Med Oncol (2015) 32:96
DOI 10.1007/s12032-014-0425-6

ORIGINAL PAPER

Meta-analysis of gene expression profiles indicates genes in spliceosome pathway are up-regulated in hepatocellular carcinoma (HCC)

Weijin Xu · Huixing Huang · Long Yu · Lihuan Cao

Received: 28 November 2014 / Accepted: 29 November 2014 / Published online: 3 March 2015

Abstract Hepatocellular carcinoma (HCC) is among the commonest kind of malignant tumors, which accounts for more than 500,000 cases of newly diagnosed cancer annually. Many microarray studies for identifying differentially expressed genes (DEGs) in HCC have been conducted, but results have varied across different studies. Here, we performed a meta-analysis of publicly available microarray Gene Expression Omnibus datasets, which covers five independent studies, containing 753 HCC samples and 638 non-tumor liver samples. We identified 192 DEGs that were consistently up-regulated in HCC vs. normal liver tissue. For the 192 up-regulated genes, we performed Kyoto Encyclopedia of Genes and Genomes pathway analysis. To our surprise, besides several cell growth-related pathways, spliceosome pathway was also up-regulated in HCC. For further exploring the relationship between spliceosome pathway and HCC, we investigated the expression data of spliceosome pathway genes in 15 independent studies in Nextbio database (https://www.nextbio.com/b/nextbioCorp.nb). It was found that many genes of spliceosome pathway such as HSPA1A, SNRPE, SF3B2, SF3B4 and TRA2A genes which we identified to be up-regulated in our meta-analysis were generally overexpressed in HCC. At last, using real-time PCR, we also found that BUD31, SF3B2, SF3B4, SNRPE, SPINK1, TPA2A and HSPA1A genes are significantly up-regulated in clinical HCC samples when compared to the corresponding non-tumorous liver tissues. Our study for the first time indicates that many genes of spliceosome pathway are up-regulated in HCC. This finding might put new insights for people's understanding about the relationship of spliceosome pathway and HCC.

Keywords Meta-analysis · Hepatocellular carcinoma (HCC) · Spliceosome pathway

Weijin Xu and Huixing Huang have contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s12032-014-0425-6) contains supplementary material, which is available to authorized users.

W. Xu · H. Huang · L. Yu (✉) · L. Cao (✉)
State Key Laboratory of Genetic Engineering, Institute of Genetics, School of Life Sciences, Fudan University, 220 Handan Road, Shanghai 200433, People's Republic of China
e-mail: longyu@fudan.edu.cn

L. Cao
e-mail: lihuancao@fudan.edu.cn

W. Xu
e-mail: xuweijin1991@gmail.com

H. Huang
e-mail: 11210700090@fudan.edu.cn

Introduction

Hepatocellular carcinoma (HCC) is among the commonest kind of malignant tumors, which accounts for more than 500,000 cases of newly diagnosed cancer annually [1]. The high death rate of HCC makes it the second most notorious killer of patients comparing to all other cancers. Even though the exact cause of HCC is unknown, extensive investigations indicate strong correlation between HCC and genomic alterations [2]. These genomic alterations include both genetic abnormalities and epigenetic alterations [3, 4]. Whereas the traditional methods to study hepatocarcinogenesis based on single gene mutation, modern approaches utilize pathway analysis as the breakthrough point [5]. Pathway alterations are believed to be a more accurate, reasonable and comprehensive description of oncogenesis than single gene alterations, which bring oncogenesis study to a whole new level [6].

However, before the invention of microarray technology, correlation between genes and cancer can only be verified in a time-consuming and highly specialized fashion. Genechips make large-scale and high-throughput gene screening a possible task [7]. Different genechips are targeted for various purposes, including gene mutation, copy-number and expression test. In our study, we focus on gene expression profiling array in HCC. A significant increase (or decrease) in mRNA concentration in tumor samples indicates protein expression alteration during tumor transformation or progression [8].

DNA microarray technology provides a searching tool for differentially expressed genes (DEGs) by investigating the gene expression profile, but the results had varied across different studies. Meta-analysis is referred to as summary analysis, which integrates information from multiple studies. Meta-analysis could overcome the limitation of small sample sizes, rare outcomes and poor quality of trials [9]. With the increasing number of public available microarray datasets, meta-analysis of multiple datasets has been widely used and proven to be a useful method in searching DEGs in cancers [10–13].

In this study, we used open-source statistics software R to meta-analyze gene expression profiling data of five independent studies covering 753 HCC samples and 638 non-tumor liver samples in total. We found that mRNA level of several pathways was up-regulated in HCC, including ribosome, proteasome and other cell growth-related pathways. These pathways are expected to be up-regulated since cancer cells are known for their high growth rate. Surprisingly, we found that spliceosome pathway was also up-regulated in HCC. We further explored Nextbio database for more information about spliceosome pathway genes overexpression in HCC. It was found that many genes of spliceosome pathway were generally overexpressed in HCC.

Spliceosome is the ribonucleoprotein (RNPs) machine removing noncoding introns from precursor messenger RNAs (pre-mRNAs), which is comprised of more than 200 kinds of protein components [14]. Dramatic compositional changes in spliceosome can be observed during assembly and splicing process [15]. Any aberrant change of spliceosome constituents may cause splicing defects or alterations, which are associated with many human diseases. Our study, for the first time, indicates that many spliceosome pathway genes are up-regulated in HCC.

Materials and methods

Data source

Genechip data used in our research are searched in and downloaded from NCBI Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) database using keyword 'hepatocellular carcinoma.' All mRNA expression profiling data series involved in this study were published results and available to public. To guarantee the fidelity of this research, we screened these studies for those with *Homo sapiens* HCC sample size more than 50. After all these scrutiny, we finally identified five series of gene expression microarray data as our analyzing subject [16–23]. Among these datasets, GSE4024 was designed as control group for GSE1898 in the original study, which exclusively contains non-tumor samples. We consider GSE4024 and GSE1898 to be one dataset for convenience in later reference. Detailed sample information can be seen in Table 1.

Preprocessing data

Statistical analyzing software utilized in this research is R (http://www.R-project.org/) along with several additional packages available in bioconductor (http://www.bioconductor.org). Package *GEOquery* was used to download genechip data. *GetGEO*, accessible in *GEOquery* package, is the algorithm used to download data from GEO database and convert data into data frame readable by R.

After integrating data from the same study into a single matrix, preprocessing is necessary before further analysis. First of all, many missing values were created in microarrays. k-Nearest neighbors (KNN) method imputes the missing values by averaging non-missing values of its neighbor, which is found by using Euclidean metric. We chose KNN method to perform data imputation because of its efficiency and accuracy [24]. Package impute available in bioconductor includes an algorithm impute.knn using KNN method to impute missing values.

Secondly, due to uncontrollable factors, values of the same gene on different genechips vary significantly. This is a serious problem obstructing cross-analysis between genechips. Fortunately, normalization can be a possible solution. Linear model for microarray analysis package is

Table 1 HCC gene expression datasets used in meta-analysis of microarrays

GEO series	HCC sample	Non-tumor sample	Total
GSE1898	182	0	182
GSE4024	0	98	98
GSE14520	247	241	488
GSE14811	56	56	112
GSE25097	268	243	511
Total	753	638	1,391

All the studies used are solely focusing on expression difference between HCC samples and normal liver samples without concerning different stages of the cancer or any other related factors

one of the most commonly used normalization package in Bioconductor, which is built upon Bayes Linear method [25]. We used algorithm normalizeBetweenArrays to perform normalization between different genechips in the same series. 'Scale' is chosen as the method for normalization. This step ensures that later analysis will not be influenced by factors other than expression level of genes.

Profiling of abnormally expressed genes

In order to determine which genes are up-regulated in HCC samples, we performed Welch two sample *t* test between tumor and non-tumor samples within series. We did not choose paired *t* test because these researches were not designed as paired study originally. In order to overcome the great number of false-positive results due to multiple *t* test, adjusted *P* value was introduced. By using p.adjust algorithm in R, we used false discovery rate method, a relatively strict but more tolerant method than Bonferroni, to control the test at 0.05 significance level [26]. The adjustment means that only 5 % of the positive results might be false-positive.

Having lists of genes which are up-regulated in each series, we cross-matched these lists to identify common up-regulated gene expression. That is to say, we searched for gene with expression aberration in all four series. If a positive result is shared by four series, there will be only 6.25×10^{-6} probability left for it to be false-positive.

Pathway analysis

The long list of genes can hardly be interpreted into characteristics of HCC unless categorized into different pathways. We chose DAVID (http://david.abcc.ncifcrf.gov) [27] to perform Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis [28]. KEGG pathway is the most utilized and dependable pathway database in the world. Pathway analysis functions by identifying several most relevant pathways according to the list of genes. After the analysis, we can not only see which pathways are involved in HCC, but also know which genes are being expressed aberrantly. In our case, spliceosome and five others pathways were identified by KEGG pathway analysis.

Analysis expression data of spliceosome genes in Nextbio database

For further information, we collected expression data of spliceosome genes in Nextbio database (https://www.nextbio.com/b/nextbioCorp.nb). We searched gene names in 'Disease Atlas' section in Nextbio database and used filter to select RNA expression studies in human. All these expression data were generated from expression profiling array between HCC and normal liver samples. Fifteen studies and 25 comparisons in total are selected and summarized in Fig. 2.

Human tissue samples

Primary HCC tissues and the corresponding non-tumorous liver tissues were freshly collected from HCC patients who underwent hepatectomy at Zhongshan Hospital (Shanghai, China). Written informed consent was obtained from each patient before tissue acquisition. Samples were snap-frozen in liquid nitrogen immediately after surgery and stored at −80 °C for further use.

Reverse transcription polymerase chain reaction and real-time polymerase chain reaction (RT-PCR)

Total RNA was extracted from HCC samples with Trizol reagent (Invitrogen), and the first-strand cDNA was synthesized using ReverTra Ace qPCR RT Kit (TOYOBO) following the manufacturers' instructions. PCR amplification reactions were performed by an iCycler detection system (Bio-Rad). Gene expression levels were normalized to that of the housekeeping gene β2-microglobulin (β2-MG). Primers for β2-MG were as follows: sense, 5′-ATGAG TATGCCTGCCGTGTGAAC-3′ and antisense, 5′-TGTG GAGCAACCTGCTCAGATAC-3′. Primers for human HSPA1A were as follows: sense, 5′-AGCTGGAGCAGGTG TGTAAC-3′ and antisense, 5′-CAGCAATCTTGGAAAG GCCC-3′; for human SF3B2 were as follows: sense, 5′-AC CAAGACTGAGGAAGAAGAGAT-3′ and antisense, 5′-T CCAGCAGGCACTGATGA-3′; for human SF3B4 were as follows: sense, 5′-CAGCACCAAGGCTATGGCT-3′ and antisense, 5′-TTCACCCGTATTGGCTTCCC-3′; for human SPINK1 were as follows: sense, 5′-TTCAACTGA CCTCTGGACGC-3′ and antisense, 5′-AGAAGTCTG GCGTTTCCGAT-3′; for human BUD31 were as follows: sense, 5′-ATCCACCACCAGAAAACCCG-3′ and antisense, 5′-CTCGATGATGCGGCCCACTT-3′; for human PRPF19 were as follows: sense, 5′-CCCGTCTCACCA AGGAAGT-3′ and antisense, CTTCCCTCTCTTCTTG CGCT; for human TRA2A were as follows: sense, 5′-CA GGCATCTACATGGGCAGA-3′ and antisense, 5′-AGGT GATCGTCTTCTGTATCGG-3′; for human SNRPE were as follows: sense, 5′-CGGATTCAGGTGTGGCTCTA-3′ and antisense, 5′-TGATCCGACCCAGTTGTTTTCT-3′. The relative gene mRNA expression in paired HCC samples was calculated as previously described [29].

Results

Common gene expression alteration and pathway analysis results

After performing *t* test using gene expression data from four series, we identified genes being overexpressed in HCC samples in each series. Each datasets yielded thousands of abnormal genes. However, when we performed gene-list cross-matching by their accession number between all possible three or four series combination of the four series, the number came down to hundreds. This phenomenon indicates that even though many genes are considered to be abnormal according to a single experiment, only tiny portion of these genes is constantly up-regulated in different HCC samples. Detailed results can be seen in Fig. 1. Steps above were performed on R. Then, common up-regulated genes were sent to KEGG pathway analysis in DAVID. All of the 192 gene names are recognized by DAVID (Supplementary Data File—Table S1 and Table S2: sheet1—Table S1). Output includes six pathways with *P* value less than default cutoff *P* value set by DAVID (<0.1). Three pathways have *P* value smaller than 0.05, which are considered significantly altered in HCC samples, including ribosome, spliceosome and nucleotide excision repair (Table 2).

The expression alteration information of related spliceosome genes in Nextbio database

Next, to ensure our finding is not confined to these 735 HCC samples, we searched in Nextbio database for expression information of several spliceosome genes between HCC and normal liver tissue [30]. Based on our previous meta-analysis results, genes of spliceosome pathway, BUD31 homolog, HSPA1A, SNRPE, SF3B2, SF3B4, TRA2A and PRPF19, were selected for further analysis. To serve as a benchmark, SPINK, which is previously reported

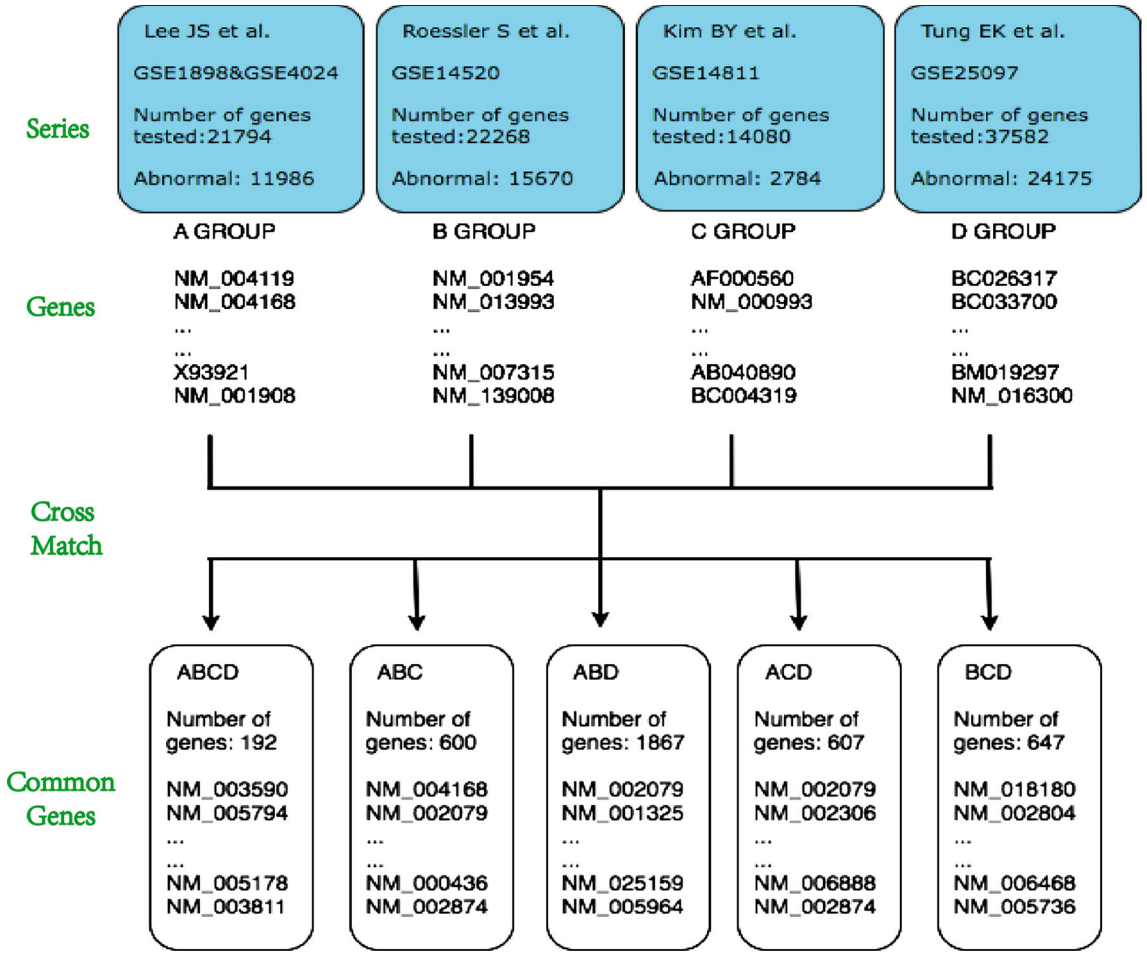


Fig. 1 Flow chart of the meta-analysis. First, Welch *t* test was performed between each gene in every HCC and normal samples in each datasets. Second, the generated lists of up-regulated genes are assigned to inter-study combination to pick out overlapping genes. The combination produced lists of common up-regulated genes

Table 2 Overexpressed KEGG pathways in HCC

Pathway	KEGG ID	Count	Percentage	*P*-value
Ribosome	map03010	7	3.6	5.4E−03
Spliceosome	map03040	7	3.6	5.4E−02
Nucleotide excision repair	map03420	4	2.1	4.8E−02
Proteasome	map03050	4	2.1	5.7E−02
Neurotrophin signaling	map04720	6	3.1	8.1E−02
RNA polymerase	map03020	3	1.6	9.5E−02

The pathways are significant up-regulated according to DAVID KEGG pathway analysis of lists of up-regulated genes. The percentage stands for the percentage of number of up-regulated genes of this pathway in the list. The *P* value stands for EASE score, a more conservative examining value than Fisher exact *P* value. The cutoff value we use is 0.1 in this analysis

to be highly up-regulated in HCC and believed to be one HCC marker [31], is also included in analysis. It was found that significant overexpression of HSPA1A, SNRPE, SF3B2, SF3B4 and TRA2A genes can be observed in a great portion of 15 studies, while overexpression of BUD31 homolog and PRPF19 is less frequent (Fig. 2). The detail information of Fig. 2 could be found Supplementary Data File—Table S1 and Table S2: sheet2—Table S2. On the other hand, though indicated as HCC marker gene, SPINK1 overexpression in HCC was found in similar or less number of studies comparing to the five genes. Thus, we thought that overexpression of spliceosome pathway genes: HSPA1A, SNRPE, SF3B2, SF3B4 and TRA2A are strongly associated with HCC. BUD31 homolog and PRPF19 genes are also overexpressed certain HCC

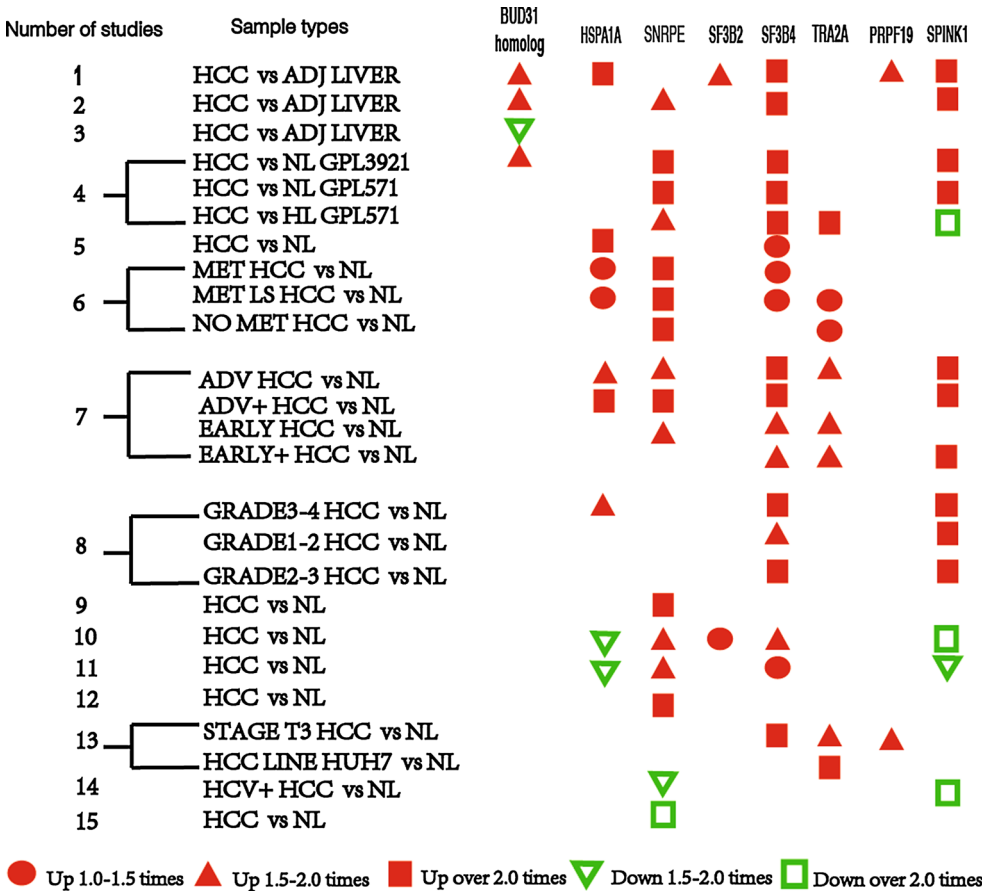
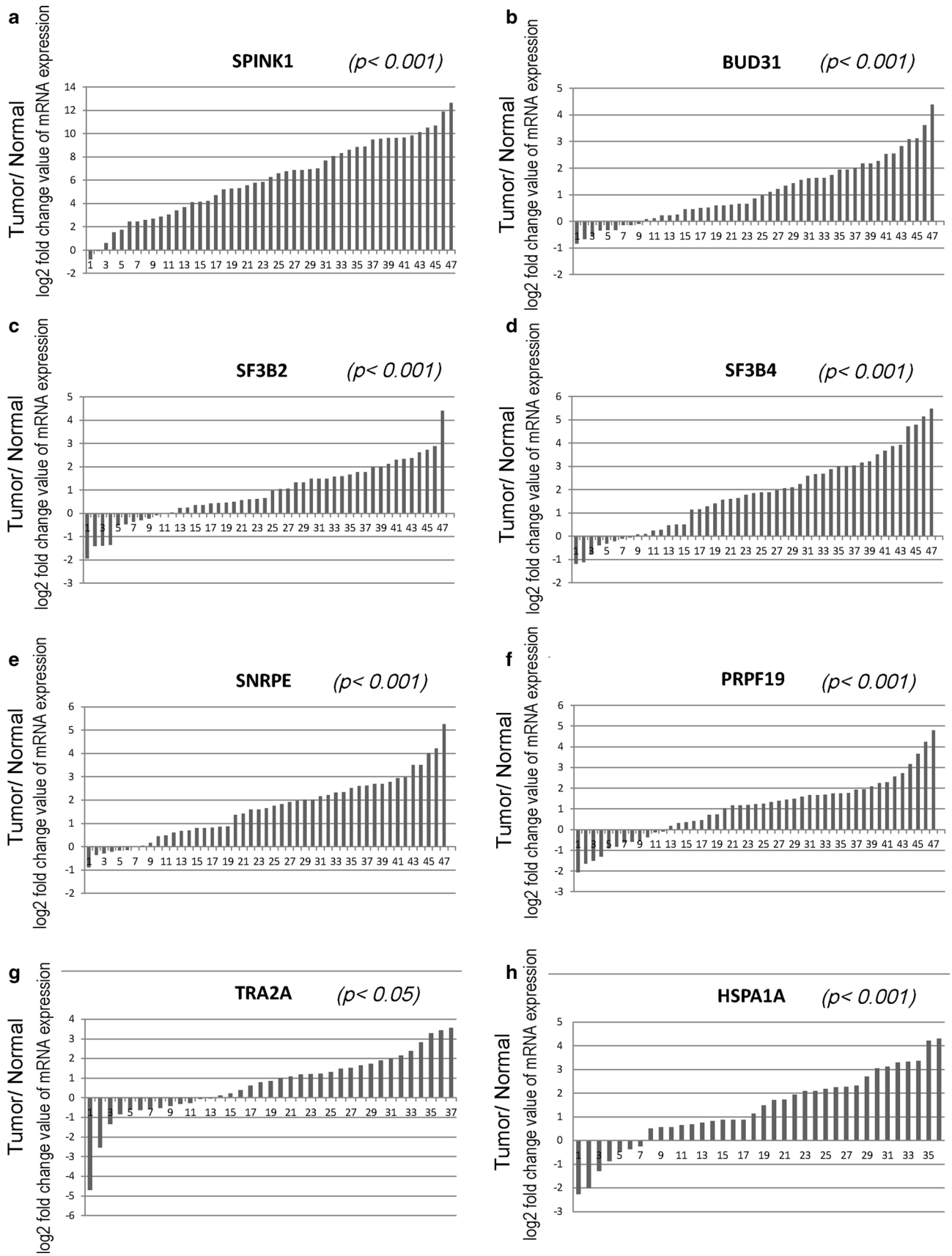


Fig. 2 Selected spliceosome pathway genes expression profiles in more HCC expression studies. All HCC and normal liver expression comparison studies available in Nextbio database are summarized in this figure. *Stuffed pattern* refers to up-regulation of the genes, while *hollow pattern* refers to down-regulation. The *shapes* of the pattern represent relative expression level of the genes in HCC comparing to normal liver. *Circle* refers to expression level 1–1.5 times difference. *Triangle* refers to 1.5–2.0 times and *square* refers to over 2.0 times difference. For example, *stuffed square* means over 2 times overexpression of this gene in HCC sample. Positions *left* for *blank* refers to no significant difference in expression between HCC and normal liver sample. Reference of corresponding study is listed below: study 1-ref [51], study 2-ref [52], study 3-ref [53], study 4-ref [18, 19], study 5- ref [54], study 6-ref [55], study 7-ref [56], study 8-no publication yet, study 9-ref [57], study 10-ref [16, 17], study 11-ref [17], study 12-ref [21–23], study 13-ref [58], study 14-ref [59], study 15-ref [60]. Detailed figures are available in Table S1. Detailed figures are available in Table S2


a
SPINK1 (p< 0.001)
b
BUD31 (p< 0.001)
c
SF3B2 (p< 0.001)
d
SF3B4 (p< 0.001)
e
SNRPE (p< 0.001)
f
PRPF19 (p< 0.001)
g
TRA2A (p< 0.05)
h
HSPA1A (p< 0.001)
Tumor/ Normal
log2 fold change value of mRNA expression

◀**Fig. 3** Real-time PCR validations of spliceosome pathway gene up-regulation in HCC samples. The relative expression levels of genes were analyzed in paired HCC tissues and adjacent non-tumor tissues by quantitative RT-PCR. **a** human SPINK1 gene; **b** human BUD31 gene; **c** human SF3B2 gene; **d** human SF3B4 gene; **e** human SNRPE gene; **f** human PRPF19 gene; **g** human TRA2A gene; **h** human HSPA1A gene

samples. Thus, our meta-analysis results were consistent with the gene expression data (from 15 independent studies) in Nextbio database.

Validation of these spliceosome pathway gene expression alterations by real-time PCR

Primary HCC tissues and the corresponding non-tumorous liver tissues were used for validation of these spliceosome gene expression alteration genes in HCC by real-time PCR. Our results indicated that BUD31, HSPA1A, SNRPE, SF3B2, SF3B4 TRA2A, PRPF19 and SPINK1 genes all are up-regulated in our analysis (Fig. 3) (The genes with log 2 ratios ≥ 1 or ≤ -1 were determined as significantly changed (up-regulated or down-regulated).

Discussion

Spliceosome is probably the most complicated RNA–protein complex in eukaryotic cells [32]. It consists of five different small nuclear RNPs, naming U1, U2, U4, U5 and U6, with many other proteins cofactors associated [33]. When catalyzing splicing reactions, spliceosomes are assembled in a step-by-step manner on their targets. This splicing process is controlled by base paring of snRNA with short splicing signal motifs located between exons and introns of the target mRNAs as well as *cis*-acting regulatory elements on the pre-mRNA and *trans*-acting splicing protein factors. Although the exact mechanism driving pre-mRNAs into different splicing outcomes is yet to be discovered, recent evidence has shown the involvement of many splicing factors [34].

On the other hand, spliceosome has already been the target of many anticancer drugs, including Spliceostatin A and pladienolides [35]. Spliceostatin A was derived from an natural product FR901464, which is isolated from bacterium *Pseudomonas* sp. When first applied to clinical treatment, the exact mechanism of Spliceostatin A to inhibit cancer progression is unknown. Later, result of Kaida et al. [36] confirms that Spliceostatin A targets an Spliceosome complex member SF3B, an subcomplex of U2 small RNP. Our study indicated that SF3B2 and SF3B4 which are subunits of U2 small RNPs are overexpressed in HCC.

Although systematic protein-level analysis of spliceosome genes expression in cancer has not been reported, several studies provided evidence that many genes in spliceosome are up-regulated in cancers. Protein-level up-regulation of spliceosome gene HSPA1A, SNRPE, TRA2B and PRPF19 in cancer tissue was reported in previous studies [33–40]. HSPA1A, also known as Heat Shock Protein 70, is overexpressed in most cancer cells [26, 36]. Inhibition of HSP70 disrupts protein degradation pathway [37] and increases sensitivity of leukemia cells to antileukemia agent [38]. Depletion of SNRPE gene by RNA interference resulted in decreased proliferation of prostate, breast, lung and melanoma cancer cells [39], while overexpression of SNRPE caused rapid proliferation of prostate cancer cells [40]. TRA2B gene is overexpressed in breast, cervical, ovarian and colon cancer and considered to be a contributory factor to their malignancy [41–43]. PRPF19, also known as pre-mRNA processing factor 19, has also been reported to be involved cell proliferation and apoptosis [44]. Overexpression of prpf19 protein in endothelial cells results in longer life span [45].

Previous genome-wide analysis suggests that 40–60 % of human genes have alternative splicing forms [46]. Alternative splicing of several cancer-related genes, such as BRCA1 [47], CD44 [48] and APC [49], can be observed specifically in cancer, though little is known about how alternative splicing is involved in cancer [50]. Although some genes of spliceosome were not shown up-regulated, our study for the first time indicated that genes of spliceosome pathway are up-regulated in HCC. Furthermore, we also confirmed that BUD31, SF3B2, SF3B4, SNRPE, SPINK1, TPA2A, HSPA1A and PRPF19 genes are all up-regulated in HCC samples. This finding suggests that spliceosome pathway might play an unidentified role in HCC. Our study should be helpful for further investigation of the relationship between spliceosome pathway regulation and HCC, as well as other cancers.

Acknowledgments This work was supported by National Natural Science Foundation of China (Grant No. 31071186, to Dr. Cao). Weijin Xu was supported by National Top Talent Undergraduate Training Program (NTTUTP).

Conflict of interest None.

References

1. Jemal A, Bray F, Eard E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011;61:69–90.
2. Coleman WB. Mechanisms of human hepatocarcinogenesis. Curr Mol Med. 2003;3:573–88.
3. Jackson PE, Kuang SY, Wang JB, Strickland PT, Munoz A, Kensler TW, Qian GS, Groopman JD. Prospective detection of

codon 249 mutations in plasma of hepatocellular carcinoma patients. Carcinogenesis. 2003;24:1657–63.
4. Kuang SY, Lekawanvijit S, Maneekarn N, Thongsawat S, Brodovicz K, Nelson K, Groopman JD. Hepatitis b 1762t/1764a mutations, hepatitis c infection, and codon 249 p53 mutations in hepatocellular carcinomas from Thailand. Cancer Epidemiol Biomark Prev. 2005;14:380–4.
5. Law PT, Ching AK, Chan AW, Wong QW, Wong CK, To KF, Wong N. Mir-145 modulates multiple components of the insulin-like growth factor pathway in hepatocellular carcinoma. Carcinogenesis. 2012;33:1134–41.
6. Ertel A, Verghese A, Byers SW, Ochs M, Tozeren A. Pathway-specific differences between tumor cell lines and normal and tumor tissue cells. Mol Cancer. 2006;5:55.
7. Hu N, Wang C, Hu Y, Yang HH, Giffen C, Tang ZZ, Han XY, Goldstein AM, Emmert-Buck MR, Buetow KH, Taylor PR, Lee MP. Genome-wide association study in esophageal cancer using genechip mapping 10 k array. Cancer Res. 2005;65:2542–6.
8. Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang CH, Angelo M, Ladd C, Reich M, Latulippe E, Mesirov JP, Poggio T, Gerald W, Loda M, Lander ES, Golub TR. Multiclass cancer diagnosis using tumor gene expression signatures. Proc Natl Acad Sci USA. 2001;98:15149–54.
9. Nordmann AJ, Kasenda B, Briel M. Meta-analyses: what they can and cannot do. Swiss Med Wkly. 2012;142:w13518.
10. Chen C, Fu X, Zhang D, Li Y, Xie Y, Li Y, Huang Y. Varied pathways of stage ia lung adenocarcinomas discovered by integrated gene expression analysis. Int J Biol Sci. 2011;7:551–66.
11. Liu W, Peng Y, Tobin DJ. A new 12-gene diagnostic biomarker signature of melanoma revealed by integrated microarray analysis. PeerJ. 2013;1:e49.
12. Ramasamy A, Mondry A, Holmes CC, Altman DG. Key issues in conducting a meta-analysis of gene expression microarray datasets. PLoS Med. 2008;5:e184.
13. Zaravinos A, Lambrou GI, Boulalas I, Delakas D, Spandidos DA. Identification of common differentially expressed genes in urinary bladder cancer. PLoS One. 2011;6:e18135.
14. Zhou Z, Licklider LJ, Gygi SP, Reed R. Comprehensive proteomic analysis of the human spliceosome. Nature. 2002;419:182–5.
15. Wahl MC, Will CL, Luhrmann R. The spliceosome: design principles of a dynamic rnp machine. Cell. 2009;136:701–18.
16. Lee JS, Chu IS, Mikaelyan A, Calvisi DF, Heo J, Reddy JK, Thorgeirsson SS. Application of comparative functional genomics to identify best-fit mouse models to study human cancer. Nat Genet. 2004;36:1306–11.
17. Lee JS, Heo J, Libbrecht L, Chu IS, Kaposi-Novak P, Calvisi DF, Mikaelyan A, Roberts LR, Demetris AJ, Sun Z, Nevens F, Roskams T, Thorgeirsson SS. A novel prognostic subtype of human hepatocellular carcinoma derived from hepatic progenitor cells. Nat Med. 2006;12:410–6.
18. Roessler S, Jia HL, Budhu A, Forgues M, Ye QH, Lee JS, Thorgeirsson SS, Sun Z, Tang ZY, Qin LX. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. Cancer Res. 2010;70:10202–12.
19. Roessler S, Long EL, Budhu A, Chen Y, Zhao X, Ji J, Walker R, Jia HL, Ye QH, Qin LX, Tang ZY, He P, Hunter KW, Thorgeirsson SS, Meltzer PS, Wang XW. Integrative genomic identification of genes on 8p associated with hepatocellular carcinoma progression and patient survival. Gastroenterology. 2012;142(957–966):e912.
20. Kim BY, Lee JG, Park S, Ahn JY, Ju YJ, Chung JH, Han CJ, Jeong SH, Yeom YI, Kim S, Lee YS, Kim CM, Eom EM, Lee DH, Choi KY, Cho MH, Suh KS, Choi DW, Lee KH. Feature genes of hepatitis b virus-positive hepatocellular carcinoma, established by its molecular discrimination approach using prediction analysis of microarray. Biochim Biophys Acta. 2004;1739:50–61.
21. Tung EK, Mak CK, Fatima S, Lo RC, Zhao H, Zhang C, Dai H, Poon RT, Yuen MF, Lai CL, Li JJ, Luk JM, Ng IO. Clinicopathological and prognostic significance of serum and tissue dickkopf-1 levels in human hepatocellular carcinoma. Liver Int. 2011;31:1494–504.
22. Lamb JR, Zhang C, Xie T, Wang K, Zhang B, Hao K, Chudin E, Fraser HB, Millstein J, Ferguson M, Suver C, Ivanovska I, Scott M, Philippar U, Bansal D, Zhang Z, Burchard J, Smith R, Greenawalt D, Cleary M, Derry J, Loboda A, Watters J, Poon RT, Fan ST, Yeung C, Lee NP, Guinney J, Molony C, Emilsson V, Buser-Doepner C, Zhu J, Friend S, Mao M, Shaw PM, Dai H, Luk JM, Schadt EE. Predictive genes in adjacent normal tissue are preferentially altered by scnv during tumorigenesis in liver cancer and may rate limiting. PLoS One. 2011;6:e20090.
23. Sung WK, Zheng H, Li S, Chen R, Liu X, Li Y, Lee NP, Lee WH, Ariyaratne PN, Tennakoon C, Mulawadi FH, Wong KF, Liu AM, Poon RT, Fan ST, Chan KL, Gong Z, Hu Y, Lin Z, Wang G, Zhang Q, Barber TD, Chou WC, Aggarwal A, Hao K, Zhou W, Zhang C, Hardwick J, Buser C, Xu J, Kan Z, Dai H, Mao M, Reinhard C, Wang J, Luk JM. Genome-wide survey of recurrent hbv integration in hepatocellular carcinoma. Nat Genet. 2012;44:765–9.
24. Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman RB. Missing value estimation methods for DNA microarrays. Bioinformatics. 2001;17:520–5.
25. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statist Appl Genet Mol Biol. 2004;3:1–25.
26. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B (Methodological) 1995:289–300.
27. da Huang W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37:1–13.
28. Kanehisa M, Goto S. Kegg: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27–30.
29. Wu Y, Zuo J, Ji G, Saiyin H, Liu X, Yin F, Cao N, Wen Y, Li JJ, Yu L. Proapoptotic function of integrin beta(3) in human hepatocellular carcinoma cells. Clin Cancer Res 2009: 60–69.
30. Kupershmid I, Su QJ, Grewal A, Sundaresh S, Halperin I, Flynn J, Shekar M, Wang H, Park J, Cui W, Wall GD, Wisotzkey R, Alag S, Akhtari S, Ronaghi M. Ontology-based meta-analysis of global collections of high-throughput public data. PloS One 2010;5.
31. Marshall A, Lukk M, Kutter C, Davies S, Alexander G, Odom DT. Global gene expression profiling reveals spink1 as a potential hepatocellular carcinoma marker. PLoS One. 2013;8:e59459.
32. Nilsen TW. The spliceosome: the most complex macromolecular machine in the cell? BioEssays. 2003;25:1147–9.
33. Matera AG, Wang Z. A day in the life of the spliceosome. Nat Rev Mol Cell Biol. 2014;15:108–21.
34. Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, Blencowe BJ, Frey BJ. Deciphering the splicing code. Nature. 2010;465:53–9.
35. Dehm SM. Test-firing ammunition for spliceosome inhibition in cancer. Clin Cancer Res. 2013;19:6064–6.
36. Kaida D, Motoyoshi H, Tashiro E, Nojima T, Hagiwara M, Ishigami K, Watanabe H, Kitahara T, Yoshida T, Nakajima H, Tani T, Horinouchi S, Yoshida M. Spliceostatin a targets sf3b and inhibits both splicing and nuclear retention of pre-mrna. Nat Chem Biol. 2007;3:576–83.
37. Leu JI, Pimkina J, Pandey P, Murphy ME, George DL. Hsp70 inhibition by the small-molecule 2-phenylethynesulfonamide

impairs protein clearance pathways in tumor cells. Molecular Cancer Res MCR. 2011;9:936–47.
38. Guo F, Rocha K, Bali P, Pranpat M, Fiskus W, Boyapalle S, Kumaraswamy S, Balasis M, Greedy B, Armitage ES, Lawrence N, Bhalla K. Abrogation of heat shock protein 70 induction as a strategy to increase antileukemia activity of heat shock protein 90 inhibitor 17-allylamino-demethoxy geldanamycin. Cancer Res. 2005;65:10536–44.
39. Quidville V, Alsafadi S, Goubar A, Commo F, Scott V, Pioche-Durieu C, Girault I, Baconnais S, Le Cam E, Lazar V, Delaloge S, Saghatchian M, Pautier P, Morice P, Dessen P, Vagner S, Andre F. Targeting the deregulated spliceosome core machinery in cancer cells triggers mtor blockade and autophagy. Cancer Res. 2013;73:2247–58.
40. Anchi T, Tamura K, Furihata M, Satake H, Sakoda H, Kawada C, Kamei M, Shimamoto T, Fukuhara H, Fukata S, Ashida S, Karashima T, Yamasaki I, Yasuda M, Kamada M, Inoue K, Shuin T. Snrpe is involved in cell proliferation and progression of high-grade prostate cancer through the regulation of androgen receptor expression. Oncol Lett. 2012;3:264–8.
41. Best A, Dagliesh C, Ehrmann I, Kheirollahi-Kouhestani M, Tyson-Capper A, Elliott DJ. Expression of tra2 beta in cancer cells as a potential contributory factor to neoplasia and metastasis. Int J Cell Biol. 2013;2013:843781.
42. Watermann DO, Tang Y, Zur Hausen A, Jager M, Stamm S, Stickeler E. Splicing factor tra2-beta1 is specifically induced in breast cancer and regulates alternative splicing of the cd44 gene. Cancer Res. 2006;66:4774–80.
43. Gabriel B, Zur Hausen A, Bouda J, Boudova L, Koprivova M, Hirschfeld M, Jager M, Stickeler E. Significance of nuclear htra2-beta1 expression in cervical cancer. Acta Obstet Gynecol Scand. 2009;88:216–21.
44. Yin J, Zhu JM, Shen XZ. New insights into pre-mrna processing factor 19: a multi-faceted protein in humans. Biol Cell. 2012;104:695–705.
45. Voglauer R, Chang MW, Dampier B, Wieser M, Baumann K, Sterovsky T, Schreiber M, Katinger H, Grillari J. Snev overexpression extends the life span of human endothelial cells. Exp Cell Res. 2006;312:746–59.
46. Modrek B, Lee C. A genomic view of alternative splicing. Nat Genet. 2002;30:13–9.
47. Liu HX, Cartegni L, Zhang MQ, Krainer AR. A mechanism for exon skipping caused by nonsense or missense mutations in brca1 and other genes. Nat Genet. 2001;27:55–8.
48. Barbour AP, Reeder JA, Walsh MD, Fawcett J, Antalis TM, Gotley DC. Expression of the cd44v2-10 isoform confers a metastatic phenotype: importance of the heparan sulfate attachment site cd44v3. Cancer Res. 2003;63:887–92.
49. Neklason DW, Solomon CH, Dalton AL, Kuwada SK, Burt RW. Intron 4 mutation in apc gene results in splice defect and attenuated fap phenotype. Fam Cancer. 2004;3:35–40.
50. Kalnina Z, Zayakin P, Silina K, Line A. Alterations of pre-mrna splicing in cancer. Genes Chromosom Cancer. 2005;42:342–57.
51. Lim HY, Sohn I, Deng S, Lee J, Jung SH, Mao M, Xu J, Wang K, Shi S, Joh JW, Choi YL, Park CK. Prediction of disease-free survival in hepatocellular carcinoma by gene expression profiling. Ann Surg Oncol. 2013;20:3747–53.
52. Stefanska B, Huang J, Bhattacharyya B, Suderman M, Hallett M, Han ZG, Szyf M. Definition of the landscape of promoter DNA hypomethylation in liver cancer. Cancer Res. 2011;71:5891–903.
53. Huang Y, Chen HC, Chiang CW, Yeh CT, Chen SJ, Chou CK. Identification of a two-layer regulatory network of proliferation-related micrornas in hepatoma cells. Nucleic Acids Res. 2012;40:10478–93.
54. Chuma M, Sakamoto M, Yamazaki K, Ohta T, Ohki M, Asaka M, Hirohashi S. Expression profiling in multistage hepatocarcinogenesis: identification of hsp70 as a molecular marker of early hepatocellular carcinoma. Hepatology. 2003;37:198–207.
55. Ye QH, Qin LX, Forgues M, He P, Kim JW, Peng AC, Simon R, Li Y, Robles AI, Chen Y, Ma ZC, Wu ZQ, Ye SL, Liu YK, Tang ZY, Wang XW. Predicting hepatitis b virus-positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. Nat Med. 2003;9:416–23.
56. Wurmbach E, Chen YB, Khitrov G, Zhang W, Roayaie S, Schwartz M, Fiel I, Thung S, Mazzaferro V, Bruix J, Bottinger E, Friedman S, Waxman S, Llovet JM. Genome-wide molecular profiles of hcv-induced dysplasia and hepatocellular carcinoma. Hepatology. 2007;45:938–47.
57. Kaposi-Novak P, Libbrecht L, Woo HG, Lee YH, Sears NC, Coulouarn C, Conner EA, Factor VM, Roskams T, Thorgeirsson SS. Central role of c-myc during malignant conversion in human hepatocarcinogenesis. Cancer Res. 2009;69:2775–82.
58. Liao YL, Sun YM, Chau GY, Chau YP, Lai TC, Wang JL, Horng JT, Hsiao M, Tsou AP. Identification of sox4 target genes using phylogenetic footprinting-based prediction from expression microarrays suggests that overexpression of sox4 potentiates metastasis in hepatocellular carcinoma. Oncogene. 2008;27:5578–89.
59. Mas VR, Maluf DG, Archer KJ, Yanek K, Kong X, Kulik L, Freise CE, Olthoff KM, Ghobrial RM, McIver P, Fisher R. Genes involved in viral carcinogenesis and tumor initiation in hepatitis c virus-induced hepatocellular carcinoma. Mol Med. 2009;15:85–94.
60. Segal E, Sirlin CB, Ooi C, Adler AS, Gollub J, Chen X, Chan BK, Matcuk GR, Barry CT, Chang HY, Kuo MD. Decoding global gene expression programs in liver cancer by noninvasive imaging. Nat Biotechnol. 2007;25:675–80.