriched in HCC subtype D. “Interleukin-19, 20, 22, 24” was one of the important pathways and its associated gene was interleukin-34 (IL-34). IL-34 is a known ligand which binds to colony-stimulating factor-1 receptor (CSF-1R) and enhances monocytes and macrophages survival. Recently, IL-34 over-expression had been shown in the sera of HCV patients with advanced liver fibrosis (74). Also, IL-34 expression levels in HCC were linked with overall survival and tumor recurrence rates (75). Another important pathway was PI3K cascade associated with kinase gene PRKAA2 (AMPK Subunit Alpha-2). It controls the levels of α -ketoglutarate and isocitrate dehydrogenase enzyme production, and stabilizes HIF-1 alpha and neutrophil survival in order to support the tumor progression. Interestingly, one of the dominant pathways enriched was biological oxidation. Important gene associated with this pathway was NNMT (Nicotinamide N-Methyltransferase) which encodes for an enzyme associated with kidney cancer and Parkinson disease. It has known implications of sinking the methylation potential of cancer cells (76). Consequently, NNMT-expressing cancer

cells alter an epigenetic state in such a way that histones and other cancer-related proteins underwent hypomethylation. It has been reported that patients with over-expressed NNMT have a smaller overall survival ($p = 0.053$) and a shorter disease-free survival ($p = 0.016$) (77). Thus, these pathways and associated genes can be investigated to find improved therapeutics for HCC subtype D patients.

Conclusions

In conclusion, this study presents comprehensive analysis of HCC transcriptomics data retrieved from TCGA. The gene sorting based on the expression and associated clinical information has exposed a number of interesting genes that can be exploited in future for diagnostic purposes as well as for improved and specific HCC therapy. The top ranked genes, specifically those associated with kinases, cancer- and immune-related pathways, can serve as potential diagnostic as well as prognostic markers of HCC. The mRNA based classifiers have clearly separated all the HCC samples from controls; however, further validation is required as classifiers were trained on a single source data, TCGA. This can be achieved by testing classifiers on the other independent datasets. Four HCC subtypes (designated here as A, B, C, D) were identified using classifier genes. The HCC subtype A was unexpectedly distinct which was dominantly enriched with kinases and the pathways enriched among top genes were surprisingly all unique. On the other hand, patients of subtype D have a higher probability of survival according to clinical data. Overall, this study

identifies heterogeneity among HCC samples and suggests subtype specific therapeutic targets. This study can be advanced to epigenetic regulation analysis, which will strengthen these findings in search of potential early stage markers as well as patient/subtype specific therapeutics for HCC.

Conflict of interest

The authors of this article declare that they have no conflicts of interest.

Acknowledgments

This study was performed over mRNA expression data of hepatocellular carcinoma retrieved from The Cancer Genome Atlas (TCGA), a collaboration between the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI). Authors thank TCGA for making their genomic data publically available.

Supplementary data

Supplementary data related to this article can be found online at [doi:10.1016/j.cancergen.2017.06.002](https://doi.org/10.1016/j.cancergen.2017.06.002).

References

1. Bosch FX, Ribes J, Diaz M, et al. Primary liver cancer: worldwide incidence and trends. *Gastroenterology* 2004;127:S5–S16.
2. El-Serag HB. Hepatocellular carcinoma. *N Engl J Med* 2011;365:1118–1127.
3. Parkin DM, Bray F, Ferlay J, et al. Global cancer statistics, 2002. *CA Cancer J Clin* 2005;55:74–108.
4. El-Serag HB. Hepatocellular carcinoma and hepatitis C in the United States. *Hepatology* 2002;36:S74–S83.
5. Okuda K, Ohtsuki T, Obata H, et al. Natural history of hepatocellular carcinoma and prognosis in relation to treatment study of 850 patients. *Cancer* 1985;56:918–928.
6. Katyal S, Oliver JH III, Peterson MS, et al. Extrahepatic metastases of hepatocellular carcinoma 1. *Radiology* 2000;216:698–703.
7. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011;12:323.
8. Wilcoxon F. Individual comparisons of grouped data by ranking methods. *J Econ Entomol* 1946;39:269.
9. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Stat Med* 1990;9:811–818.
10. Huang DW, Sherman BT, Tan Q, et al. DAVID bioinformatics resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res* 2007;35:W169–W175.
11. Fabregat A, Sidiropoulos K, Garapati P, et al. The reactome pathway knowledgebase. *Nucleic Acids Res* 2016;44:D481–D487.
12. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 2015;43:D447–D452.
13. Blagus R, Lusa L. SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics* 2013;14:106.
14. Slawski M, Daumer M, Boulesteix A-L. CMA—a comprehensive Bioconductor package for supervised classification with high dimensional data. *BMC Bioinformatics* 2008;9:439.
15. Archer KJ, Kimes RV. Empirical characterization of random forest variable importance measures. *Comput Stat Data Anal* 2008;52:2249–2260.
16. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 2010;26:1572–1573.
17. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 1987;20:53–65.
18. Maechler M, Rousseeuw P, Struyf A, et al. Cluster: cluster analysis basics and extensions. 2013.
19. Schwender H. Siggenes: multiple testing using SAM and Efron's empirical Bayes approaches. 2012.
20. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958;53:457–481.
21. Cox DR. Regression models and life tables. *J R Stat Soc B* 1972;34:187–220.
22. Tsuchiya N, Sawada Y, Endo I, et al. Biomarkers for the early diagnosis of hepatocellular carcinoma. *World J Gastroenterol* 2015;21:10573–10583.
23. Förster S, Gretscher S, Jöns T, et al. THBS4, a novel stromal molecule of diffuse-type gastric adenocarcinomas, identified by transcriptome-wide expression profiling. *Mod Pathol* 2011;24:1390–1403.
24. Asimakopoulou A, Weiskirchen S, Weiskirchen R. Lipocalin 2 (LCN2) expression in hepatic malfunction and therapy. *Front Physiol* 2016;7.
25. Haruyama Y, Kataoka H. Glypican-3 is a prognostic factor and an immunotherapeutic target in hepatocellular carcinoma. *World J Gastroenterol* 2016;22:275.
26. McClelland ML, Gardner RD, Kallio MJ, et al. The highly conserved Ndc80 complex is required for kinetochore assembly, chromosome congression, and spindle checkpoint activity. *Genes Dev* 2003;17:101–114.
27. Harao M, Hirata S, Irie A, et al. HLA-A2-restricted CTL epitopes of a novel lung cancer-associated cancer testis antigen, cell division cycle associated 1, can induce tumor-reactive CTL. *Int J Cancer* 2008;123:2616–2625.
28. Liu Q, Dai S-J, Li H, et al. Silencing of NUF2 inhibits tumor growth and induces apoptosis in human hepatocellular carcinomas. *Asian Pac J Cancer Prev* 2013;15:8623–8629.
29. Kagohara LT, Kulkarni P, Shiraishi T, et al. Cancer/testis antigen expression pattern is a potential biomarker for prostate cancer aggressiveness. *Cancer Res* 2015;75:4826.
30. Ferrín G, Ranchal I, Llamaza C, et al. Identification of candidate biomarkers for hepatocellular carcinoma in plasma of HCV-infected cirrhotic patients by 2-D DIGE. *Liver Int* 2014;34:438–446.
31. Kijima H, Sawada T, Tomosugi N, et al. Expression of hepcidin mRNA is uniformly suppressed in hepatocellular carcinoma. *BMC Cancer* 2008;8:1.
32. Costa-Matos L, Batista P, Monteiro N, et al. Liver hepcidin mRNA expression is inappropriately low in alcoholic patients compared with healthy controls. *Eur J Gastroenterol Hepatol* 2012;24:1158–1165.
33. Chang Q, Chen J, Beezhold KJ, et al. JNK1 activation predicts the prognostic outcome of the human hepatocellular carcinoma. *Mol Cancer* 2009;8:1.
34. McKenzie L, King S, Marcar L, et al. p53-dependent repression of polo-like kinase-1 (PLK1). *Cell Cycle* 2010;9:4200–4212.
35. Hu J, Wang G, Liu X, et al. Polo-like kinase 1 (PLK1) is involved in toll-like receptor (TLR)-mediated TNF- α production in monocytic THP-1 cells. *PLoS ONE* 2013;8:e78832.
36. Bussey KJ, Bapat A, Linnehan C, et al. Targeting polo-like kinase 1, a regulator of p53, in the treatment of adrenocortical carcinoma. *Clin Transl Med* 2016;5:1.

37. Gjertsen BT, Schöffski P. Discovery and development of the Polo-like kinase inhibitor volasertib in cancer therapy. *Leukemia* 2015;29:11–19.
38. Mao J, Yu H, Wang C, et al. Metallothionein MT1M is a tumor suppressor of human hepatocellular carcinomas. *Carcinogenesis* 2012;33:2568–2577.
39. Park Y, Yu E. Expression of metallothionein-1 and metallothionein-2 as a prognostic marker in hepatocellular carcinoma. *J Gastroenterol Hepatol* 2013;28:1565–1572.
40. Sun X, Niu X, Chen R, et al. Metallothionein-1G facilitates sorafenib resistance through inhibition of ferroptosis. *Hepatology* 2016;64:488–500.
41. Ji X-F, Fan Y-C, Gao S, et al. MT1M and MT1G promoter methylation as biomarkers for hepatocellular carcinoma. *World J Gastroenterol* 2014;20:4723.
42. Ding J, Lu S. Low metallothionein 1M expression association with poor hepatocellular carcinoma prognosis after curative resection. *Genet Mol Res* 2016;15.
43. Kara N, Hossain M, Prasanth SG, et al. Orc1 binding to mitotic chromosomes precedes spatial patterning during G1 phase and assembly of the origin recognition complex in human cells. *J Biol Chem* 2015;290:12355–12369.
44. Cai J, Li B, Zhu Y, et al. Prognostic biomarker identification through integrating the gene signatures of hepatocellular carcinoma properties. *EBioMedicine* 2017;19:18–30.
45. Amon A. The spindle checkpoint. *Curr Opin Genet Dev* 1999; 9:69–75.
46. Gayyed MF, El-Maqsoud NM, Tawfik ER, et al. A comprehensive analysis of CDC20 overexpression in common malignant tumors from multiple organs: its correlation with tumor grade and stage. *Tumour Biol* 2016;37:749–762.
47. Woo Seo D, Yeop You S, Chung WJ, et al. Zwint-1 is required for spindle assembly checkpoint function and kinetochore-microtubule attachment during oocyte meiosis. *Sci Rep* 2015; 5:15431.
48. Liang X-D, Dai Y-C, Li Z-Y, et al. Expression and function analysis of mitotic checkpoint genes identifies TTK as a potential therapeutic target for human hepatocellular carcinoma. *PLoS ONE* 2014;9:e97739.
49. Xie Y, Wang A, Lin J, et al. Mps1/TTK: a novel target and biomarker for cancer. *J Drug Target* 2017;25:112–118.
50. Xia H, Kong SN, Chen J, et al. MELK is an oncogenic kinase essential for early hepatocellular carcinoma recurrence. *Cancer Lett* 2016;383:85–93.
51. Buchbinder EI, Desai A. CTLA-4 and PD-1 pathways: similarities, differences, and implications of their inhibition. *Am J Clin Oncol* 2016;39:98–106.
52. Bowyer S, Prithviraj P, Lorigan P, et al. Efficacy and toxicity of treatment with the anti-CTLA-4 antibody ipilimumab in patients with metastatic melanoma after prior anti-PD-1 therapy. *Br J Cancer* 2016;114:1084–1089.
53. Jain A, Kaczanowska S, Davila E. IL-1 receptor-associated kinase signaling and its role in inflammation, cancer progression, and therapy resistance. *Front Immunol* 2014;5:553.
54. Rhyasen GW, Bolanos L, Fang J, et al. Targeting IRAK1 as a therapeutic approach for myelodysplastic syndrome. *Cancer Cell* 2013;24:90–104.
55. Srivastava R, Geng D, Liu Y, et al. Augmentation of therapeutic responses in melanoma by inhibition of IRAK-1,-4. *Cancer Res* 2012;72:6209–6216.
56. Barbie TU, Alexe G, Aref AR, et al. Targeting an IKBKE cytokine network impairs triple-negative breast cancer growth. *J Clin Invest* 2014;124:5411–5423.
57. Fitzgerald KA, McWhirter SM, Faia KL, et al. IKKepsilon and TBK1 are essential components of the IRF3 signaling pathway. *Nat Immunol* 2003;4:491–496.
58. Mai CW, Kang YB, Pichika MR. Should a Toll-like receptor 4 (TLR-4) agonist or antagonist be designed to treat cancer? TLR-4: its expression and effects in the ten most common cancers. *Oncotargets Ther* 2013;6:1573–1587.
59. Yang J. Emerging role of Toll-like receptor 4 in hepatocellular carcinoma. *J Hepatocell Carcinoma* 2015;2:11–17.
60. Nishimura M, Naito S. Tissue-specific mRNA expression profiles of human toll-like receptors and related genes (Biopharmacy). *Biol Pharm Bull* 2005;28:886–892.
61. Jing Y-Y, Han Z-P, Sun K, et al. Toll-like receptor 4 signaling promotes epithelial-mesenchymal transition in human hepatocellular carcinoma induced by lipopolysaccharide. *BMC Med* 2012;10:98.
62. Zu Y, Ping W, Deng T, et al. Lipopolysaccharide-induced toll-like receptor 4 signaling in esophageal squamous cell carcinoma promotes tumor proliferation and regulates inflammatory cytokines expression. *Dis Esophagus* 2017;30:1–8.
63. Kadomatsu T, Tabata M, Oike Y. Angiopoietin-like proteins: emerging targets for treatment of obesity and related metabolic diseases. *FEBS J* 2011;278:559–564.
64. Verdeguer F, Soustek MS, Hatting M, et al. Brown adipose YY1 deficiency activates expression of secreted proteins linked to energy expenditure and prevents diet-induced obesity. *Mol Cell Biol* 2016;36:184–196.
65. Gross AM, Kreisberg JF, Ideker T. Analysis of matched tumor and normal profiles reveals common transcriptional and epigenetic signals shared across cancer types. *PLoS ONE* 2015;10: e0142618.
66. Pépin M, Kleinjan A, Hajage D, et al. ADAMTS-13 and von Willebrand factor predict venous thromboembolism in patients with cancer. *J Thromb Haemost* 2016;14:306–315.
67. Borsig L. VWF fibers induce thrombosis during cancer. *Blood* 2015;125:3042–3043.
68. Olson KE, Kosloski-Bilek LM, Anderson KM, et al. Selective VIP receptor agonists facilitate immune transformation for dopaminergic neuroprotection in MPTP-intoxicated mice. *J Neurosci* 2015;35:16463–16478.
69. Wang Y-H, Cheng T-Y, Chen T-Y, et al. Plasmalemmal Vesicle Associated Protein (PLVAP) as a therapeutic target for treatment of hepatocellular carcinoma. *BMC Cancer* 2014;14: 815.
70. Li J, Weinberg MS, Zerbini L, et al. The oncogenic TBX3 is a downstream target and mediator of the TGF- β 1 signaling pathway. *Mol Biol Cell* 2013;24:3569–3576.
71. Perkhof L, Walter K, Costa IG, et al. Tbx3 fosters pancreatic cancer growth by increased angiogenesis and activin/nodal-dependent induction of stemness. *Stem Cell Res* 2016;17:367–378.
72. Sikdar S, Datta S, Datta S. Exploring the importance of cancer pathways by meta-analysis of differential protein expression networks in three different cancers. *Biol Direct* 2016;11:65.
73. Gysin S, Salt M, Young A, et al. Therapeutic strategies for targeting Ras proteins. *Genes Cancer* 2011;1947601911412376.
74. Shoji H, Yoshio S, Mano Y, et al. Interleukin-34 as a fibroblast-derived marker of liver fibrosis in patients with non-alcoholic fatty liver disease. *Sci Rep* 2016;6.
75. Zhou SL, Hu ZQ, Zhou ZJ, et al. miR-28-5p-IL-34-macrophage feedback loop modulates hepatocellular carcinoma metastasis. *Hepatology* 2016;63:1560–1575.
76. Ulanovskaya OA, Zuhl AM, Cravatt BF. NNMT promotes epigenetic remodeling in cancer by creating a metabolic methylation sink. *Nat Chem Biol* 2013;9:300–306.
77. Kim J, Hong SJ, Lim EK, et al. Expression of nicotinamide N-methyltransferase in hepatocellular carcinoma is associated with poor prognosis. *J Exp Clin Cancer Res* 2009;28:1.