

Identification of important long non-coding RNAs and highly recurrent aberrant alternative splicing events in hepatocellular carcinoma through integrative analysis of multiple RNA-Seq datasets

Lu Zhang^{1,2} · Xiaoqiao Liu¹ · Xuegong Zhang^{2,3} · Ronghua Chen^{1,4}

Received: 27 July 2015 / Accepted: 16 December 2015 / Published online: 28 December 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Hepatocellular carcinoma (HCC) is an aggressive and deadly cancer. The molecular pathogenesis of the disease remains poorly understood. To better understand HCC biology and explore potential biomarkers and therapeutic targets, we investigated the whole transcriptome of HCC. Considering the genetic heterogeneity of HCC, four datasets from four studies consisting of 15 pairs of HCC and adjacent normal samples were analyzed. We observed that the number of lncRNAs expressed in each HCC sample was consistently greater than the adjacent normal sample. Moreover, 15 lncRNAs were identified expressed in five to seven HCC tissues but were not detected in any adjacent normal tissue. Differential expression analysis detected 35 up- and 80 down-regulated lncRNAs in HCC samples compared with adjacent normal samples. In addition, five differentially expressed lncRNAs were predicted to play a role in oxidation and reduction process. With regard to splicing

alterations, we identified nine highly recurrent differential splicing events belonging to eight genes *USO1*, *RPS24*, *CCDC50*, *THNSL2*, *NUMB*, *FN1* (two events), *SLC39A14* and *NR1I3*. Of them, splicing alterations of *SLC39A14* and *NR1I3* were reported for the association with HCC for the first time. The splicing dysregulation in HCC may be influenced by three splicing factors *ESRP2*, *CELF2* and *SRSF5* which were significantly down-regulated in HCC samples. This study revealed uncharacterized aspects of HCC transcriptome and identified important lncRNAs and splicing isoforms with the potential to serve as biomarkers and therapeutic targets for the disease.

Keywords HCC · Long noncoding RNA (lncRNA) · Alternative splicing · RNA-Seq

Communicated by S. Hohmann.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-015-1163-y) contains supplementary material, which is available to authorized users.

✉ Xuegong Zhang
zhangxg@tsinghua.edu.cn

✉ Ronghua Chen
ronghua_chen@merck.com

- ¹ Informatics, MSD China R&D, Beijing, China
- ² MOE Key Laboratory of Bioinformatics, Bioinformatics Division and Center for Synthetic and Systems Biology, TNLIST and Department of Automation, Tsinghua University, Beijing, China
- ³ School of Life Sciences, Tsinghua University, Beijing, China
- ⁴ Research IT, Merck and Co., Inc., Boston, MA, USA

Introduction

HCC is a prevalent solid-organ tumor representing the third leading cause of cancer mortality in the world with growing incidence rates (Jemal et al. 2011). The disease is primarily induced by hepatitis B virus (HBV) infection particularly in epidemic regions of Asia and Africa. Other pathogenic factors include hepatitis C virus (HCV) infection, alcoholism and aflatoxin B1 exposure (Parkin 2006; Jemal et al. 2011). To date, there are still very limited treatments for HCC urging the development of new diagnostic and prognostic biomarkers and therapeutic targets. A comprehensive exploration of the transcriptomic alterations in HCC is critical for a better understanding of the biology of HCC and provides possibilities for the identification of new biomarkers and therapeutic targets. With the rapid development of next-generation sequencing, RNA-Seq provides a powerful way to study transcriptome (Wang et al. 2009).

An lncRNA is a RNA molecule which has a length of more than 200 bp and lacks protein-coding potential. In cancer biology, lncRNAs are emerging as important regulators influencing a wide range of biological processes such as transcription (Wang et al. 2008b; Tsai et al. 2010; Bi et al. 2013), translation (Carrieri et al. 2012), gene expression (Wang and Chang 2011), cell cycle (Yochum et al. 2007) and cellular differentiation (Young et al. 2005). Several lncRNAs have been reported dysregulated in HCC such as *MALAT1* (Lai et al. 2012), *HOTAIR* (Ishibashi et al. 2013), *H19* (Matouk et al. 2007; Zhang et al. 2013), *MEG3* (Braconi et al. 2011), *MVIH* (Yuan et al. 2012) and *HULC* (Panzitt et al. 2007). Despite these findings, the expression profiles and function of lncRNAs in HCC remain poorly understood (Zhao et al. 2014), further investigation of which may shed new light on hepatopathogenesis and the utility of lncRNAs as biomarkers.

Alternative splicing of precursor mRNA of a gene can produce multiple isoforms, which results in a single gene coding for multiple proteins. In addition, inefficient splicing due to inclusion of premature STOP codons can lead to mRNA degradation through nonsense-mediated decay (NMD). Studies estimated that more than 90 % of multi-exon human genes undergo alternative splicing (Pan et al. 2008; Wang et al. 2008a). This phenomenon is a major post-transcriptional regulatory mechanism involved in the development of cancers by affecting key aspects of cancer cell biology including cell proliferation, cancer metabolism, angiogenesis, apoptosis, invasiveness and metastasis (Ghigna et al. 2008; David and Manley 2010; Biamonti et al. 2012). Splicing dysregulation of genes such as *FNI* (Oyama et al. 1989, 1993; Matsui et al. 1997), *CD44* (Harn et al. 1994), *FGFR2* (Lin et al. 2014b), *NT5E* (Snider et al. 2014) and *Sulf1* (Gill et al. 2012) have been reported to be associated with HCC. However, the knowledge on aberrant splicing in HCC is rather incomplete compared to other types of tumors (Berasain et al. 2010). Therefore, it is essential to thoroughly explore the splicing alterations generated by HCC through RNA-Seq.

This study characterized lncRNAs and differential splicing in HCC at whole-transcriptome level by integrative analysis of four sets of RNA-Seq data. Of note, we identified 15 lncRNAs only detectable in a number of HCC samples in this study and 115 lncRNAs differentially expressed between HCC and adjacent normal samples. In addition, function of five lncRNAs was predicted. On the other hand, nine highly recurrent differential splicing events were identified. Our findings provided important and novel insight into the transcriptional changes in HCC, justifying further investigation into the pathogenic and translational impact of these changes in HCC.

Materials and methods

Datasets

At the time of this study, four datasets on HCC were publicly available in NCBI Sequence Read Archive (SRA) database (Shumway et al. 2010) which were designated as dataset A [accession number SRP018008 (Kang et al. 2015)], B [accession number SRP009123 (Chan et al. 2014)], C [accession number SRP007560 (Lin et al. 2014b)] and D [accession number SRP004768 (Huang et al. 2011)] in this study, respectively (Table 1). Dataset A was derived from Asian Cancer Research Group including RNA-Seq data of nine HCC tissues and nine adjacent normal tissues. Dataset B is comprised of RNA-Seq data of three pairs of primary HCC and adjacent normal liver tissues. Dataset C contains RNA-Seq data sequenced in great depth of one pair of liver tumor and adjacent normal samples. Dataset D contains RNA-Seq data from 10 matched pairs of HCC and adjacent normal liver tissues.

Primary processing and alignment of RNA-Seq reads

First, quality of the RNA-Seq reads of the four datasets was checked using FastQC v0.10.1 (<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc>). FastQC showed that the quality of RNA-Seq data in dataset D is relatively low compared with the quality reached by recent sequencing technology. Software Trimmomatic v0.32 (Bolger et al. 2014) was then applied on the fastq files of dataset D to (1) remove adapters, (2) remove leading and trailing bases with quality below 3, (3) scan each read with a 4-base wide sliding window and cut when the average quality per base lower than 15 and (4) discard reads shorter than 36 bases. After trimming and eliminating short reads, between 37.73 and 73.76 % (median: 47.87 %) of total reads were kept for each sample in dataset D for downstream analysis.

Afterwards, all the RNA-Seq reads were mapped to the human reference genome (release hg19) using tophat (v2.0.10) (Kim et al. 2013) with default settings and the following options ‘-p 4 -g 10 - -keep-fasta-order - -no-coverage-search’. In addition, a combined gene annotation file was fed into tophat through the ‘-G’ parameter. The annotation file contained annotations for 296,680 transcript IDs created by merging the RefSeq genes (downloaded from UCSC Table Browser on Mar 3, 2014), UCSC known genes (downloaded from UCSC Table Browser on Mar 1, 2014) and Ensembl (v75) genes.

The statistics on the four sequencing datasets were presented in Online Resource 1a. The total number of reads per sample in dataset A ranged between 63,720,281 and 79,543,083, with a median of 73,190,808. Of them, from

Table 1 Descriptive information on the four datasets used for analysis

Dataset	Accession no.	Sample (T/N) ^a	Patient ID ^b	Sequence length	Sequencing type	Viral status	Reference
A	SRP018008	9 HCC samples and 9 adjacent normal samples	A_11, A_180, A_200, A_22, A_28, A_30, A_65, A_70, A_82	90 bp	Paired end	HBV ^c	Kan et al. (2013); Kang et al. (2015)
B	SRP009123	3 HCC samples and 3 adjacent normal samples	B_448, B_473, B_510	58 bp	Single end	HBV+	Chan et al. (2014)
C	SRP007560	1 HCC sample and 1 adjacent normal sample	C_1428	75 bp	Paired end	HBV–	Lin et al. (2014b)
D	SRP004768	10 HCC samples and 10 adjacent normal samples	D_A13, D_A39, D_A82, D_A63, D_A60, D_A57, D_A35, D_A31, D_A19, D_A05	40 bp	Single end	HBV+	Huang et al. (2011)

^a T represents HCC samples. N represents adjacent normal samples

^b Each patient ID in the present study was created by first giving the dataset ID, followed by an underscore and the patient ID used in the original study

^c Eight of the nine patients except patient A_28 in this cohort are positive for HBV. There is no record on the viral status for patient A_28

80.1 to 93.3 % were aligned to the reference genome. For samples in dataset B, the total read counts ranged from 29,479,423 to 31,701,232 (median 30,700,686). The median mapping rate is 94.5 %. The total read counts for the one pair of samples in dataset C is 126,533,165 and 127,161,324 with alignment rates of 91.5 and 93 %, respectively. For dataset D, after trimming and dropping short reads, only two pairs of samples from patients D_A13 and D_A39 with higher read counts retained (from 18,980,456 to 30,588,144) and alignment rates (between 90.5 and 94.9 %) were kept for downstream analysis.

Differential gene expression analysis

To identify differentially expressed genes between HCC and adjacent normal samples meta-analyses using two different *p* value combination techniques (Inverse normal and Fisher methods) were performed on the multiple datasets according to Rau et al. (2014). Meta-analyses were used to increase detection power by increasing available sample size. Before performing meta-analyses, a differential expression analysis was performed on each individual dataset.

We performed a paired sample test with a generalized linear model (GLM) method on each individual dataset for differential expression using edgeR (v3.2.4) program (Robinson et al. 2010). The input to edgeR is a matrix of read count with each row corresponding to a gene and each column corresponding to a sample. Read count per gene per sample was generated using htseq-count python script v0.6.1 (Anders et al. 2015) in union mode with the tophat aligned

reads and GENCODE v19 annotations for a total of 57,820 genes (Derrien et al. 2012; Harrow et al. 2012). For each dataset we used genes that achieve one count per million for at least half of the samples in the dataset to perform differential expression analysis. The significance level for per-study differential analysis was set at Benjamini-Hochberg (BH) false discover rate (FDR) of 0.05. A total of 2878, 2587, 670 and 6 differentially expressed genes were detected for dataset A, B, C and D, respectively. We found that the number of differentially expressed genes detected by dataset D is very low compared with the other three datasets. We checked the distribution of *p* values in the four datasets and found that the null distribution of *p* values from dataset D diverged from the expected uniform distribution greatly and was different from those of the other three datasets. This indicates that the model for detecting differentially expressed genes for dataset A, B and C did not fit dataset D well, probably due to the additional preprocessing of dataset D during quality control. Therefore, we decided not to include the *p* values from dataset D in the differential expression meta-analyses.

Meta-analyses were performed by combining the raw *p* values of all genes from individual differential expression analyses of dataset A, B and C through Inverse normal and Fisher approaches. The significance threshold for meta-analyses is BH FDR <0.05 for either Fisher or Inverse normal method. There were 744 genes displaying differential expression in contradictory directions in individual expression analyses which were removed from the list of genes identified as differentially expressed via meta-analyses.

After obtaining the list of differentially expressed genes, hierarchical clustering based on the list was

performed across the three datasets using log-transformed normalized gene counts $\log N$. The $\log N$ were calculated as $\log N = \ln\left(\frac{Y_{gk}}{S_k} \cdot \bar{S} + 1\right)$, where Y_{gk} was defined as the total reads for gene g for sample k ; S_k is derived from multiplying the library size of sample k by the scale factor for the sample k given by the trimmed mean of M values normalization method implemented in edgeR (Robinson and Oshlack 2010); \bar{S} represents the arithmetic mean of S_k values. The distances between the samples of the three datasets were calculated using $\log N$ by Spearman correlation method, while the distances between the differentially expressed genes were calculated using $\log N$ by Pearson correlation method. Clustering was formed using the average linkage method for samples and genes, respectively.

Identification of important lncRNAs in HCC by expression and correlation analysis

To identify important lncRNAs in HCC, we used the GENCODE v19 annotation (Derrien et al. 2012). The annotation contains 13,870 lncRNA genes involving six lncRNA types that are long intervening ncRNA (7114), antisense RNA (5276), processed transcript (515), 3' overlapping ncRNA (21), sense intronic ncRNA (742) and sense overlapping ncRNA (202). Furthermore, we found 41 genes tagged as processed transcript from GENCODE v19 lncRNAs annotation are, however, annotated as protein coding genes according to RefSeq annotation and were treated as protein coding genes in the present study. This resulted in an annotation of 13,829 lncRNA genes and 20,386 protein coding genes for downstream analysis.

First, we counted the number of lncRNA genes expressed (number of reads mapped to the lncRNA ≥ 10) in each HCC and adjacent normal samples. We also investigated the existence of lncRNAs which were only expressed (number of reads mapped to the lncRNA ≥ 10) in HCC tissues but were not expressed (number of reads mapped to the gene = 0) in any of the normal tissues. In addition, differentially expressed lncRNA genes were obtained from the list of differentially expressed genes derived from meta-analyses described in the above section in combination with GENCODE v19 lncRNA annotation.

The correlation between the differentially expressed lncRNAs and the differentially expressed protein coding genes were explored in this study. To obtain the Spearman correlation coefficients R the expression level of differentially expressed lncRNAs and protein coding genes represented by normalized counts calculated as mentioned above $\frac{Y_{gk}}{S_k} \cdot \bar{S}$ was used. A $|R| > 0.9$ and $FDR < 0.001$ was set as the significance threshold for correlation. Moreover, as highly connected genes tend to be involved in similar biological functions, we predicted the function of lncRNAs based on

the function of their co-expressed protein coding genes. Functional annotation was performed on each set of protein coding genes co-expressed with a lncRNA using biological processes (BP-FAT) annotation implemented in DAVID (Database for Annotation, Visualization and Integrated Discovery <http://david.abcc.ncifcrf.gov/>). The threshold for a significant gene ontology (GO) term was BH FDR < 0.1 .

Identification of highly recurrent splicing alterations in HCC through comparison with adjacent normal samples

The detection of differential alternative splicing events was carried out using program MISO (Mixture-of-Isoforms) v0.5.2 exon-centric analysis (Katz et al. 2010). As MISO does not handle replicates/groups of samples, we performed MISO program on each pair of HCC and adjacent normal samples independently and then summarized the results across all 15 pairwise comparisons of the four datasets (Online Resource 2a). A total of five types of alternative splicing events were investigated, including alternative 3' splice sites (A3SS), alternative 5' splice sites (A5SS), mutually exclusive exons (MXE), retained introns (RI) and skipped exons (SE) using the annotation file provided by MISO (hg19 v2.0) (<http://miso.readthedocs.org/en/fast-miso/annotation.html>). The annotation for each alternative splicing event contains two isoforms. MISO utilized these annotations and tophat aligned bam files to detect differential alternative splicing events between each HCC and adjacent normal samples through calculating $\Delta\Psi$ and Bayes factors which determine the magnitude and statistical significance of splicing differences, respectively.

The differential splicing events detected by an HCC sample-adjacent normal sample pairwise comparison were filtered by applying MISO default cutoff of $|\Delta\Psi| \geq 0.2$ and Bayes factor ≥ 10 . Moreover, for an event to be retained we required the sum of junction counts supporting the first or the second isoform ≥ 5 in at least one of the two samples. After filtering, each differential splicing event was summarized across all the pairwise comparisons detecting the event by counting the number of detections with positive $\Delta\Psi$ values ($\# \Delta\Psi (+)$) and the number of detections with negative $\Delta\Psi$ values ($\# \Delta\Psi (-)$), respectively. For a splicing event different signs (+ or -) of $\Delta\Psi$ represent different directions of splicing change (inclusion or exclusion of a splicing segment in HCC versus adjacent normal). We called a differential splicing event as recurrent only if the event was detected by several pairwise comparisons and the number of detections in one direction is larger than that of the other direction by a given number 3. The definition for recurrent differential splicing events can be denoted by $|\# \Delta\Psi (+) - \# \Delta\Psi (-)| \geq 3$. In this study, a highly recurrent differential splicing event entails the event occurring in

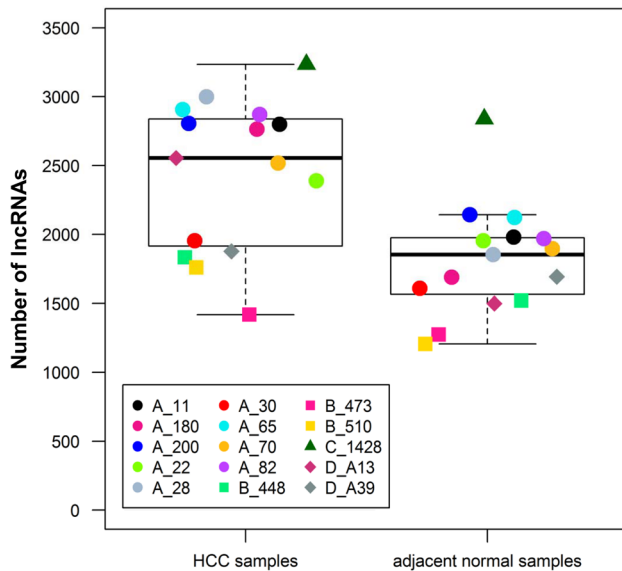


Fig. 1 Boxplots displaying the number of lncRNAs expressed in HCC and adjacent normal samples. The y-axis reflects the number of lncRNAs. The cutoff for claiming an lncRNA being expressed is 10 reads mapped to the gene. The number of lncRNAs expressed in each sample was also plotted on the boxplots as a dot with specific color and shape for each patient as shown in the legend

more than half of the 15 HCC tissues (i.e., recurrence times ≥ 8).

To understand whether the highly recurrent differential splicing genes are biologically connected, we used software IPA (ingenuity pathway analysis, Ingenuity Systems, www.ingenuity.com) which can generate networks enriched by the user-provided genes (focus genes). IPA produces the networks by applying an algorithm on the list of user-specified genes and IPA knowledge base global molecular network and gives p scores to rank the networks. In a network containing n genes of which f are focus genes, the p score is the negative logarithm (base 10) of the p value representing the probability of obtaining at least f focus genes in a set of n genes randomly taken from the global molecular network calculated using Fisher's exact test.

Results

Preliminary characterization of lncRNA expression profiles in HCC

lncRNAs have been shown to act as critical components of cancer biology because of their important roles in regulating key biological processes. In the present study we found that the number of lncRNAs expressed in each HCC sample was consistently greater than the adjacent normal sample across the four datasets (Fig. 1). By Wilcoxon

signed-rank test, we confirmed that the number of lncRNA genes in tumor samples was significantly greater than that in adjacent normal samples ($p = 6.10e-5$). Moreover, we identified 15 lncRNAs that were not detected in any of the adjacent normal samples but were expressed in between five and seven of the 15 HCC samples including *NOVA1-AS1*, *CTC-261N6.1*, *XX-C2158C6.3*, *RP11-284G10.1*, *RP11-199O14.1*, *RP11-346D6.6*, *RP1-90L14.1*, *RP11-608O21.1*, *RP11-565A3.2*, *RP11-400D2.2*, *RP11-6N13.1*, *RP11-962G15.1*, *RP11-1038A11.3*, *RP11-103J17.2* and *RP11-109M17.2* (Fig. 2). Of the 15 lncRNAs, only *RP11-109M17.2* is single-exonic gene according to GENCODE annotation. The 15 lncRNAs are worthy of further investigation for their roles in HCC tumorigenesis.

Detection of differentially expressed lncRNAs in HCC in comparison with adjacent normal samples

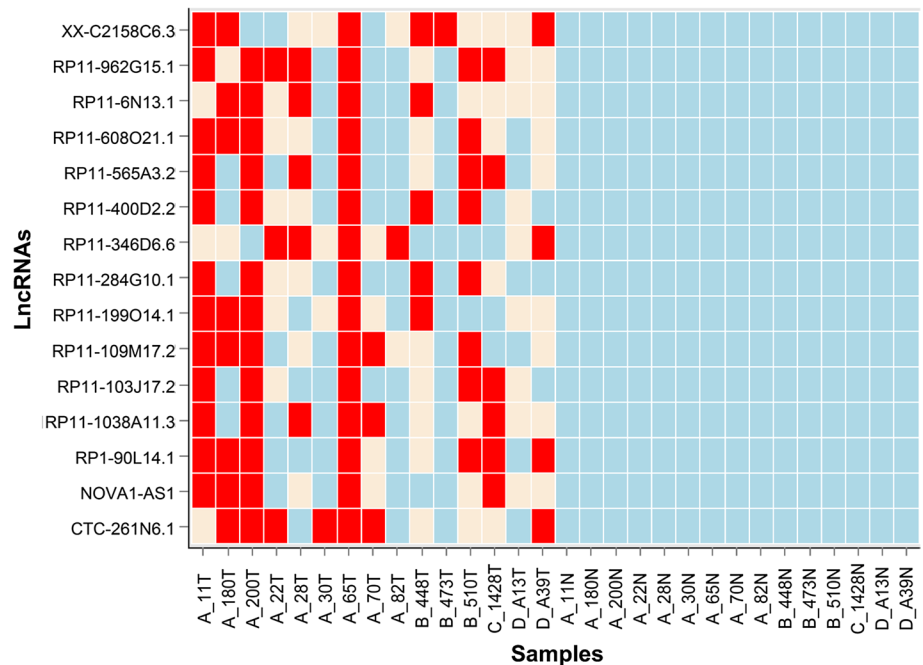
Before detecting differentially expressed lncRNA genes between HCC and adjacent normal samples, we first identified 3112 differentially expressed genes through p value combination approaches (Inverse normal and Fisher methods) using dataset A, B and C (Online Resource 2b). Hierarchical clustering was performed on the three datasets based on the 3112 genes and showed a distinguishable gene expression profiling between HCC and adjacent normal samples (Online Resource 2c).

Based on the 3112 differentially expressed genes and GENCODE v19 lncRNA annotation, we identified 35 up- and 80 down-regulated lncRNA genes in HCC samples in comparison with adjacent normal samples (Online Resource 1b). Several of these differentially expressed lncRNAs have been well-studied including *H19*, *MEG3*, *HAND2-AS1*, *RN7SK*, *LINC00261* and *TP53TG1* down-regulated in HCC samples and *GAS5*, *LINC00152*, *PVT1* and *SNHG1* up-regulated in HCC samples in this study. The 20 most significantly differentially expressed lncRNA genes in HCC samples according to the FDR obtained by Fisher method are listed in Table 2, which contains lncRNA *HAND2-AS1*.

Co-expression analysis of lncRNAs and protein coding genes

This study investigated the effects of expression changes of the 115 differentially expressed lncRNAs on the expression of differentially expressed protein coding genes in HCC. As a result, we identified 212 pairs of co-expressed lncRNAs and protein coding genes formed by 33 lncRNAs (3 were up-regulated in HCC samples and 30 were down-regulated in HCC samples) and 173 protein coding genes with 210 pairs presented as positive correlation and only two pairs presented as negative correlation (Fig. 3; Online

Fig. 2 Expression matrix of 15 lncRNAs which were only expressed in HCC tissues but were not expressed in any adjacent normal samples. Being expressed was defined as having mapped read count ≥ 10 illustrated by red color. Not being expressed was defined as having mapped read count = 0 represented by light blue color. Expression values in the middle of 0 to 10 mapped read counts were depicted using antique white color. The x-axis represents sample names. Sample names ending in *T* represent HCC samples. Sample names ending in *N* represent adjacent normal samples. The y-axis represents lncRNAs (color figure online)



Resource 1c). The two negatively co-expressed pairs are *LINC01093–TTK* ($R = -0.912$, $FDR = 4.14e-07$) and *LINC01093–NDC80* ($R = -0.902$, $FDR = 7.91e-07$). Moreover, there were five differentially expressed lncRNAs positively correlated with six nearby differentially expressed protein coding genes (distance <300 kb) forming co-expressed pairs including the top two significantly positively co-expressed pairs *CTD-2044J15.2–SRD5A1* ($R = 0.979$, $FDR = 4.26e-05$) and *CTB-167B5.2–STEAP4* ($R = 0.976$, $FDR = 6.95e-12$) along with *TNRC6C-AS1–TMC8* ($R = 0.931$, $FDR = 8.40e-08$), *CTD-2337J16.1–LILRB5* ($R = 0.953$, $FDR = 7.09e-09$), *CTD-2337J16.1–LILRB2* ($R = 0.904$, $FDR = 6.91e-07$) and *RP11-42O15.3–CTH* ($R = 0.922$, $FDR = 2.13e-07$). In addition, we found that *MEG3* was co-expressed with *LAMC3* ($R = 0.904$, $FDR = 0.0001$) and *C8orf58* ($R = 0.905$, $FDR = 6.54e-07$).

Based on the coding–non coding co-expression pairs, we also performed functional prediction of the lncRNAs according to GO biological processes (BP-FAT) terms enriched by their co-expressed protein coding genes using DAVID. In total function of five lncRNAs *LINC00261*, *RP11-119D9.1*, *AC004538.3*, *CTD-2044J15.2* and *CTC-505O3.2* was predicted. All of the five lncRNAs were down-regulated in HCC samples (Online Resource 1b). The five sets of protein coding genes correlated with the five lncRNAs were all enriched in the biological process of oxidation and reduction (Table 3). In addition, functional annotation also showed the association between *LINC00261* and lipid metabolism and the correlation between *AC004538.3* and immunity.

Identification of highly recurrent differential splicing events between HCC and adjacent normal samples

Aberrant splicing is a hallmark of cancer which was investigated by MISO exon-centric analysis. An average of 82 A3SS, 57 A5SS, 83 MXE, 135 RI and 324 SE differential splicing events were detected by the 15 HCC-adjacent normal pairwise comparisons after filtering (Fig. 4). Differential exon skipping and differential intron retention seem to be the predominant differential splicing types. Moreover, 43 A3SS, 27 A5SS, 37 MXE, 84 RI and 199 SE differential splicing events were identified as recurrent differential splicing events based on our definition ($|\# \Delta\Psi (+) - \# \Delta\Psi (-)| \geq 3$) (Fig. 4; Online Resource 1d-h).

Of note, we identified nine aberrant alternative splicing events occurring in at least 8 of the 15 HCC samples (incidence >50 %) (Fig. 5; Table 4). The highly recurrent aberrant splicing events took place in eight genes *USO1*, *RPS24*, *CCDC50*, *THNSL2*, *SLC39A14*, *NR1I3*, *FNI* (two events) and *NUMB*. To the best of our knowledge, the aberrant splicing occurring in *SLC39A14* and *NR1I3* was reported for the association with HCC for the first time. The other seven events have been shown to be present in HCC before (Oyama et al. 1989, 1993; Matsui et al. 1997; Huang et al. 2011; Danan-Gotthold et al. 2015; Lu et al. 2015; Zhang et al. 2015). Among the eight genes, *SLC39A14* (Fisher $FDR = 0.006$) and *NR1I3* (Fisher $FDR = 0.002$) were down-regulated in HCC samples compared to adjacent normal samples. The other six genes were not differentially expressed between HCC and adjacent normal samples.

Table 2 The top 20 significantly differentially expressed lncRNAs in HCC samples in comparison with adjacent normal samples according to the false discovery rate (FDR) produced by meta-analysis using Fisher method

LncRNA ID	Gene name	LncRNA type	FDR (A) ^a	logFC (A) ^a	FDR (B) ^b	logFC (B) ^b	FDR (C) ^c	logFC (C) ^c	FDR (Fisher) ^d	FDR (inverse normal) ^e
ENSG00000204583.5	<i>LRCOL1</i>	lincRNA	1.49E-08	-4.3	1.77E-09	-6.5	0.4359	-2.1	0	0
ENSG00000223991.1	<i>AC104809.2</i>	antisense	0.0003	-5.2	6.19E-11	-8.3	1.79E-05	-7.4	0	0
ENSG00000237949.1	<i>LINC00844</i>	lincRNA	2.46E-11	-5.7	1.31E-05	-4.9	0.3342	-2.2	0	0
ENSG00000243694.2	<i>RP11-6B4.1</i>	lincRNA	0.0153	-2.6	5.92E-13	-10.3	0.0037	-4.6	0	2.22E-11
ENSG00000249173.1	<i>LINC01093</i>	lincRNA	3.32E-11	-5.9	9.19E-17	-8.7	0.0001	-6.4	0	0
ENSG00000261012.2	<i>RP11-116D2.1</i>	lincRNA	0.0184	-1.0	7.09E-17	-5.4	0.6757	-1.6	0	0
ENSG00000272520.1	<i>CTD-2044J15.2</i>	lincRNA	1.84E-11	-2.2	6.17E-09	-3.3	0.9995	-0.6	0	0
ENSG00000264868.1	<i>CTB-167B5.2</i>	lincRNA	2.74E-06	-3.2	4.32E-08	-4.8	0.0549	-3.3	7.80E-15	2.43E-14
ENSG00000248709.1	<i>CTC-505O3.2</i>	lincRNA	0.0003	-2.6	7.53E-11	-6.5	0.5403	-1.9	5.02E-14	1.10E-11
ENSG00000254211.1	<i>CTB-43E15.4</i>	lincRNA	0.0304	-1.9	1.35E-13	-5.0	0.9657	-1.1	6.29E-14	1.31E-08
ENSG00000237125.4	<i>HAND2-AS1</i>	antisense	0.0022	-3.4	2.06E-07	-5.3	0.0004	-5.7	2.16E-13	8.48E-11
ENSG00000248740.1	<i>RP11-328K4.1</i>	lincRNA	1.70E-07	-4.3	0.1044	-1.4	0.0004	-5.7	9.32E-12	4.67E-11
ENSG00000234456.3	<i>MAGI2-AS3</i>	processed transcribed	0.0003	-1.8	4.40E-06	-3.6	0.0050	-4.4	1.06E-11	1.14E-10
ENSG00000241224.2	<i>RP11-59E19.1</i>	lincRNA	2.06E-05	-3.7	0.0056	-7.1	0.0002	-6.0	2.95E-11	1.68E-10
ENSG00000225765.1	<i>AC068535.3</i>	lincRNA	0.0040	-2.0	3.88E-07	-6.3	0.0103	-4.2	2.96E-11	1.94E-09
ENSG00000251637.2	<i>RP11-119D9.1</i>	lincRNA	0.0002	-2.8	6.22E-07	-4.9	0.3644	-2.2	9.37E-11	2.80E-10
ENSG00000232667.5	<i>AC004862.6</i>	antisense	8.95E-08	-4.0	0.0009	-3.1	0.9995	-0.6	6.61E-10	2.59E-10
ENSG00000230333.2	<i>AC004538.3</i>	processed transcribed	2.71E-06	-3.5	0.0003	-4.4	0.5012	-2.0	1.10E-09	2.08E-10
ENSG00000249096.2	<i>RP11-290F5.1</i>	lincRNA	5.91E-05	-2.1	0.0012	-3.2	0.0131	-4.0	1.30E-09	1.08E-09
ENSG00000269976.1	<i>RP11-130L8.2</i>	lincRNA	0.2661	-0.9	5.30E-08	-5.7	0.0251	-3.7	1.56E-09	1.37E-05

The table listed the FDR and log transformed fold change (logFC) derived from independent dataset analysis and FDR from meta-analysis for each of the 20 lncRNAs

^a A represents dataset A

^b B represents dataset B

^c C represents dataset C

^d FDR produced by meta-analysis using Fisher method

^e FDR produced by meta-analysis using Inverse normal method

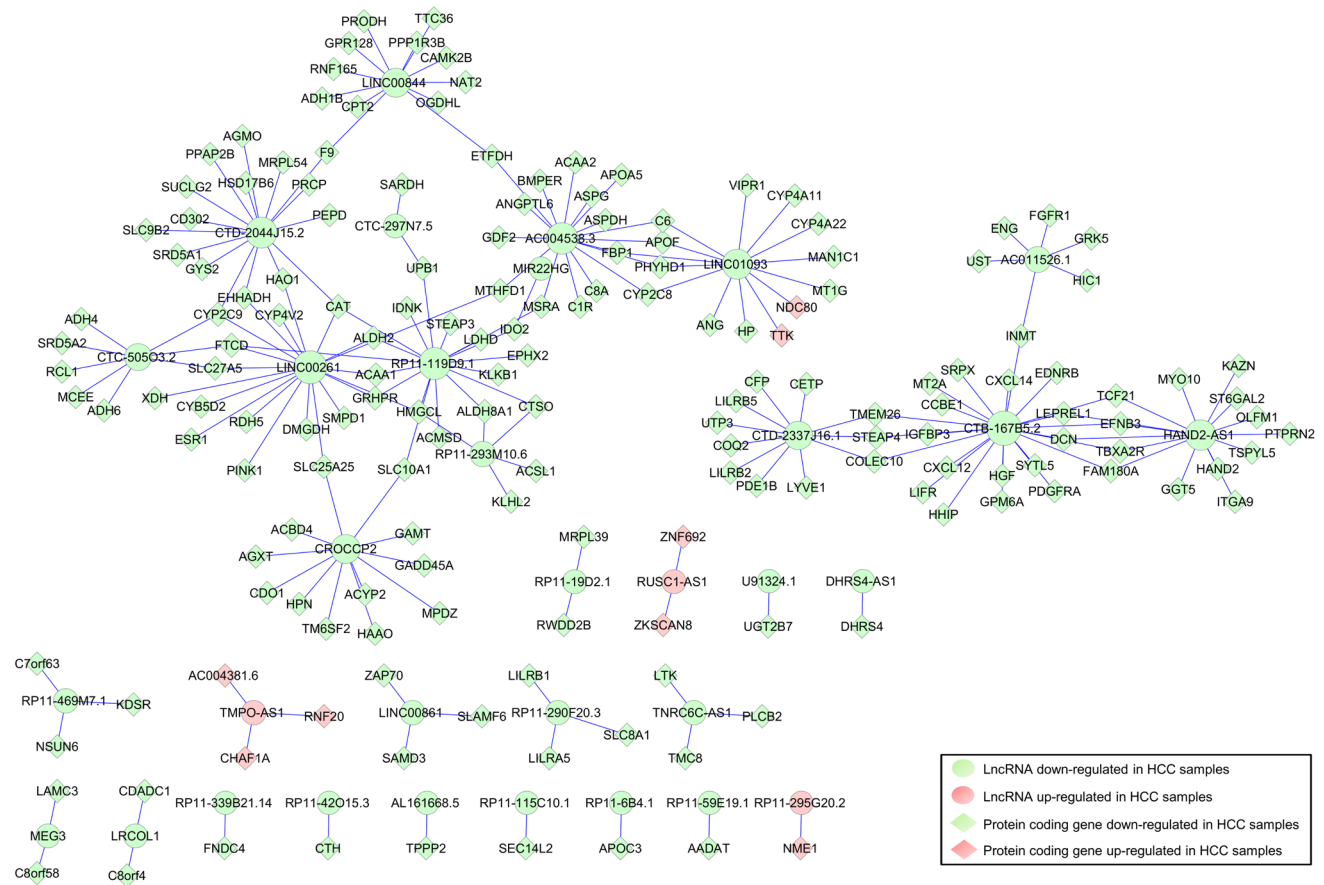


Fig. 3 Coding-non coding co-expression network. The plot exhibited the co-expression network formed by 33 differentially expressed lncRNAs and 173 differentially expressed protein coding genes. Each lncRNA was represented by an ellipse and each protein coding gene

was denoted by a diamond. Pink color represents the gene up-regulated in HCC samples compared with the adjacent normal tissues, and green color indicates the gene down-regulated in HCC tissues compared with the adjacent normal tissues (color figure online)

The most frequent differential splicing gene is *USO1* identified in 11 out of 15 HCC tissues. In the 11 HCC tissues the exon (exon 15 of NM_001290049)-excluding isoform of *USO1* was significantly up-regulated in isoform percentage (Fig. 6). In addition, we observed that *FNI* isoforms containing extra-domain A (ED-A) region (Online Resource 2d) or type III connecting segment 1 (CS1) (Online Resource 2e) region were up-regulated in relative abundance in 53.3 and 66.7 % HCC samples relative to adjacent normal samples, respectively. Seven out of the eight patients undergoing differential splicing of *FNI* at the ED-A region also went through differential splicing at the CS1 region indicating a coordinated dysregulation of splicing of *FNI* at these two regions in HCC tissues compared with adjacent normal tissues. The remaining six highly recurrent differential splicing events were illustrated in Online Resource 2f-k.

The eight highly recurrent differential splicing genes were involved in a network with the function cell-to-cell signaling and interaction, cell-mediated immune response

and cellular development though IPA network analysis (score = 24) (Online Resource 2l). The specific functions of these HCC-related aberrant splicing events are worthy of further investigation for their roles in the development of HCC.

Detection of three splicing factors down-regulated in HCC

Alternative splicing is regulated through the interplay between cis-acting sequence elements of the pre-mRNA and trans-acting splicing factor proteins that bind to them. We checked whether any of the 71 splicing factors from SpliceAid-F database (Giulietti et al. 2013) was in the list of differentially expressed genes detected by differential expression meta-analyses described above. We found that three splicing factors *ESRP2* (Fisher FDR = $2.41e-05$), *CELF2* (Fisher FDR = 0.06) and *SRSF5* (Fisher FDR = 0.005) were down-regulated in HCC samples compared with the adjacent normal samples (Fig. 7) suggesting

Table 3 Predicted function of lncRNAs based on the co-expressed protein coding genes

LncRNA ^a	Term	Count ^b	P value	Benjamini ^c
<i>LINC00261</i> (20)	GO:0055114~oxidation reduction	11	1.60E-09	3.68E-07
	GO:0046395~carboxylic acid catabolic process	5	1.20E-05	0.0014
	GO:0016054~organic acid catabolic process	5	1.20E-05	0.0014
	GO:0044242~cellular lipid catabolic process	4	1.31E-04	0.0100
	GO:0019395~fatty acid oxidation	3	0.0012	0.0539
	GO:0034440~lipid oxidation	3	0.0012	0.0539
	GO:0016042~lipid catabolic process	4	0.0015	0.0543
	GO:0009062~fatty acid catabolic process	3	0.0010	0.0573
	GO:0006631~fatty acid metabolic process	4	0.0021	0.0681
<i>RP11-119D9.1</i> (18)	GO:0055114~oxidation reduction	8	2.34E-06	5.51E-04
<i>AC004538.3</i> (17)	GO:0055114~oxidation reduction	6	6.54E-04	0.0582
	GO:0002541~activation of plasma proteins involved in acute inflammatory response	3	0.0012	0.0614
	GO:0006956~complement activation	3	0.0011	0.0728
	GO:0019724~B cell mediated immunity	3	0.0019	0.0737
	GO:0002455~humoral immune response mediated by circulating immunoglobulin	3	5.98E-04	0.0789
	GO:0016064~immunoglobulin mediated immune response	3	0.0018	0.0797
	GO:0002449~lymphocyte mediated immunity	3	0.0030	0.0988
<i>CTD-2044J15.2</i> (16)	GO:0055114~oxidation reduction	6	1.39E-04	0.0220
<i>CTC-50503.2</i> (8)	GO:0051186~cofactor metabolic process	3	0.0041	0.0758
	GO:0055114~oxidation reduction	4	0.0032	0.0868

^a The number of co-expressed protein coding genes of the lncRNA was enclosed in the parentheses

^b Count of protein coding genes involved in the biological process

^c Benjamini–Hochberg false discovery rate

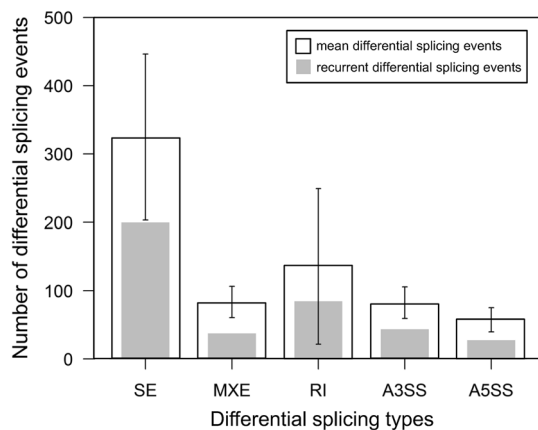


Fig. 4 Number of differential splicing events detected between HCC and adjacent normal samples. There are five types of differential splicing including differential alternative 3' splice sites (A3SS), differential alternative 5' splice sites (A5SS), differential mutually exclusive exons (MXE), differential retained introns (RI) and differential skipped exons (SE). The height of a *hollow bar* represents the average number of differential splicing events of a specific type detected by the 15 HCC-adjacent normal pairwise comparisons. The *error bar* of a *hollow bar* represents the standard deviation. The height of a *filled bar* represents the number of recurrent differential splicing events of a certain type

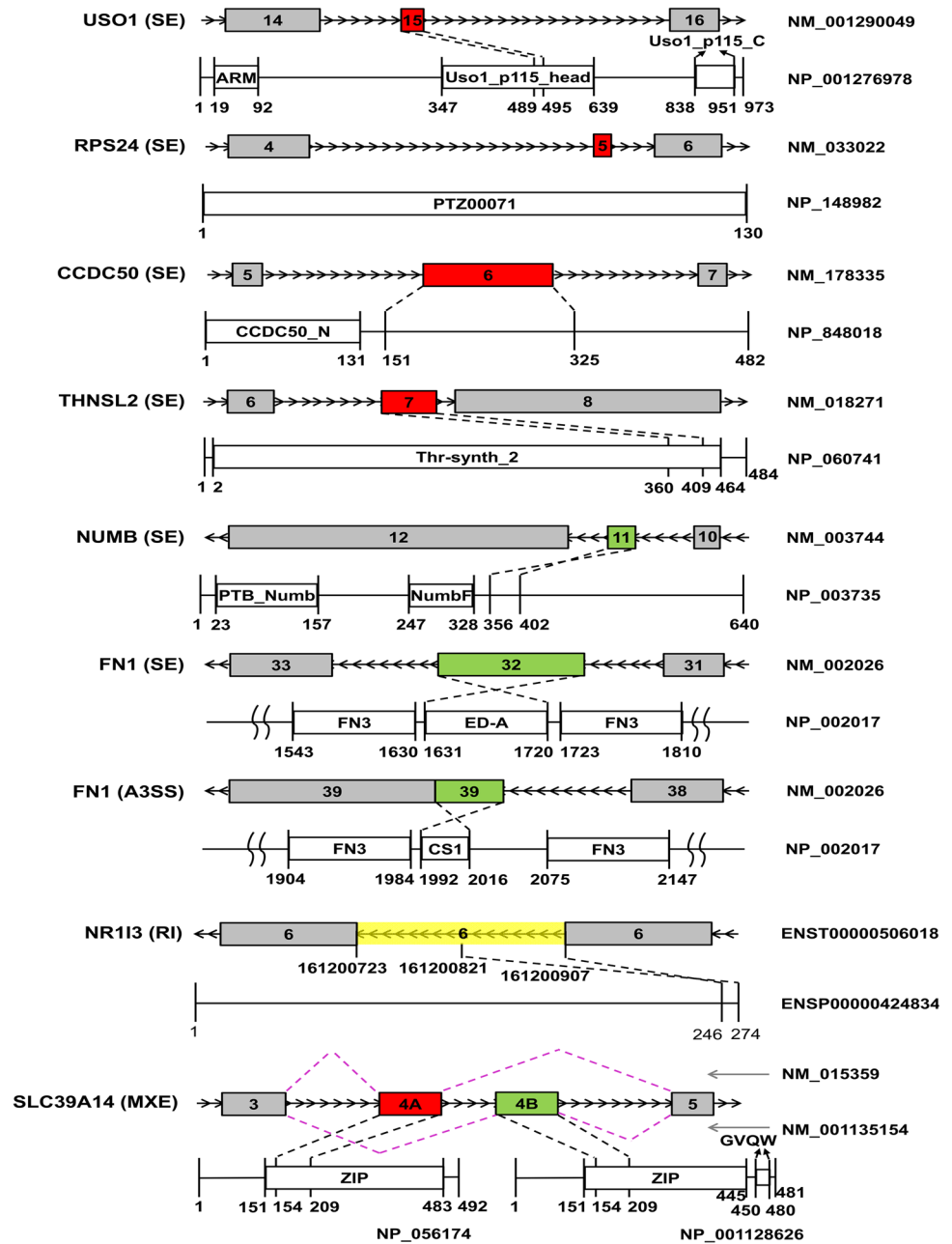
that the splicing dysregulation in HCC might be influenced by these splicing factors. Figure 7 was created based on the normalized counts of the three splicing factors.

Discussion

HCC is an aggressive and deadly cancer. To date, still there has been a startling lack of effective treatments for the disease. To further elucidate the mechanism of HCC tumorigenesis and identify potential biomarkers and therapeutic targets, this study investigated the transcriptome of HCC with a focus on lncRNAs and alternative splicing through integrative analysis of multiple datasets which revealed important transcriptional changes in HCC.

Accumulating evidence suggests the existence of vital regulatory roles of lncRNAs in cancer (Young et al. 2005; Yochum et al. 2007; Wang et al. 2008b; Tsai et al. 2010; Wang and Chang 2011; Carrieri et al. 2012; Bi et al. 2013). In the present study, we found that there are more lncRNAs expressed in HCC samples than in the adjacent normal samples as shown by all pairwise comparisons, which suggested a direct correlation of the number of expressed

Fig. 5 Illustration of nine alternative splicing isoforms up-regulated in relative abundance in the majority of the HCC samples compared with each corresponding adjacent normal sample. The isoforms were derived from *USO1*, *RPS24*, *CCDC50*, *THNSL2*, *NUMB*, *FN1*, *SLC39A14* and *NR1I3*. A red rectangle represents the exon is skipped from the isoform, while a green rectangle denotes the exon is included in the isoform. A yellow rectangle represents an intron retained in the isoform. For each isoform, the reference transcript and exon number were denoted in the plot. The translated amino acid sequence of each isoform was also provided with domain structure plotted and corresponding coordinates indicated to scale. The amino acid regions translated by the splicing sequences were shown in the plot. ARM, Armadillo/ beta-catenin-like repeats; *Uso1_p115_head*, *Uso1/p115* like vesicle tethering protein, head region; *Uso1_p115_c*, *Uso1/p115* like vesicle tethering protein, C terminal region; PTZ00071, 40S ribosomal protein S24; *CCDC50_N*, coiled-coil domain-containing protein 50 N-terminus; *Thr-synth_2*, threonine synthase; PTB_Numb, Numb phosphotyrosine-binding (PTB) domain; NumbF, NUMB domain; ED-A, extra-domain A; CS1, type III connecting segment 1; FN3, fibronectin type 3 domain; ZIP, ZIP zinc transporter; GVQW, putative binding domain (color figure online)



lncRNAs and HCC development. Further analysis identified 15 lncRNA genes which were only detectable in a number of HCC samples and may be potential biomarkers of HCC.

More than a hundred lncRNAs were detected differentially expressed between HCC and adjacent normal tissues. Some well-studied lncRNAs including *H19*, *MEG3*, *HAND2-AS1*, *RN7SK*, *LINC00261* and *TP53TG1* were identified down-regulated in HCC tissues in this study. Consistently, Zhang et al. (2013) found under-expressed *H19* in intratumoral HCC tissues (T) vs peritumoral tissues (L) and low *T/L* ratio of *H19* associated with poor prognosis.

Besides, several studies also displayed the down-regulation of tumor suppressor *MEG3* in HCC which is due to hypermethylation of *MEG3* in promoter region (Braconi et al. 2011; Zhuo et al. 2015). Of interest, this study detected *MEG3* co-expressed with two protein coding genes *LAMC3* and *C8orf58*. Significant methylation was found in the promoter region of *LAMC3* in breast cancer (Kuznetsova et al. 2007). *HAND2-AS1*, also known as *DEIN*, was first identified by Voth et al. (2007) displaying high expression level in stage IVS neuroblastoma. Similarly, Lin and Chuang (2012) found repressed expression of *LINC00261* in primary cultured invasive phenotype HCC cells compared

Table 4 List of highly recurrent aberrant alternative splicing events identified in HCC tissues in comparison with adjacent normal tissues

Gene symbol	Gene function	Splicing type	Affected protein domain	Incidence
<i>USO1</i>	Important in Golgi structure and mitosis progression (Nakajima et al. 1991; Radulescu et al. 2011)	SE ^a	Uso1_p115_head ^e	11/15
<i>RPS24</i>	Essential for processing of pre-rRNA and maturation of 40S ribosomal subunits (Choemsel et al. 2008)	SE		9/15
<i>CCDC50</i>	Involved in <i>EGFR</i> signaling (Blagoev et al. 2004)	SE		9/15
<i>THNSL2</i>	Encodes a threonine synthase-like protein	SE	Thr-synth_2 ^f	8/15
<i>NUMB</i>	Determines neuronal cell fate during development (Uemura et al. 1989)	SE		9/15
<i>FN1</i>	Involved in cell adhesion and migration processes (Hynes 1985; Yamada 1989)	SE	ED-A ^g	8/15
<i>SLC39A14</i>	Zinc transporter (Taylor et al. 2005)	MXE ^b	ZIP ^h	8/15
<i>NR1I3</i>	Key regulator of xenobiotic and endobiotic metabolism (Wei et al. 2000; Kodama et al. 2004; Konno et al. 2008)	RI ^c		10/15
<i>FN1</i>	Involved in cell adhesion and migration processes (Hynes 1985; Yamada 1989)	A3SS ^d	CS1 ⁱ	10/15

^a SE exon skipping

^b MXE mutually exclusive exons

^c RI retained intron

^d A3SS alternative 3' splice sites

^e *Uso1_p115_head* *Uso1/p115* like vesicle tethering protein, head region

^f *Thr-synth_2* threonine synthase

^g ED-A extra-domain A

^h ZIP ZIP zinc transporter

ⁱ CS1 type III connecting segment 1

to their corresponding parent cells. Gene *TP53TG1* was originally isolated from a colon cancer cell line which may have an effect on the signaling pathway of *TP53* and the response to cellular damage (Takei et al. 1998). In addition, we found some lncRNAs over-expressed in HCC tissues such as *GAS5*, *LINC00152*, *PVT1* and *SNHG1*. *GAS5* may have a pro-apoptotic attribute due to its repressive action on glucocorticoid receptor during starvation (Kino et al. 2010). Neumann et al. (2012) detected *LINC00152* as differentially hypomethylated during hepatocarcinogenesis. Studies by Wang et al. (2014) and Ding et al. (2015) supported our finding of up-regulation of *PVT1* in HCC and showed that *PVT1* positively regulates HCC cell proliferation and stemness by stabilizing *NOP2* nucleolar protein. Yan et al. (2015) reported a breakpoint between *c-Myc* and *PVT1* commonly detected in early onset HCC leading to the overexpression of *c-Myc* and *PVT1* in tumors. In addition, *SNHG1* was found up-regulated in prostate cancer as well (Berretta and Moscato 2010).

This study also investigated the co-expression profile between differentially expressed lncRNAs and differentially expressed protein coding genes. The rationale is that lncRNAs have a regulatory role in the transcription of many protein coding genes (Wang et al. 2008b; Tsai et al. 2010; Bi et al. 2013) and many co-expressed genes are found related in function such as involved in the same

signal transduction pathway. Consistent with Ren et al. (2012), most of the co-expression pairs in this study displayed positive correlation suggesting an enhancer-like role of these lncRNAs in regulating the transcription of the protein coding genes. Only two co-expression pairs were presented as negative correlation including one lncRNA *LINC01093* correlated with *TTK* and *NDC80*. The mRNA level of *TTK* and *NDC80* were elevated in HCC samples in this study. Both of the two genes participate in the regulation of mitosis (Chen et al. 1997; Dou et al. 2004; Huang et al. 2009; Sundin et al. 2011). *TTK* encoding a protein kinase was shown to be a promising prognostic marker in HCC (Miao et al. 2014). *NDC80* could be a treatment target which was suggested to be implicated in the pathogenesis of HCC (Liu et al. 2015). The negative correlation between the expression level of *LINC01093* and the expression level of *TTK* and *NDC80* found in this study indicated that *LINC01093* could serve as a prognostic biomarker and therapeutic target for HCC. *LINC01093* was nominated as a cancer-associated lncRNA in a recent study comprehensively delineating the landscape of human lncRNAs (Iyer et al. 2015). Generally, lncRNAs show higher cancer- and tissue-specificity compared to protein coding genes, suggesting that they can be powerful biomarkers and drug targets (Iyer et al. 2015; Sahu et al. 2015).

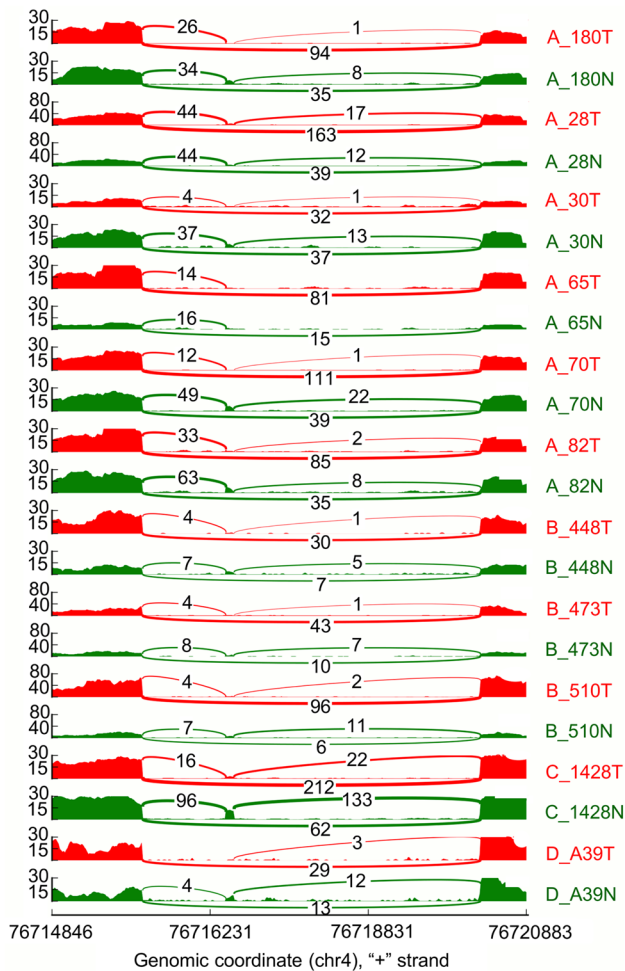


Fig. 6 Sashimi plot illustrating the differential exon skipping event of *USO1* in 11 patients undergoing this event. In the plot, red color represents HCC samples and green color represents adjacent normal samples. The exon excluding isoform of *USO1* is up-regulated in isoform percentage in each of the 11 HCC samples compared with the corresponding adjacent normal sample. The thickness of each arc is in proportion to the number of reads spanning the particular junction. The actual number of junction reads was also plotted in each arc. The x-axis represents the genomic coordinates. The y-axis represents the RPKM (Reads Per Kilobase per Million mapped reads) values (color figure online)

In many cases, lncRNAs seem to regulate the expression of their neighboring protein coding genes through different mechanisms (Guttman et al. 2009; Orom et al. 2010; Kambara et al. 2015). This study identified five lncRNAs correlated with six vicinal protein coding genes, among which *TMC8*, *LILRB5* and *LILRB2* were involved in immunity (Borges et al. 1997; Shiroishi et al. 2006; Crequer et al. 2013).

The correlation analysis also provided us with novel insights into the function of five lncRNAs, all of which were predicted to be involved in oxidation–reduction process.

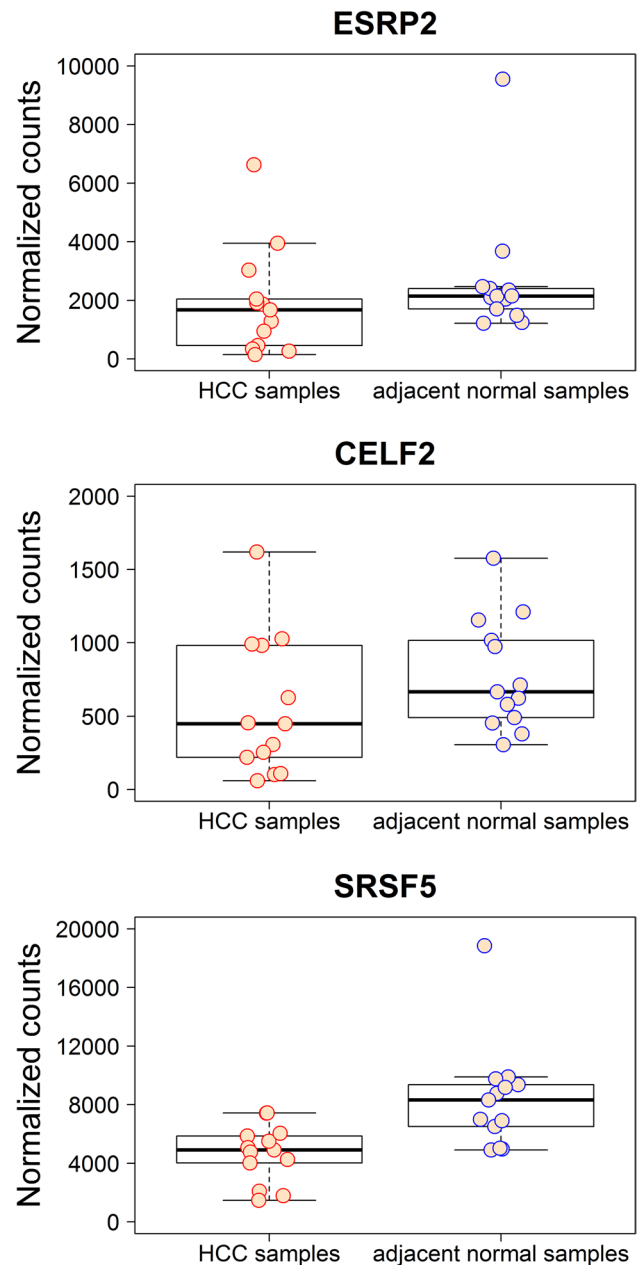


Fig. 7 Boxplots illustrating the expression level of three splicing factors *ESRP2*, *CELF2* and *SRSF5*. The y-axis reflects normalized counts of a splicing factor. The x-axis represents HCC and adjacent normal samples. For each splicing factor the normalized count of each sample was also plotted in the plot as represented by a red circle (filled circle) for each HCC sample and a blue circle (filled circle) for each adjacent normal sample (color figure online)

Studies have demonstrated the association between the dysregulation of redox and the carcinogenesis of HCC (Vali et al. 2008; Zhao et al. 2011; Lin et al. 2014a), which warrants further study of these five lncRNAs in the pathogenesis of HCC.

Another important aspect of HCC transcriptome examined in this study is differential splicing. The alternative

splicing profile of several genes is recurrently altered in HCC arguing for a direct role of specific splicing isoforms in HCC development. In the present study, nine highly recurrent aberrant splicing events were identified associated with HCC. As far as we know, this is the first study reporting the association between the splicing alterations of *SLC39A14* and *NR1I3* and HCC.

We identified an MXE event of *SLC39A14* with splicing of this gene shifted towards including exon 4B in the majority of HCC samples. The aberrant splicing is regulated by the Wnt pathway proposed by Thorsen et al. (2011) who found the alternative splicing of *SLC39A14* in colorectal tumors changing in the same way as our finding. Franklin et al. (2012) found the gene expression of *SLC39A14* was down-regulated in the hepatoma cells, which was verified by this study. They also deduced that the absence of *SLC39A14* may induce the depletion of zinc in hepatoma.

This study also identified an intron 7-retained isoform of gene *NR1I3* up-regulated in isoform percentage in the majority of HCC samples. The retention of intron 7 in *NR1I3* was first reported by Choi et al. (2013), which may lead to the production of proteins fail to transactivate the *CYP2B6* reporter gene (SV1, SV3, SV6) or produce protein with enhanced transactivation activity (SV2). *NR1I3* encodes a constitutive androstane receptor (CAR) regulating hepatic drug metabolism (Wei et al. 2000) and hepatic energy homeostasis (Kodama et al. 2004; Konno et al. 2008). Activation of CAR promotes liver injury and the development of HCC (Yamazaki et al. 2011; Kamino and Negishi 2012).

Another splicing isoform significantly up-regulated in isoform percentage in the majority of HCC samples is the oncofetal *FNI* variant containing the ED-A splice-in segment. The inclusion of ED-A exon was regulated by *PI3K/Akt/mTOR* (Blaustein et al. 2005; White et al. 2010) and multiple *MAPK* pathways (Al-Ayoubi et al. 2012). The alternative ED-A domain serves as a vascular marker of solid tumor and metastasis (Rybak et al. 2007) as well as contributes to the lymphangiogenesis of colorectal tumors (Ou et al. 2010). A high-affinity human anti-ED-A monoclonal antibody F8 has been generated targeting tumor neovasculature in vivo (Villa et al. 2008). Based on antibody F8 several potent anti-cancer biopharmaceuticals have been developed such as immunocytokines F8-IL13 (Hess and Neri 2015) and F8-IL2 (Pretto et al. 2014) and immunocytokine drug conjugate F8-IL2-SS-DM1 (List et al. 2014). On the other hand, splicing of *FNI* towards producing the isoform with CS1 exon was up-regulated in isoform percentage in the majority of HCC samples. The CS1 domain has a selective affinity for some tumor cells (Humphries et al. 1987) and lymphoid cells (Wayner et al. 1989).

The *NUMB* PRR^L isoform significantly increased in relative abundance in three-fifths of HCC samples in this

study was found linked to tumorigenesis of lung cancer (Misquitta-Ali et al. 2011). This isoform is formed by including an exon which reduces the level of *NUMB* protein expression and activates Notch signaling pathway (Misquitta-Ali et al. 2011). Recently, Lu et al. (2015) also observed a strong expression of PRR^L isoform in HCC. Bechara et al. (2013) found the *NUMB* PRR alternative splicing is regulated by RNA-binding motif proteins *RBM5*, *RBM6* and *RBM10* in the control of cancer cell proliferation.

Based on these splicing isoforms significantly up-regulated in isoform percentage in the majority of HCC tissues, it is possible to targetedly deliver anti-cancer molecules (i.e., cytotoxic drugs, cytokines, radionuclides, etc.) to the HCC site by binding molecules such as human antibodies specific to the isoforms (Schrama et al. 2006; List et al. 2014; Hess and Neri 2015). Another way to treat HCC might be through correcting these splicing isoforms by drugs which lead to the activation of NMD pathway or generation of inactive cell cycle genes as demonstrated by Kaida et al. (2007), Kotake et al. (2007), Chang et al. (2011) and Corrionero et al. (2011).

Alternative splicing is regulated by the interaction between sequence elements of the pre-mRNA and splicing factor proteins binding to them. Three splicing factors including *ESRP2*, *CELF2* and *SRSF5* were found down-regulated in HCC samples compared with adjacent normal samples, suggesting that the splicing alterations of HCC were influenced by these splicing factors. *ESRP2* and *ESRP1* were shown to be down-regulated in cells during an EMT (Warzecha et al. 2009) which is a sign of cancer progression. In addition, Xiao et al. (2014) suggested that *CELF2* induces apoptosis of HCC cells.

In summary, this study characterized HCC transcriptome with regard to lncRNAs and alternative splicing in a comprehensive way. By applying an integrative approach for the analysis of four RNA-Seq datasets, we observed that the number of lncRNAs expressed in each HCC sample was consistently greater than the adjacent normal sample. Furthermore, 15 lncRNAs were found expressed in five to seven HCC samples but were not detected in any adjacent normal sample. Based on differential expression analysis we detected 35 up- and 80 down-regulated lncRNAs in HCC samples compared with adjacent normal samples, among which five lncRNAs were predicted to be involved in oxidation and reduction process. Differential splicing analysis revealed nine highly recurrent differential splicing events belonging to eight genes *USO1*, *RPS24*, *CCDC50*, *THNSL2*, *NUMB*, *FNI*, *SLC39A14* and *NR1I3*. As far as we know, this is the first study reporting that aberrant splicing of *SLC39A14* and *NR1I3* is associated with HCC. Alternative splicing in HCC may be influenced by three splicing factors *ESRP2*, *CELF2* and *SRSF5* which were

significantly down-regulated in HCC samples. The findings of this study will add new information to aid in understanding the pathogenesis of HCC. The important molecules identified in this study are worthy of further investigation as potential biomarkers and possible therapeutic targets for the disease.

Acknowledgments This research was sponsored by Merck Sharp & Dohme (MSD) postdoc fellowship. We thank I-Ming Wang for good comments and Yunfei Pei, Hongchao Lu and Lan Chen for valuable discussion.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval For this type of study formal consent is not required.

References

- Al-Ayoubi AM, Zheng H, Liu Y, Bai T, Eblen ST (2012) Mitogen-activated protein kinase phosphorylation of splicing factor 45 (SPF45) regulates SPF45 alternative splicing site utilization, proliferation, and cell adhesion. *Mol Cell Biol* 32:2880–2893
- Anders S, Pyl PT, Huber W (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169
- Bechara EG, Sebestyen E, Bernardis I, Eyraas E, Valcarcel J (2013) RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation. *Mol Cell* 52:720–733
- Berasain C, Goni S, Castillo J, Latasa MU, Prieto J, Avila MA (2010) Impairment of pre-mRNA splicing in liver disease: mechanisms and consequences. *World J Gastroenterol* 16:3091–3102
- Berretta R, Moscato P (2010) Cancer biomarker discovery: the entropic hallmark. *PLoS One* 5:e12262
- Bi HS, Yang XY, Yuan JH, Yang F, Xu D, Guo YJ, Zhang L, Zhou CC, Wang F, Sun SH (2013) H19 inhibits RNA polymerase II-mediated transcription by disrupting the hnRNP U-actin complex. *Biochim Biophys Acta* 1830:4899–4906
- Biamonti G, Bonomi S, Gallo S, Ghigna C (2012) Making alternative splicing decisions during epithelial-to-mesenchymal transition (EMT). *Cell Mol Life Sci* 69:2515–2526
- Blagojev B, Ong SE, Kratchmarova I, Mann M (2004) Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat Biotechnol* 22:1139–1145
- Blaustein M, Pelisch F, Tanos T, Munoz MJ, Wengier D, Quadrana L, Sanford JR, Muschietti JP, Kornbliht AR, Caceres JF, Coso OA, Srebrow A (2005) Concerted regulation of nuclear and cytoplasmic activities of SR proteins by AKT. *Nat Struct Mol Biol* 12:1037–1044
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120
- Borges L, Hsu ML, Fanger N, Kubin M, Cosman D (1997) A family of human lymphoid and myeloid Ig-like receptors, some of which bind to MHC class I molecules. *J Immunol* 159:5192–5196
- Braconi C, Kogure T, Valeri N, Huang N, Nuovo G, Costinean S, Negrini M, Miotto E, Croce CM, Patel T (2011) microRNA-29 can regulate expression of the long non-coding RNA gene MEG3 in hepatocellular cancer. *Oncogene* 30:4750–4756
- Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, Fedele S, Pesce E, Ferrer I, Collavin L, Santoro C, Forrest AR, Carninci P, Biffo S, Stupka E, Gustincich S (2012) Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* 491:454–457
- Chan TH, Lin CH, Qi L, Fei J, Li Y, Yong KJ, Liu M, Song Y, Chow RK, Ng VH, Yuan YF, Tenen DG, Guan XY, Chen L (2014) A disrupted RNA editing balance mediated by ADARs (Adenosine Deaminases that act on RNA) in human hepatocellular carcinoma. *Gut* 63:832–843
- Chang KH, Li R, Papari-Zareei M, Watumull L, Zhao YD, Auchus RJ, Sharifi N (2011) Dihydrotestosterone synthesis bypasses testosterone to drive castration-resistant prostate cancer. *Proc Natl Acad Sci USA* 108:13728–13733
- Chen Y, Riley DJ, Chen PL, Lee WH (1997) HEC, a novel nuclear protein rich in leucine heptad repeats specifically involved in mitosis. *Mol Cell Biol* 17:6049–6056
- Choemsel V, Fribourg S, Aguisa-Toure AH, Pinaud N, Legrand P, Gazda HT, Gleizes PE (2008) Mutation of ribosomal protein RPS24 in Diamond-Blackfan anemia results in a ribosome biogenesis disorder. *Hum Mol Genet* 17:1253–1263
- Choi EJ, Jang YJ, Cha EY, Shin JG, Lee SS (2013) Identification and characterization of novel alternative splice variants of human constitutive androstane receptor in liver samples of Koreans and Caucasians. *Drug Metab Dispos* 41:888–896
- Corrionero A, Minana B, Valcarcel J (2011) Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A. *Genes Dev* 25:445–459
- Crequer A, Picard C, Pedergrana V, Lim A, Zhang SY, Abel L, Majewski S, Casanova JL, Jablonska S, Orth G, Jouanguy E (2013) EVER2 deficiency is associated with mild T-cell abnormalities. *J Clin Immunol* 33:14–21
- Danan-Gotthold M, Golan-Gerstl R, Eisenberg E, Meir K, Karni R, Levanon EY (2015) Identification of recurrent regulated alternative splicing events across human solid tumors. *Nucleic Acids Res* 43:5130–5144
- David CJ, Manley JL (2010) Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev* 24:2343–2364
- Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan X, Ruan Y, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M, Davis CA, Shiekhattar R, Gingeras TR, Hubbard TJ, Notredame C, Harrow J, Guigo R (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 22:1775–1789
- Ding C, Yang Z, Lv Z, Du C, Xiao H, Peng C, Cheng S, Xie H, Zhou L, Wu J, Zheng S (2015) Long non-coding RNA PVT1 is associated with tumor progression and predicts recurrence in hepatocellular carcinoma patients. *Oncol Lett* 9:955–963
- Dou Z, Ding X, Zereshki A, Zhang Y, Zhang J, Wang F, Sun J, Huang H, Yao X (2004) TTK kinase is essential for the centrosomal localization of TACC2. *FEBS Lett* 572:51–56
- Franklin RB, Levy BA, Zou J, Hanna N, Desouki MM, Bagasra O, Johnson LA, Costello LC (2012) ZIP14 zinc transporter down-regulation and zinc depletion in the development and progression of hepatocellular cancer. *J Gastrointest Cancer* 43:249–257
- Ghigna C, Valacca C, Biamonti G (2008) Alternative splicing and tumor progression. *Curr Genom* 9:556–570
- Gill RB, Day A, Barstow A, Zaman G, Chenu C, Dhoot GK (2012) Mammalian Sulf1 RNA alternative splicing and its significance to tumour growth regulation. *Tumour Biol* 33:1669–1680

- Giulietti M, Piva F, D'Antonio M, D'Onorio De Meo P, Paoletti D, Castrignano T, D'Erchia AM, Picardi E, Zambelli F, Principato G, Pavesi G, Pesole G (2013) SpliceAid-F: a database of human splicing factors and their RNA-binding sites. *Nucleic Acids Res* 41:D125–D131
- Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458:223–227
- Harn HJ, Ho LI, Yu CP, Wang MW, Lee HS, Lin JJ, Lee WH, Isola NR, Cooper DL (1994) The variant mRNA isoform of human metastasis gene (CD44 V) detected in the cell lines of human hepatocellular carcinoma. *Biochem Mol Biol Int* 32:233–238
- Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G, Saunders G, Steward C, Harte R, Lin M, Howald C, Tanzer A, Derrien T, Chrast J, Walters N, Balasubramanian S, Pei B, Tress M, Rodriguez JM, Ezkurdia I, van Baren J, Brent M, Haussler D, Kellis M, Valencia A, Reymond A, Gerstein M, Guigo R, Hubbard TJ (2012) GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 22:1760–1774
- Hess C, Neri D (2015) The antibody-mediated targeted delivery of interleukin-13 to syngeneic murine tumors mediates a potent anticancer activity. *Cancer Immunol Immunother*
- Huang YF, Chang MD, Shieh SY (2009) TTK/hMps1 mediates the p53-dependent postmitotic checkpoint by phosphorylating p53 at Thr18. *Mol Cell Biol* 29:2935–2944
- Huang Q, Lin B, Liu H, Ma X, Mo F, Yu W, Li L, Li H, Tian T, Wu D, Shen F, Xing J, Chen ZN (2011) RNA-Seq analyses generate comprehensive transcriptomic landscape and reveal complex transcript patterns in hepatocellular carcinoma. *PLoS One* 6:e26168
- Humphries MJ, Komoriya A, Akiyama SK, Olden K, Yamada KM (1987) Identification of two distinct regions of the type III connecting segment of human plasma fibronectin that promote cell type-specific adhesion. *J Biol Chem* 262:6886–6892
- Hynes R (1985) Molecular biology of fibronectin. *Annu Rev Cell Biol* 1:67–90
- Ishibashi M, Kogo R, Shibata K, Sawada G, Takahashi Y, Kurashige J, Akiyoshi S, Sasaki S, Iwaya T, Sudo T, Sugimachi K, Mimori K, Wakabayashi G, Mori M (2013) Clinical significance of the expression of long non-coding RNA HOTAIR in primary hepatocellular carcinoma. *Oncol Rep* 29:946–950
- Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, Barrette TR, Prensner JR, Evans JR, Zhao S, Poliakov A, Cao X, Dhanasekaran SM, Wu YM, Robinson DR, Beer DG, Feng FY, Iyer HK, Chinnaiyan AM (2015) The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet* 47:199–208
- Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D (2011) Global cancer statistics. *CA Cancer J Clin* 61:69–90
- 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 (2007) Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nat Chem Biol* 3:576–583
- Kambara H, Gunawardane L, Zebrowski E, Kostadinova L, Jobava R, Krokowski D, Hatzoglu M, Anthony DD, Valadkhan S (2015) Regulation of interferon-stimulated gene BST2 by a lncRNA transcribed from a shared bidirectional promoter. *Front Immunol* 5:676
- Kamino H, Negishi M (2012) The nuclear receptor constitutive active/androstane receptor arrests DNA-damaged human hepatocellular carcinoma Huh7 cells at the G2/M phase. *Mol Carcinog* 51:206–212
- Kan Z, Zheng H, Liu X, Li S, Barber TD, Gong Z, Gao H, Hao K, Willard MD, Xu J, Hauptschein R, Rejto PA, Fernandez J, Wang G, Zhang Q, Wang B, Chen R, Wang J, Lee NP, Zhou W, Lin Z, Peng Z, Yi K, Chen S, Li L, Fan X, Yang J, Ye R, Ju J, Wang K, Estrella H, Deng S, Wei P, Qiu M, Wulur IH, Liu J, Ehsani ME, Zhang C, Loboda A, Sung WK, Aggarwal A, Poon RT, Fan ST, Wang J, Hardwick J, Reinhard C, Dai H, Li Y, Luk JM, Mao M (2013) Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome Res* 23:1422–1433
- Kang L, Liu X, Gong Z, Zheng H, Wang J, Li Y, Yang H, Hardwick J, Dai H, Poon RT, Lee NP, Mao M, Peng Z, Chen R (2015) Genome-wide identification of RNA editing in hepatocellular carcinoma. *Genomics* 105:76–82
- Katz Y, Wang ET, Airoidi EM, Burge CB (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods* 7:1009–1015
- Kim D, Perte G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14:R36
- Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP (2010) Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci Signal* 3:ra8
- Kodama S, Koike C, Negishi M, Yamamoto Y (2004) Nuclear receptors CAR and PXR cross talk with FOXO1 to regulate genes that encode drug-metabolizing and gluconeogenic enzymes. *Mol Cell Biol* 24:7931–7940
- Konno Y, Negishi M, Kodama S (2008) The roles of nuclear receptors CAR and PXR in hepatic energy metabolism. *Drug Metab Pharmacokinet* 23:8–13
- Kotake Y, Sagane K, Owa T, Mimori-Kiyosue Y, Shimizu H, Uesugi M, Ishihama Y, Iwata M, Mizui Y (2007) Splicing factor SF3b as a target of the antitumor natural product pladienolide. *Nat Chem Biol* 3:570–575
- Kuznetsova EB, Kekeeva TV, Larin SS, Zemliakova VV, Babenko OV, Nemtsova MV, Zaletaev DV, Strel'nikov VV (2007) Novel methylation and expression markers associated with breast cancer. *Mol Biol (Mosk)* 41:624–633
- Lai MC, Yang Z, Zhou L, Zhu QQ, Xie HY, Zhang F, Wu LM, Chen LM, Zheng SS (2012) Long non-coding RNA MALAT-1 overexpression predicts tumor recurrence of hepatocellular carcinoma after liver transplantation. *Med Oncol* 29:1810–1816
- Lin ZY, Chuang WL (2012) Genes responsible for the characteristics of primary cultured invasive phenotype hepatocellular carcinoma cells. *Biomed Pharmacother* 66:454–458
- Lin B, Tan X, Liang J, Wu S, Liu J, Zhang Q, Zhu R (2014a) A reduction in reactive oxygen species contributes to dihydromyricetin-induced apoptosis in human hepatocellular carcinoma cells. *Sci Rep* 4:7041
- Lin KT, Shann YJ, Chau GY, Hsu CN, Huang CY (2014b) Identification of latent biomarkers in hepatocellular carcinoma by ultra-deep whole-transcriptome sequencing. *Oncogene* 33:4786–4794
- List T, Casi G, Neri D (2014) A chemically defined trifunctional antibody–cytokine–drug conjugate with potent antitumor activity. *Mol Cancer Ther* 13:2641–2652
- Liu B, Yao Z, Hu K, Huang H, Xu S, Wang Q, Yang Y, Ren J (2015) ShRNA-mediated silencing of the Ndc80 gene suppress cell proliferation and affected hepatitis B virus-related hepatocellular carcinoma. *Clin Res Hepatol Gastroenterol*
- Lu Y, Xu W, Ji J, Feng D, Sourbier C, Yang Y, Qu J, Zeng Z, Wang C, Chang X, Chen Y, Mishra A, Xu M, Lee MJ, Lee S, Trepel J, Linehan WM, Wang X, Yang Y, Neckers L (2015) Alternative

- splicing of the cell fate determinant Numb in hepatocellular carcinoma. *Hepatology*
- Matouk IJ, DeGroot N, Mezan S, Ayesh S, Abu-lail R, Hochberg A, Galun E (2007) The H19 non-coding RNA is essential for human tumor growth. *PLoS One* 2:e845
- Matsui S, Takahashi T, Oyanagi Y, Takahashi S, Boku S, Takahashi K, Furukawa K, Arai F, Asakura H (1997) Expression, localization and alternative splicing pattern of fibronectin messenger RNA in fibrotic human liver and hepatocellular carcinoma. *J Hepatol* 27:843–853
- Miao R, Luo H, Zhou H, Li G, Bu D, Yang X, Zhao X, Zhang H, Liu S, Zhong Y, Zou Z, Zhao Y, Yu K, He L, Sang X, Zhong S, Huang J, Wu Y, Miksad RA, Robson SC, Jiang C, Zhao Y, Zhao H (2014) Identification of prognostic biomarkers in hepatitis B virus-related hepatocellular carcinoma and stratification by integrative multi-omics analysis. *J Hepatol* 61:840–849
- Misquitta-Ali CM, Cheng E, O'Hanlon D, Liu N, McGlade CJ, Tsao MS, Blencowe BJ (2011) Global profiling and molecular characterization of alternative splicing events misregulated in lung cancer. *Mol Cell Biol* 31:138–150
- Nakajima H, Hirata A, Ogawa Y, Yonehara T, Yoda K, Yamasaki M (1991) A cytoskeleton-related gene, *uso1*, is required for intracellular protein transport in *Saccharomyces cerevisiae*. *J Cell Biol* 113:245–260
- Neumann O, Kesselmeier M, Geffers R, Pellegrino R, Radlwimmer B, Hoffmann K, Ehemann V, Schemmer P, Schirmacher P, Lorenzo Bermejo J, Longrich T (2012) Methyloome analysis and integrative profiling of human HCCs identify novel protumorigenic factors. *Hepatology* 56:1817–1827
- Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Busotti G, Lai F, Zytnicki M, Notredame C, Huang Q, Guigo R, Shiekhattar R (2010) Long noncoding RNAs with enhancer-like function in human cells. *Cell* 143:46–58
- Ou JJ, Wu F, Liang HJ (2010) Colorectal tumor derived fibronectin alternatively spliced EDA domain exerts lymphangiogenic effect on human lymphatic endothelial cells. *Cancer Biol Ther* 9:186–191
- Oyama F, Hirohashi S, Shimosato Y, Titani K, Sekiguchi K (1989) Deregulation of alternative splicing of fibronectin pre-mRNA in malignant human liver tumors. *J Biol Chem* 264:10331–10334
- Oyama F, Hirohashi S, Sakamoto M, Titani K, Sekiguchi K (1993) Coordinate oncogene developmental modulation of alternative splicing of fibronectin pre-messenger RNA at ED-A, ED-B, and CS1 regions in human liver tumors. *Cancer Res* 53:2005–2011
- Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 40:1413–1415
- Panzitt K, Tschernatsch MM, Guelly C, Moustafa T, Stradner M, Strohmaier HM, Buck CR, Denk H, Schroeder R, Trauner M, Zatloukal K (2007) Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as non-coding RNA. *Gastroenterology* 132:330–342
- Parkin DM (2006) The global health burden of infection-associated cancers in the year 2002. *Int J Cancer* 118:3030–3044
- Pretto F, Elia G, Castioni N, Neri D (2014) Preclinical evaluation of IL2-based immunocytokines supports their use in combination with dacarbazine, paclitaxel and TNF-based immunotherapy. *Cancer Immunol Immunother* 63:901–910
- Radulescu AE, Mukherjee S, Shields D (2011) The Golgi protein p115 associates with gamma-tubulin and plays a role in Golgi structure and mitosis progression. *J Biol Chem* 286:21915–21926
- Rau A, Marot G, Jaffrezic F (2014) Differential meta-analysis of RNA-seq data from multiple studies. *BMC Bioinform* 15:91
- Ren S, Peng Z, Mao JH, Yu Y, Yin C, Gao X, Cui Z, Zhang J, Yi K, Xu W, Chen C, Wang F, Guo X, Lu J, Yang J, Wei M, Tian Z, Guan Y, Tang L, Xu C, Wang L, Gao X, Tian W, Wang J, Yang H, Wang J, Sun Y (2012) RNA-seq analysis of prostate cancer in the Chinese population identifies recurrent gene fusions, cancer-associated long noncoding RNAs and aberrant alternative splicings. *Cell Res* 22:806–821
- Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 11:R25
- Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140
- Rybak JN, Roesli C, Kaspar M, Villa A, Neri D (2007) The extracellular domain A of fibronectin is a vascular marker of solid tumors and metastases. *Cancer Res* 67:10948–10957
- Sahu A, Singhal U, Chinnaiyan AM (2015) Long noncoding RNAs in cancer: from function to translation. *Trends Cancer* 1:93–109
- Schrama D, Reisfeld RA, Becker JC (2006) Antibody targeted drugs as cancer therapeutics. *Nat Rev Drug Discov* 5:147–159
- Shiroishi M, Kuroki K, Rasubala L, Tsumoto K, Kumagai I, Kurimoto E, Kato K, Kohda D, Maenaka K (2006) Structural basis for recognition of the nonclassical MHC molecule HLA-G by the leukocyte Ig-like receptor B2 (LILRB2/LIR2/ILT4/CD85d). *Proc Natl Acad Sci USA* 103:16412–16417
- Shumway M, Cochrane G, Sugawara H (2010) Archiving next generation sequencing data. *Nucleic Acids Res* 38:D870–D871
- Snider NT, Altshuler PJ, Wan S, Welling TH, Cavalcoli J, Omary MB (2014) Alternative splicing of human NT5E in cirrhosis and hepatocellular carcinoma produces a negative regulator of ecto-5'-nucleotidase (CD73). *Mol Biol Cell* 25:4024–4033
- Sundin LJ, Guimaraes GJ, Deluca JG (2011) The NDC80 complex proteins Nuf2 and Hec1 make distinct contributions to kinetochore-microtubule attachment in mitosis. *Mol Biol Cell* 22:759–768
- Takei Y, Ishikawa S, Tokino T, Muto T, Nakamura Y (1998) Isolation of a novel TP53 target gene from a colon cancer cell line carrying a highly regulated wild-type TP53 expression system. *Genes Chromosomes Cancer* 23:1–9
- Taylor KM, Morgan HE, Johnson A, Nicholson RI (2005) Structure-function analysis of a novel member of the LIV-1 subfamily of zinc transporters, ZIP14. *FEBS Lett* 579:427–432
- Thorsen K, Mansilla F, Schepeler T, Oster B, Rasmussen MH, Dyrskjot L, Karni R, Akerman M, Krainer AR, Laurberg S, Andersen CL, Orntoft TF (2011) Alternative splicing of SLC39A14 in colorectal cancer is regulated by the Wnt pathway. *Mol Cell Proteom* 10(M110):002998
- Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY (2010) Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329:689–693
- Uemura T, Shepherd S, Ackerman L, Jan LY, Jan YN (1989) Numb, a gene required in determination of cell fate during sensory organ formation in *Drosophila* embryos. *Cell* 58:349–360
- Vali L, Hahn O, Kupcsulik P, Drahos A, Sarvary E, Szentmihalyi K, Pallai Z, Kurucz T, Sipos P, Blazovics A (2008) Oxidative stress with altered element content and decreased ATP level of erythrocytes in hepatocellular carcinoma and colorectal liver metastases. *Eur J Gastroenterol Hepatol* 20:393–398
- Villa A, Trachsel E, Kaspar M, Schliemann C, Somavilla R, Rybak JN, Rosli C, Borsi L, Neri D (2008) A high-affinity human monoclonal antibody specific to the alternatively spliced EDA domain of fibronectin efficiently targets tumor neo-vasculature in vivo. *Int J Cancer* 122:2405–2413
- Voth H, Oberthuer A, Simon T, Kahlert Y, Berthold F, Fischer M (2007) Identification of DEIN, a novel gene with high expression levels in stage IV neuroblastoma. *Mol Cancer Res* 5:1276–1284
- Wang KC, Chang HY (2011) Molecular Mechanisms of Long Non-coding RNAs. *Mol Cell* 43:904–914

- Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB (2008a) Alternative isoform regulation in human tissue transcriptomes. *Nature* 456:470–476
- Wang X, Arai S, Song X, Reichart D, Du K, Pascual G, Tempst P, Rosenfeld MG, Glass CK, Kurokawa R (2008b) Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* 454:126–130
- Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10:57–63
- Wang F, Yuan JH, Wang SB, Yang F, Yuan SX, Ye C, Yang N, Zhou WP, Li WL, Li W, Sun SH (2014) Oncofetal long noncoding RNA PVT1 promotes proliferation and stem cell-like property of hepatocellular carcinoma cells by stabilizing NOP2. *Hepatology* 60:1278–1290
- Warzecha CC, Sato TK, Nabet B, Hogenesch JB, Carstens RP (2009) ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. *Mol Cell* 33:591–601
- Wayner EA, Garcia-Pardo A, Humphries MJ, McDonald JA, Carter WG (1989) Identification and characterization of the T lymphocyte adhesion receptor for an alternative cell attachment domain (CS-1) in plasma fibronectin. *J Cell Biol* 109:1321–1330
- Wei P, Zhang J, Egan-Hafley M, Liang S, Moore DD (2000) The nuclear receptor CAR mediates specific xenobiotic induction of drug metabolism. *Nature* 407:920–923
- White ES, Sagana RL, Booth AJ, Yan M, Cornett AM, Bloomheart CA, Tsui JL, Wilke CA, Moore BB, Ritzenthaler JD, Roman J, Muro AF (2010) Control of fibroblast fibronectin expression and alternative splicing via the PI3K/Akt/mTOR pathway. *Exp Cell Res* 316:2644–2653
- Xiao Z, Ching Chow S, Han Li C, Chun Tang S, Tsui SK, Lin Z, Chen Y (2014) Role of microRNA-95 in the anticancer activity of Brucein D in hepatocellular carcinoma. *Eur J Pharmacol* 728:141–150
- Yamada KM (1989) Fibronectin domains and receptors. In: Mosher DF (ed) *Fibronectin*. Academic Press, San Diego, pp 47–121
- Yamazaki Y, Moore R, Negishi M (2011) Nuclear receptor CAR (NR1H3) is essential for DDC-induced liver injury and oval cell proliferation in mouse liver. *Lab Invest* 91:1624–1633
- Yan H, Yang Y, Zhang L, Tang G, Wang Y, Xue G, Zhou W, Sun S (2015) Characterization of the genotype and integration patterns of hepatitis B virus in early- and late-onset hepatocellular carcinoma. *Hepatology*
- Yochum GS, Cleland R, McWeeney S, Goodman RH (2007) An antisense transcript induced by Wnt/beta-catenin signaling decreases E2F4. *J Biol Chem* 282:871–878
- Young TL, Matsuda T, Cepko CL (2005) The noncoding RNA taurine upregulated gene 1 is required for differentiation of the murine retina. *Curr Biol* 15:501–512
- Yuan SX, Yang F, Yang Y, Tao QF, Zhang J, Huang G, Yang Y, Wang RY, Yang S, Huo XS, Zhang L, Wang F, Sun SH, Zhou WP (2012) Long noncoding RNA associated with microvascular invasion in hepatocellular carcinoma promotes angiogenesis and serves as a predictor for hepatocellular carcinoma patients poor recurrence-free survival after hepatectomy. *Hepatology* 56:2231–2241
- Zhang L, Yang F, Yuan JH, Yuan SX, Zhou WP, Huo XS, Xu D, Bi HS, Wang F, Sun SH (2013) Epigenetic activation of the MiR-200 family contributes to H19-mediated metastasis suppression in hepatocellular carcinoma. *Carcinogenesis* 34:577–586
- Zhang H, Ye J, Weng X, Liu F, He L, Zhou D, Liu Y (2015) Comparative transcriptome analysis reveals that the ECM-receptor interaction contributes to the venous metastases of hepatocellular carcinoma. *Cancer Genet*
- Zhao J, Zhao Y, Wang H, Gu X, Ji J, Gao C (2011) Association between metabolic abnormalities and HBV related hepatocellular carcinoma in Chinese: a cross-sectional study. *Nutr J* 10:49
- Zhao J, Greene CM, Gray SG, Lawless MW (2014) Long noncoding RNAs in liver cancer: what we know in 2014. *Expert Opin Ther Targets* 18:1207–1218
- Zhuo H, Tang J, Lin Z, Jiang R, Zhang X, Ji J, Wang P, Sun B (2015) The aberrant expression of MEG3 regulated by UHRF1 predicts the prognosis of hepatocellular carcinoma. *Mol Carcinog*