

# Invasion and metastasis-related long noncoding RNA expression profiles in hepatocellular carcinoma

Yunzhen Gao<sup>1,2</sup> · Geng Chen<sup>1,2</sup> · Yongyi Zeng<sup>1,2,3</sup> · Jinhua Zeng<sup>1,2,3</sup> · Minjie Lin<sup>1,2</sup> · Xiaolong Liu<sup>1,2</sup> · Jingfeng Liu<sup>1,2,3</sup>

Received: 25 December 2014 / Accepted: 29 March 2015 / Published online: 22 April 2015  
© International Society of Oncology and BioMarkers (ISOBM) 2015

**Abstract** Recurrence, invasion, and metastasis are the major reasons of the low 5-year survival of hepatocellular carcinoma. However, the mechanisms of recurrence, invasion, and metastasis are still poorly understood. Long noncoding RNAs (LncRNAs, >200 nt) have been demonstrated to play important roles in both tumor suppressive and oncogenic signaling pathways. Here, we employed the LncRNAs microarray technology to study the LncRNAs expression profiles at genome-wide in hepatocellular carcinoma (HCC) tissue samples with early recurrence (less than 1 year, with invasion and metastasis out of liver) and late recurrence (longer than 2 years, without invasion and metastasis out of liver), which had different recurrent/metastatic potentials, by using normal liver tissue as control to screen the dysregulated LncRNAs which are potentially involved in the recurrence, invasion, and metastasis

process of HCC. Overall, 1170 LncRNAs were identified to differentially expressed between the early and late recurrence samples. These differentially expressed LncRNAs were further characterized by integrating examination of genomic context, co-expression network analysis, and gene ontology (GO) enrichment of their associated protein-coding genes. Furthermore, 15 LncRNAs selected randomly from top 50 differentially expressed LncRNAs were validated by quantitative PCR (qPCR) in cell lines MHCC97H and MHCC97L, which have exactly the same genetic background but with different invasion potentials. Meanwhile, the prognostic potential of three verified LncRNAs at cell line level was further validated in 59 HCC samples. Therefore, our results demonstrated that the aberrant expression of LncRNAs might be responsible for the HCC invasion and metastasis and provide fundamental information for further study the LncRNAs involved molecular mechanisms of the invasion and metastasis of HCC.

**Electronic supplementary material** The online version of this article (doi:10.1007/s13277-015-3408-0) contains supplementary material, which is available to authorized users.

✉ Xiaolong Liu  
xiaoloong.liu@gmail.com

✉ Jingfeng Liu  
drjingfeng@126.com

<sup>1</sup> The United Innovation of Mengchao Hepatobiliary Technology Key Laboratory of Fujian Province, Mengchao Hepatobiliary Hospital of Fujian Medical University, Xihong Road 312, Fuzhou 350025, Fujian Province, People's Republic of China

<sup>2</sup> The Liver Center of Fujian Province, Fujian Medical University, Fuzhou 350025, People's Republic of China

<sup>3</sup> Liver Disease Center, The First Affiliated Hospital of Fujian Medical University, Fuzhou 350005, People's Republic of China

**Keywords** Long noncoding RNA · Expression profiles · Hepatocellular carcinoma · Metastasis · Recurrence · Invasion

## Introduction

Hepatocellular carcinoma (HCC) is one of the most common cancers and is currently the third common cause of cancer-related mortality worldwide [1]. The highest HCC incidence is found in the sub-Saharan Africa and East Asia and HCC amounted to 90 % of primary liver cancer in China [2]. Currently, the therapeutic strategies to treat HCC are including surgical resection, loco-regional ablation, and liver transplantation. Although the 5-year survival has a gradual increase in recent decades due to the advances in surgical techniques and medical treatments,

**Table 1** The primers used in qPCR analysis

LncRNA ID	Forward primer	Reverse primer	Product length (bp)	Tm (°C)
P23099	ACAGCAGGTTGGCACAAAAC	GTCTGCCTAATGTCCGGGTC	299	60
P6488	CTTGATGGAGGAACAGGAA	GGAGCTCCACTGATGTCTCT	192	60
P700	GTGATGACACACTCTGTGCT	GAGGAAGTGCTGTCAGCAGA	166	60
P14695	TCTTGCTCACAACAAGGCAG	AGGATCTTTCATGTGCTGTGT	179	60
P6391	CACAGAATGTCCGGTCAGGGT	CCGCTGAAAAGCAACTGGTA	174	60
P8611	GCCATGTCAATTCAGGTCTC	CTGGACGGAGAATGAGATTG	232	60
P24363	CCCCAAAGTGGTATCAGGC	GTGCTGGCTTGCTGATGAAC	156	60
P8860	CAATGGCCAGCACACAATGA	GAGAGAACAACACCCTGCT	199	60
P16984	ACTCAACTTCCCCATGTGACT	GGTTTTGTTTGCTGCTACGG	175	60
P19780	GATGCTGAACAGCGGAACAC	CTTGGGATACATGGGCAGGG	190	60
P28210	AGTACAGCTCTCCTTCCCTGA	CAATCGCAAGCCAGATGTCA	216	60
P33863	CAATCGCAAGCCAGATGTCA	ACTCACCCCGATCTCTCTCC	177	60
P4091	CACCGGAAGGAACCAATTCT	GCTCTTGCTCAAGGCCTGTT	138	60
P8725	GCCATGGTCTGCGATCCTA	GTTCTTGTTGTCGTGTGCT	203	60
P9745	TCACCTTCCAACCAAGTCTCA	GCTGGTCCATTGGCATAACA	235	60

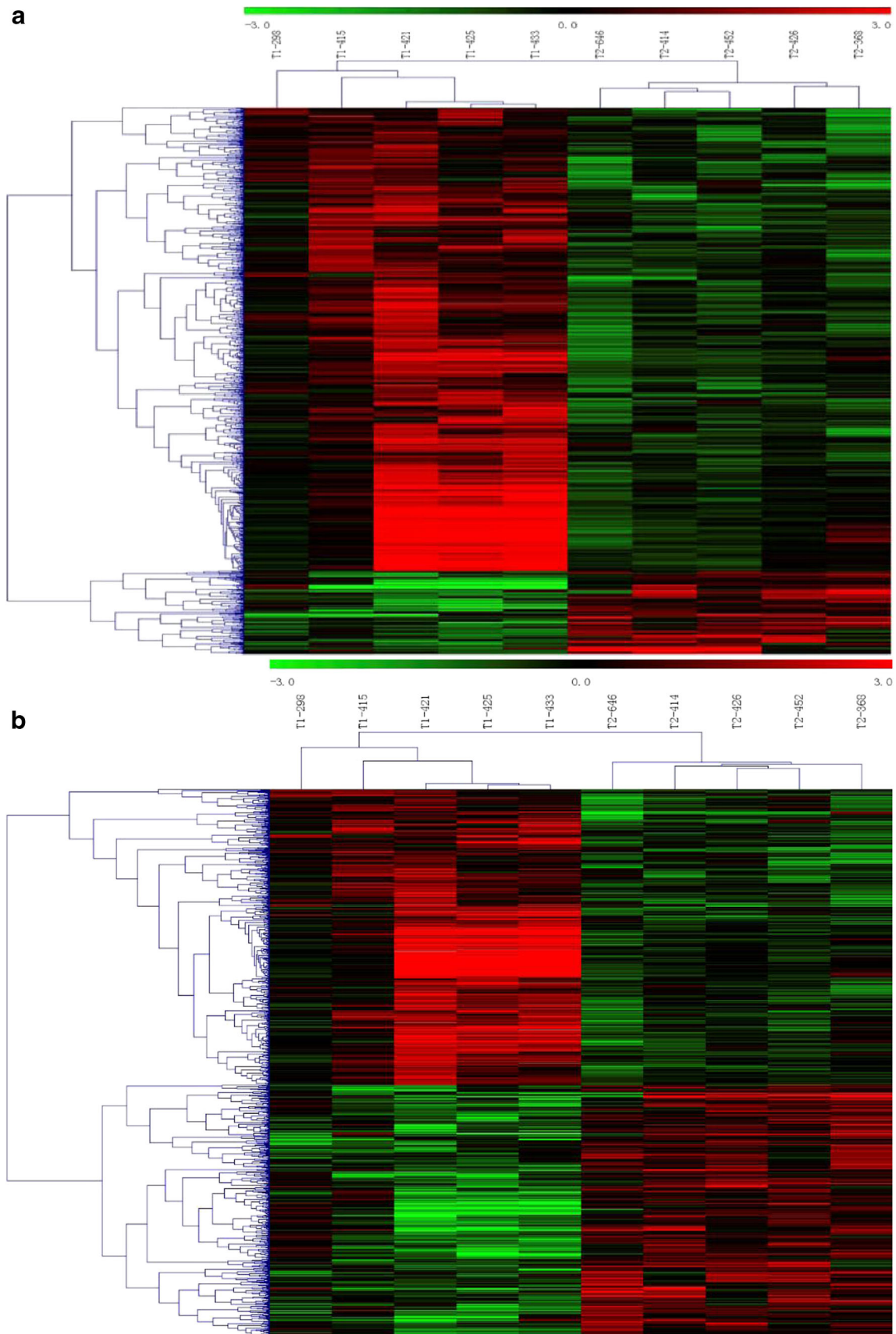
for the majority of HCC patients, it was less than 20 % [3]. Recurrence, invasion, and metastasis are the major causes of the low 5-year survival rate [4], approximately 50–70 % of HCC patients suffer from recurrence/metastasis within 12 months after curative resection [5]. Therefore, it is an urgent need to discover novel molecular biomarkers of the recurrence, invasion, and metastasis, which can help in early diagnosis and prognosis assessment, and better understand the underlying molecular mechanisms.

Long noncoding RNAs (LncRNAs), previously thought as noise of transcription and generally defined as having a size bigger than 200 nucleotides, have come into the limelight as functional molecules [6]. Recent analysis and experiments suggest that LncRNAs exhibit tissue-specific expression patterns and may play important roles in a broad range of biological process [7]. LncRNAs could regulate gene expression at epigenetic, transcription, and post-transcriptional level such as DNA methylation, gene silencing, and histone modifications [8]. Dysregulation of LncRNAs is closely associated with many human diseases, including various cancers [9]. More recently, there are several reports on the functions of LncRNAs in HCC such as HOXA transcript at the distal tip (HOTTIP) [10], highly upregulated in liver cancer (HULC) [11, 12], and LncRNA-High Expressed In HCC (HEIH) [13]. It has been reported that the LncRNA could activate by TGF- $\beta$  (LncRNA-ATB), upregulate in HCC metastases and associate with poor prognosis, induce EMT and invasion through upregulating ZEB1 and ZEB2 [14]. LncRNA-hPVT1 could promote cell

proliferation, cell cycling, and acquisition of stem cell-like properties in HCC cells by stabilizing NOP2 [15]. Opposite to these tumor promotion functions of LncRNAs, it has also been reported that the LncRNA metallothionein 1D pseudogene (MT1DP) could act as a tumor suppressor, overexpression of which resulted in reduced cell proliferation and colony formation in sofa agar and increased apoptosis in liver cancer cells; furthermore, MT1DP could negatively regulate the expression of alpha-fetoprotein through inhibiting protein synthesis of forkhead box A1, an important transcription factor in liver development and cancer progression [16]. Since several LncRNAs have been reported to dysregulate in HCC and then facilitate the tumor growth, the LncRNAs might be potential valuable diagnostic biomarkers even therapeutic targets for HCC treatment. However, the roles of LncRNAs in the recurrence, invasion, and metastasis of HCC are largely unknown.

Here, we have investigated the roles of LncRNAs in the recurrence and invasion/metastasis of HCC by comparing the LncRNAs expression profiles between HCC tissues that have recurrence within 1 year and metastasis outside the liver (T1) and HCC tissues that have recurrence after 2 years and

**Fig. 1** Hierarchical clustering of the differentially expressed LncRNAs and mRNAs in the primary HCC tissues from patients group T1 and T2; T1 represents patients who had recurrence within 1 year after surgical resection and with outside liver metastasis; T2 represents patients who had recurrence after 2 years of surgical resection and without outside liver metastasis. **a** Differentially expressed LncRNAs; **b** differentially expressed mRNAs. *Red color* indicates the upregulation, and *green color* indicates the down regulation



**Table 2** Summary of microarray analysis results

Group	Probe class	Total	Expression above background (%)
T1	LncRNA	37,581	27,618 (73.49)
	mRNA	34,235	29,850 (87.19)
T2	LncRNA	37,581	26,864 (71.48)
	mRNA	34,235	29,587 (86.42)
NT	LncRNA	37,581	27,576 (73.38)
	mRNA	34,235	29,778 (86.98)

without outside liver metastasis (T2), with the normal liver tissues (NT) as control.

## Methods

### Ethics statement

The project was approved for the using of human biopsy by the Institution Review Board of Mengchao Hepatobiliary Hospital of Fujian Medical University. The written consent was received from all participants in this study at the time of surgery.

### Clinical samples

Total of 10 fresh-frozen primary HCC tissues and 5 fresh-frozen normal liver tissues were included in the microarray assay. Those tissues in microarray assay were divided into the following three groups: primary HCC tissues from patients who had recurrence within 1 year after surgical resection and with outside liver metastasis (T1), primary HCC tissues from patients who had recurrence after 2 years (or even more) after surgical resection and without outside liver metastasis (T2), and the normal liver tissues (NT). The clinical samples for quantitative PCR (qPCR) validation experiments were divided into the following two groups: primary HCC tissues from patients who had outside liver invasion and metastasis ( $n=25$ ) and primary HCC tissues from patients without outside liver invasion and metastasis ( $n=34$ ). Fresh-frozen HCC tissues and normal liver tissue samples were collected during the

surgical resection at Mengchao Hepatobiliary Hospital of Fujian Medical University and stored in the tissue bank for further usage.

### RNA extraction

TRIzol reagent (Invitrogen, Carlsbad, CA, USA) was used to extract the total RNAs from tissues, and mirVana miRNA Isolation Kit (Ambion, Austin, TX, USA) was used to purify small RNAs according to the manufacturer's protocol. The concentration and purity of RNAs were determined by OD260/280 readings using spectrophotometer (NanoDrop ND-2000). By using the RNA 6000 Nano Lab-on-a-Chip kit and the Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA), RNA integrity was determined by capillary electrophoresis. Only RNA extracts with RNA integrity number higher than 6 undergone further analysis.

### Fabrication of DNA microarray

Microarray assay was performed by CapitalBio Corporation, Beijing, China. The Agilent Human LncRNA + messenger RNA (mRNA) Array V3.0 was used in the assay which was designed with four identical arrays per slide ( $4 \times 180\text{-K}$  format). This microarray contains 37,582 human LncRNAs probes collected from GENCODE/Ensembl, Human LncRNA Catalog [17], RefSeq, USCS, ncRNA Expression Database (NRED), H-InvDB, LncRNAs-a (enhancer-like), RNadb, CombinedLit, Antisense ncRNA pipeline, HoxncRNAs, UCRs, and Chen Runsheng lab (Institute of Biophysics, Chinese Academy of Science), 34,235 human mRNAs probes, and 4974 Agilent control probes. Each RNA was detected by corresponding probes repeated for two times.

### RNA amplification, labeling, and hybridization

By using Eberwine's linear RNA amplification method and subsequent enzymatic reaction, fluorescent dye (Cy3) labeled complementary DNAs (cDNAs) were produced according to the manufacturer's protocol of CapitalBio cRNA Amplification and Labeling Kit (CapitalBio).

**Table 3** Summary of differentially expressed LncRNAs and mRNAs in the three comparisons as indicated (P1, T1 vs. NT; P2, T2 vs. NT; P3, T1 vs. T2)

Group	Probe class	Differentially expressed (%)	Upregulated	Downregulated
P1	LncRNA	1780 (6.45)	362	1428
	mRNA	3313 (11.1)	1436	1877
P2	LncRNA	2720 (10.13)	392	2328
	mRNA	4160 (13.94)	2269	1891
P3	LncRNA	1170 (4.24)	992	168
	mRNA	953 (3.2)	515	438

**Table 4** Overview of the differentially expressed LncRNAs and mRNAs in P1 (T1 vs. NT) and P2 (T2 vs. NT) but not in P3 (T1 vs. T2)

Group	Probe class	Differentially expressed	Upregulated	Downregulated
P1	LncRNA	1428	317	1111
	mRNA	2967	1390	1577
P2	LncRNA	2108	352	1756
	mRNA	3682	2141	1543

In detail, 1  $\mu\text{g}$  total RNA was used to synthesize the double-stranded cDNAs (containing the T7 RNA polymerase promoter sequence) using the CbcScript reverse transcriptase with cDNA synthesis system according to the manufacturer's protocol (Capitalbio) with the T7 Oligo (dT) and T7 Oligo (dN). PCR NucleoSpin Extract II Kit (MN) was used to purify dsDNA by eluting in 30  $\mu\text{L}$  elution buffer. Then, the eluted double-stranded DNA was vacuum concentrated to 16  $\mu\text{L}$ . Subsequently, the cRNA was amplified using a T7 Enzyme Mix in a 40- $\mu\text{L}$  in vitro transcription reactions system at 37  $^{\circ}\text{C}$  for 14 h and then purified by the RNA Clean-up Kit (MN). Two micrograms amplified RNA was further reverse transcribed using CbcScript reverse transcriptase, then the reverse transcribed cDNA products were purified by PCR NucleoSpin Extract II Kit (MN) and vacuum concentrated to 14  $\mu\text{L}$ . Klenow enzyme was used to label the cDNA at 37  $^{\circ}\text{C}$  for 90 min. The labeled cDNA was further purified by a PCR NucleoSpin Extract II Kit (MN). Test samples labeled with Cy3-dCTP, as well as the labeled controls were dissolved in 80  $\mu\text{L}$  hybridization solution (3 $\times$  saline-sodium citrate (SSC), 0.2 % sodium dodecyl sulfate (SDS), 5 $\times$  Denhardt's solution and 25 % formamide) and denatured at 95  $^{\circ}\text{C}$  for 3 min prior to loading onto a microarray. At a rotation speed of 20 rpm and a temperature of 42  $^{\circ}\text{C}$  overnight, array hybridization was performed in an Agilent Hybridization Oven. Afterwards, the hybridized arrays were washed with two consecutive solutions

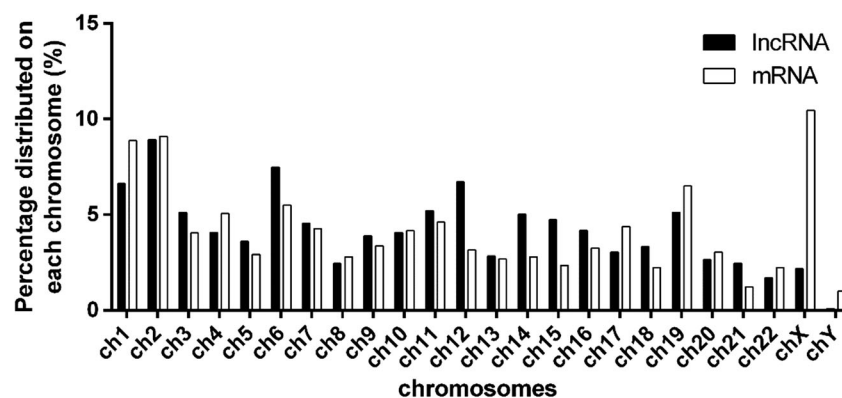
(0.2 % SDS, 2 $\times$  SSC at 42  $^{\circ}\text{C}$  for 5 min, followed by 0.2 $\times$  SSC for 5 min at room temperature).

### Microarray imaging and data analysis

GeneSpring software V12.0 (Agilent) was used to analyze the LncRNAs+mRNA array data for data summarization, normalization, and quality control. Threshold values of  $\geq 2$ - and  $\leq -2$ -fold change and a Benjamini-Hochberg corrected  $P$  value of  $\leq 0.05$  were used to select the differentially expressed genes. The data was log 2 transformed and median centered by genes using the Adjust Data function of Multiexperiment Viewer (MeV) software (Dana-Farber Cancer Institute, Boston, MA, USA), then further analyzed with hierarchical clustering with average linkage. The microarray data has been deposited in the NCBI Gene Expression Omnibus (GEO) database (accession number: GSE67260).

### Construction of the coding-noncoding gene co-expression network

The coding-noncoding gene co-expression network (CNC network) was constructed based on the correlation analysis between the differential expressed LncRNAs and mRNAs. The Pearson correlation was calculated for each



**Fig. 2** The distribution of differentially expressed LncRNAs and mRNAs on each chromosome in the comparison between T1 and T2; T1 represents patients who had recurrence within 1 year after surgical resection and with outside liver metastasis; T2 represents patients who had recurrence after 2 years of surgical resection and without outside liver

metastasis. The percentage of LncRNAs and mRNAs distributed on each chromosome was calculated by the numbers of LncRNAs and mRNAs on each chromosome dividing the total numbers of differentially expressed LncRNAs and mRNAs; *ch* represents chromosome

pair of genes, and the significant correlation pairs were chosen to construct the network. The network was drawn through the open source bioinformatics software Cytoscape using LncRNAs and mRNAs with Pearson correlation coefficients higher than 0.99. Degree centrality is defined as the link numbers one node has to the other in a network analysis. A degree is the simplest and most important measure of a gene centrality within a network that determines the relative importance [18].

### Gene ontology and pathway analysis

The Molecular Annotation System (CB-MAS) V3.0 (<http://bioinfo.capitalbio.com/mas3/>) was used to perform gene ontology and pathway analysis.

### qPCR analysis

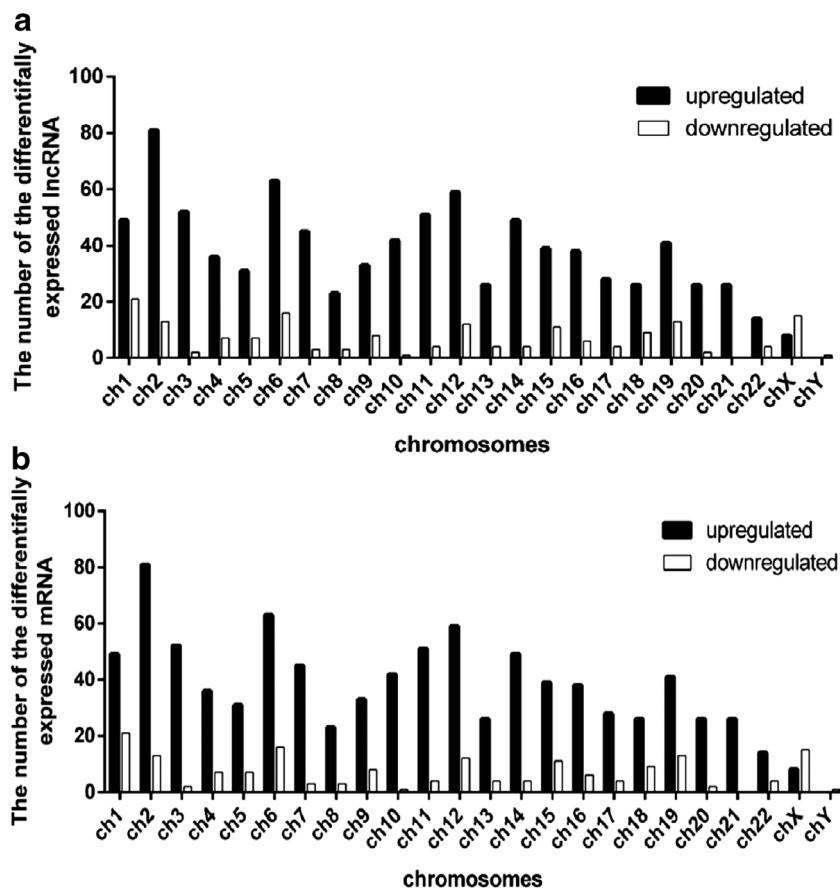
Total RNAs of MHCC97H and MHCC97L, as well as 59 clinical samples were isolated by using TRIzol reagent (Invitrogen, Carlsbad, CA, USA); the RNA purity and concentration were determined by using spectrophotometer (NanoDrop ND-2000). The samples with OD260/280 ratio higher than 2 were then reversely transcribed using a

GoScript™ Reverse Transcription System Kit (Promega, Madison, WI, USA) in accordance with the manufacturer's instructions. The expressions of selected LncRNAs were analyzed using qPCR with a Go Taq®qPCR Master Mix kit (Promega, Madison, WI, USA) on StepOne Plus PCR System (Applied Biosystems, Carlsbad, CA, USA) with following cycling parameters: 95 °C for 2 min in the pre-denature stage; 15 s at 95 °C, 20 s at 60 °C, 20 s at 72 °C (collect the signature at this step), with a total of 40 cycles; melt curve stage (95 °C for 15 s, 60 °C for 1 min, and 90 °C for 30 s, with a reading signature of per 0.3 °C from 60 to 90 °C, system default).  $\beta$ -Actin was used as an internal control, and  $2^{-\Delta\Delta CT}$  method was used to calculate the expression of LncRNAs. The primers are listed in Table 1.

### Statistical analysis

All the data were represented as mean  $\pm$  standard deviation. Student's *t* test was used to compare two variables of microarray data, and  $P < 0.01$  was considered statistically significant. qPCR data were analyzed using Student's *t* test with SPSS (version 15) and  $P < 0.05$  was considered as statistically significant.

**Fig. 3** The distribution of up- and downregulated LncRNAs (a) and mRNAs (b) on each chromosome in the comparison between T1 and T2; T1 represents patients who had recurrence within 1 year after surgical resection and with outside liver metastasis; T2 represents patients who had recurrence after 2 years of surgical resection and without outside liver metastasis



**Table 5** Overview of the subgroups of the differentially expressed LncRNAs in three comparisons as indicated (P1, T1 vs. NT; P2, T2 vs. NT, P3, T1 vs. T2)

Group	Expression levels	Sense LncRNAs	Antisense LncRNAs	Intronic LncRNAs	Bidirection LncRNAs	Intergenic LncRNAs
P1	Upregulated	63	97	11	28	138
	Downregulated	195	289	284	123	404
P2	Upregulated	49	79	16	45	131
	Downregulated	126	190	123	80	249
P3	Upregulated	65	126	110	55	173
	Downregulated	26	24	5	13	45

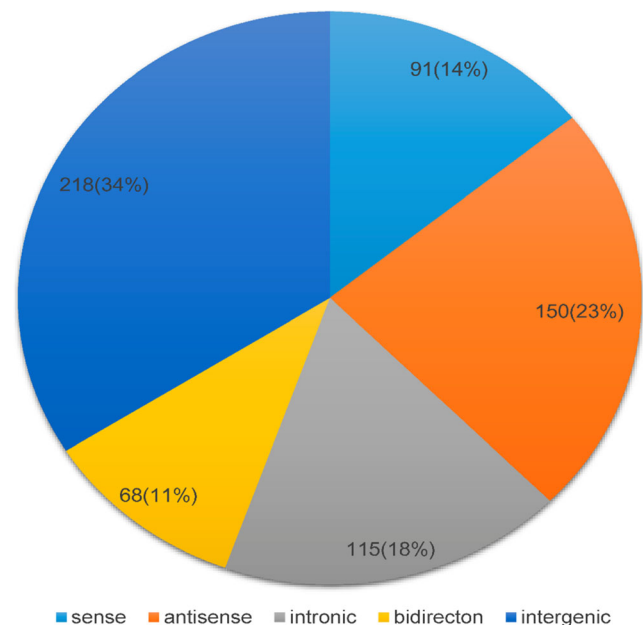
## Results

### Overview of the LncRNA and mRNA profiles in the early and late recurrent/metastatic HCC tissues

Here, we applied a commercial human LncRNA microarray (CapitalBio, Beijing, China) to study the characteristic of LncRNA expression profiles related to the recurrence/metastasis behavior, by using total RNA isolated from different HCC tissues with different recurrence/metastasis potentials, including the primary HCC tissues from patients who had recurrence within 1 year after surgical resection and with outside liver metastasis (group T1), the primary HCC tissues from patients who had recurrence after 2 years of surgical resection and without outside liver metastasis (group T2), and the normal liver tissues (group NT, as the control). The characteristic of LncRNA profiles related to the recurrence/metastasis was carefully analyzed by comparing T1 group versus NT group (P1), T2 group versus NT group (P2), and the T1 group versus T3 group (P3).

Based on the expression level of the LncRNAs, a hierarchical clustering analysis was performed to group LncRNAs and mRNAs (Fig. 1 and Supplemental Fig. 1). The analyzed results of LncRNA expression profiles are summarized in Table 2. The comparison of T1 versus NT (P1) and T2 versus NT (P2) was performed to analyze the tumor-specific LncRNA profiles, while the identification of recurrence/metastasis-specific LncRNA profiles was performed through comparing the T1 and T2 group (P3). The alternation of the expression levels of LncRNAs was evaluated by fold change. As shown in Table 2, although there were similar percentages of LncRNAs and protein-coding mRNAs expressed above the background, P1 and P2 had more significantly differentially expressed LncRNAs and protein-coding mRNAs compared with P3. In the comparison between T1 and T2 group (P3), 73.38 % (27,576 out of 37,581) LncRNAs on the microarray expressed above background and 4.24 % (1170 out of 27,576) of it were significantly differentially expressed between the T1 and T2 groups (absolute fold change  $\geq 2$ ,  $P \leq 0.05$ , Table 3). In those significantly differentially expressed LncRNAs, 84.79 % (992 out of 1170, Table 3) was upregulated. By contrast, 86.98 % (29,778 out of 34,235) of protein-coding

mRNA transcripts on the microarray was expressed above background, and 3.2 % (953 out of 34,235) of it were significantly differentially expressed between the T1 and T2 groups; meanwhile, 54.04 % (515 out of 953) mRNA was significantly upregulated in those differentially expressed protein-coding mRNAs (Table 3). Interestingly, compared with 20.34 % (362 out of 1780) and 14.41 % (392 out of 2720) upregulated LncRNAs in P1 and P2, respectively, P3 had 84.79 % (992 out of 1160) upregulated LncRNAs (Table 3). With regard to the differentially expressed mRNAs (absolute fold change  $\geq 2$ ,  $P \leq 0.05$ ), upregulated mRNAs were more common than downregulated mRNAs in P2 and P3. The P1 had more downregulated protein-coding mRNAs (56.66 %, 1877 out of 3313, Table 3). Furthermore, our data revealed that less LncRNAs was detected above the background compared to protein-coding genes. This result was similar to previous observations [19, 20] and indicated that the expression of LncRNAs



**Fig. 4** The subgroups of the differentially expressed LncRNAs in the comparison between T1 and T2; T1 represents patients who had recurrence within 1 year after surgical resection and with outside liver metastasis; T2 represents patients who had recurrence after 2 years of surgical resection and without outside liver metastasis

exhibited a larger temporal and spatial specificity than protein-coding genes.

To make a deep insight into the tumorigenesis process of HCC, we analyzed the LncRNAs which had significant difference in P1 and P2, but without significant difference in P3 (Table 4). Here, we discovered 1428 and 2108 these kind of significantly differentially expressed LncRNAs in P1 and P2, respectively. This portion of LncRNAs may participate in the tumorigenesis of HCC.

Although, the LncRNAs and mRNA expressed above background in human HCC tissues were widely scattered on all chromosomes, it seems that LncRNAs and protein-coding mRNAs related to the recurrence/metastasis of HCC were not distributed equally on the all chromosomes. After analysis of the distribution patterns of differentially expressed LncRNAs and mRNAs on chromosomes, we found that chromosome 2 (ch2) had the highest percentage of the expression level, altered LncRNAs and protein-coding mRNAs in all the three comparisons (Figs. 2 and 3, Supplemental Fig. 2 and Supplemental Fig. 3).

Interestingly, we found that the number of LncRNAs transcribed from X chromosome is much larger in the downregulated group than the upregulated group in all three groups. In contrast, the number of LncRNAs from all other chromosomes seems to be randomly distributed in the up- and downregulated groups.

The P3 group had lesser significantly differentially expressed LncRNAs and protein-coding mRNAs; this may be because there are more similarities between the early-metastasis and late-metastasis tissues than that of the cancer and normal tissues. So, we focus on the P3 group in further analysis.

### LncRNAs classification and subgroup analysis

According to the difference in transcription form, LncRNAs can be classified or subgrouped to sense LncRNAs, antisense LncRNAs, intronic LncRNAs, intergenic LncRNAs, and bidirectional LncRNAs. Many sense LncRNAs, which have the same transcriptional direction and overlap with exons of protein-coding genes, can be considered as noncoding transcript variants of protein-coding genes [21], and these LncRNAs can regulate the expression of their associated protein-coding genes [22]. We found that there were 65 upregulated sense LncRNAs and 26 downregulated sense LncRNAs comparing the T1 and T2 groups (P3, Table 5 and Fig. 4).

Anti-sense LncRNAs, which were transcribed against and overlapping with the protein-coding genes, regulate their protein-coding counterparts via multiple mechanisms, including alternative splicing, chromatin remodeling, translational interference, translational promotion, and promoter targeting [23]. As shown in Table 5 and Fig. 4, 126 anti-sense LncRNAs were significantly upregulated and 24 anti-sense

**Fig. 5** The co-expression network of the upregulated LncRNAs (a) and downregulated LncRNAs (b) from the comparison between the patient groups T1 and T2, with all differentially expressed mRNAs. The correlation  $>0.99$  or correlation  $<-0.99$  and  $P<0.05$  were recognized as co-expression. *Yellow ring* indicates LncRNAs, *green ring* indicates mRNAs, *red line* indicates positive association, and *green line* indicates negative association. The size of the node indicates the node degrees (the number of its neighbors)

LncRNAs were significantly downregulated comparing the T1 and T2 groups (P3).

Intronic LncRNAs, which transcribed from intronic regions of protein-coding genes, reside in a large portion of mammalian transcriptional units [24], could regulate the expression of their neighborhood genes or host protein-coding genes via a range of mechanisms, including RNA interference, microRNA, alternative splicing, chromatin modification, and transcriptional disruption [25]. In the current study, we identified 110 upregulated and 5 downregulated intronic LncRNAs comparing the T1 and T2 groups (P3, Table 5 and Fig. 4).

There are approximately 10 % of known uncoding genes as bidirectional genes, which have pairs of transcription initiation sites from two different transcripts that have opposite orientation but in close proximity ( $<1000$  bp) [26]. Bidirectional LncRNAs can repress or promote the expression of their neighboring protein-coding genes by epigenetic modification [27]. Among the significantly expressed LncRNAs, we found that there were 55 upregulated LncRNAs and 13 downregulated LncRNAs that possessed a bidirectional pair in the comparison of T1 and T2 groups (P3, Table 5 and Fig. 4).

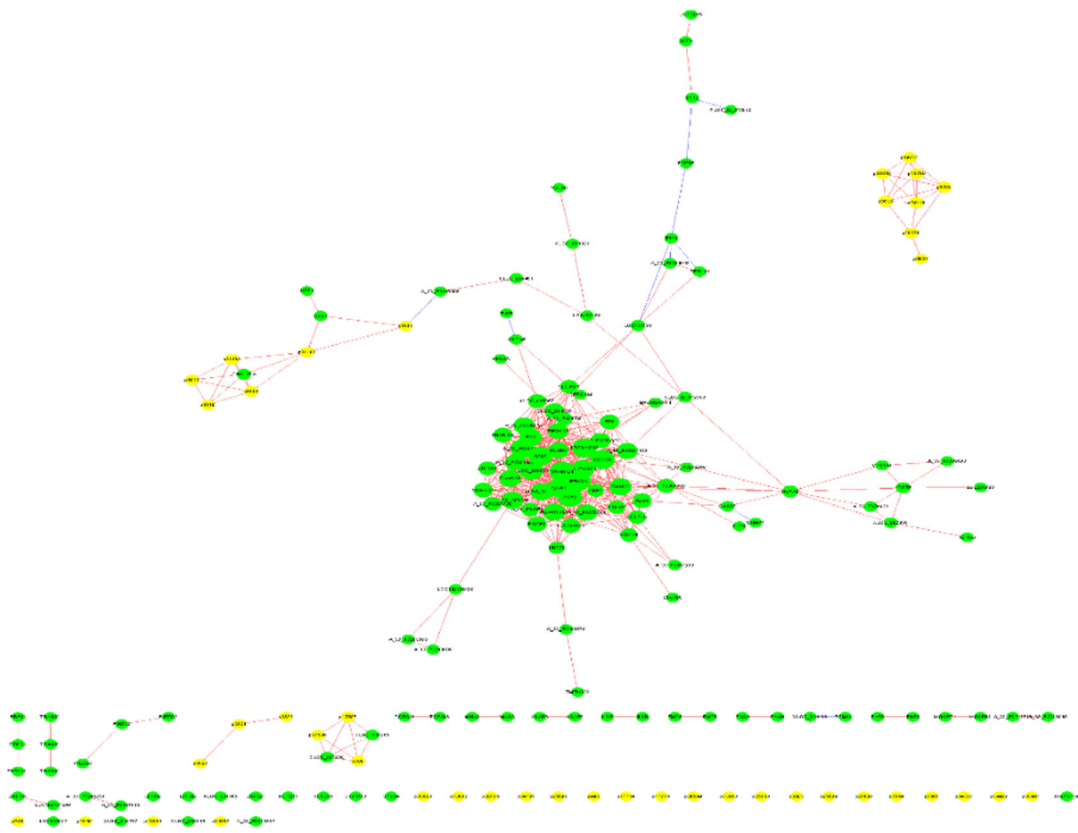
Long intergenic noncoding RNAs (LincRNAs) that are 10-kb apart from protein-coding genes can modulate the expression of target genes via directly recruiting histone-modifying enzymes to the chromatin [28]. Their target genes can scatter across the genome. We identify 173 upregulated LincRNAs and 45 downregulated LincRNAs in comparison to the T1 and T2 groups (P3, Table 5 and Fig. 4).

### Construction of the coding-noncoding gene co-expression network

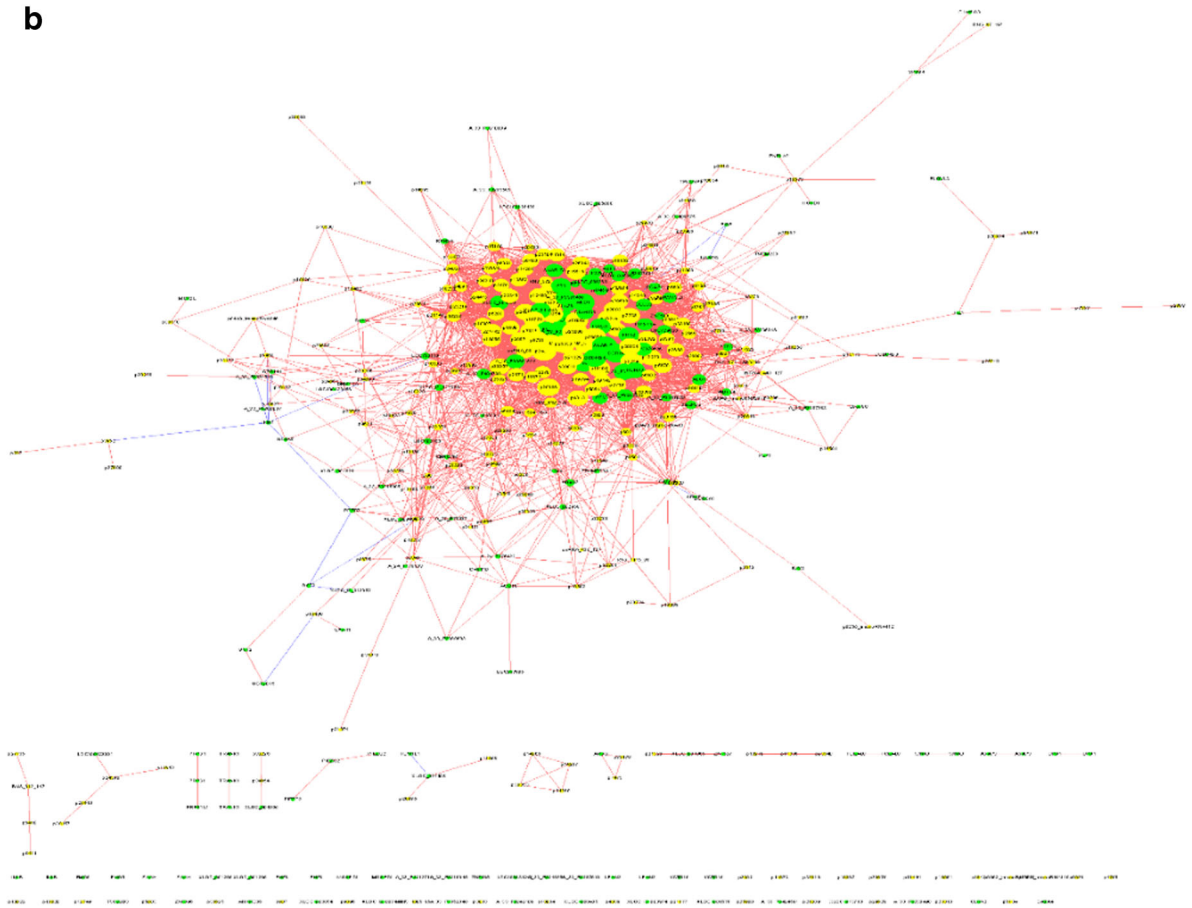
Based on the correlation analysis between the significantly differentially expressed LncRNAs and mRNAs, we constructed a coding-noncoding gene co-expression network (CNC network). CNC network was drawn through the open source bioinformatics software Cytoscape using LncRNAs and mRNAs with Pearson correlation coefficients higher than 0.99. We constructed two CNC networks, using up- and downregulated LncRNAs in each comparison, upregulated CNC network, and downregulated CNC network. Among the upregulated CNC network in P3, 315 upregulated LncRNAs and 111 mRNAs composed the CNC network node. These 426 network nodes made 852 associated network pairs of co-expression LncRNAs and mRNAs (Fig. 5a). One

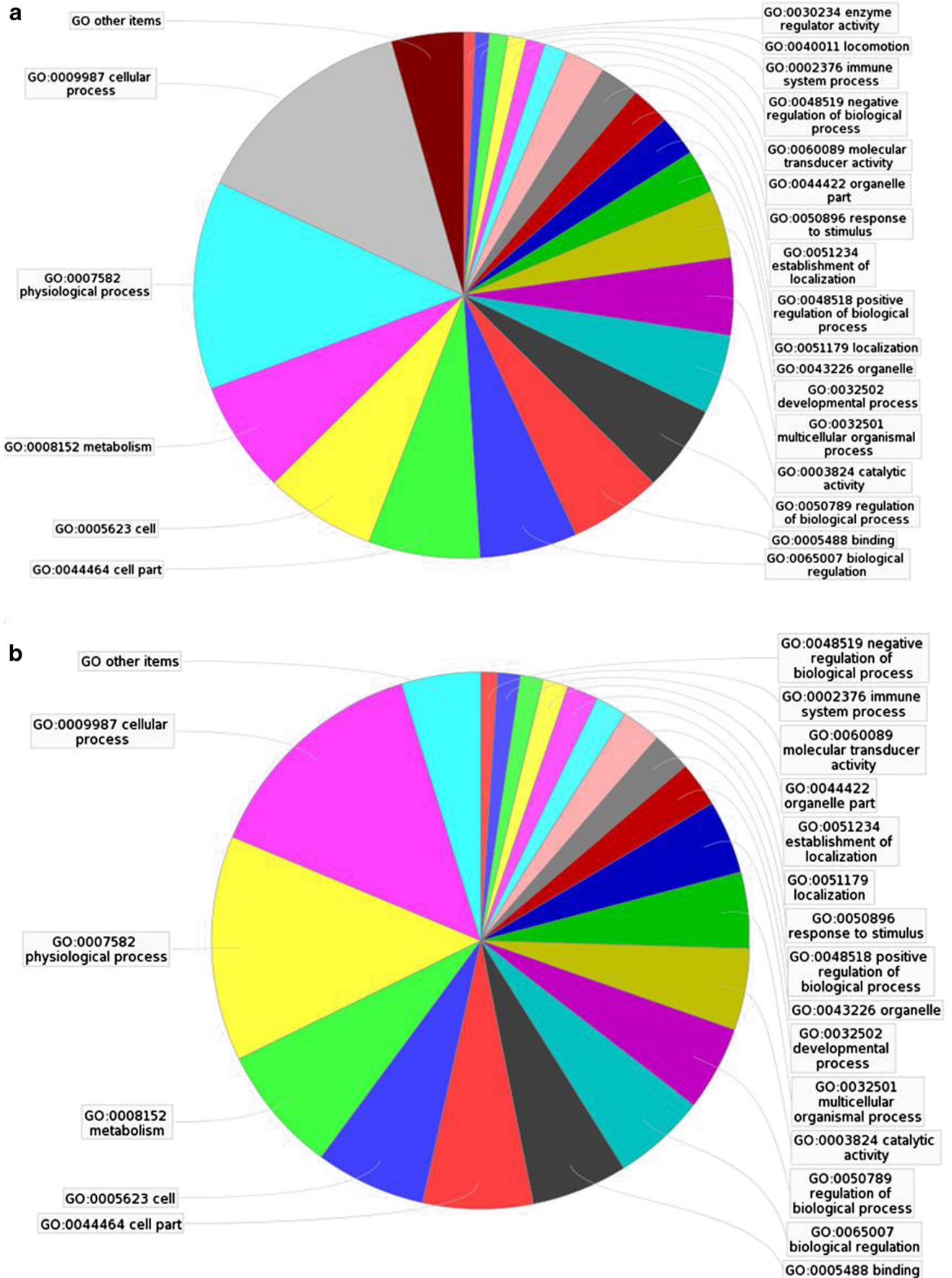


**a**



**b**





◀ **Fig. 6** Gene ontology (GO) enrichment analysis of the mRNAs which was co-expressed with the upregulated (a) and down regulated (b) LncRNAs from the comparison between the patient groups T1 and T2. The analysis was performed by the Molecular Annotation System (CBMAS) V3.0 (<http://bioinfo.capitalbio.com/mas3/>)

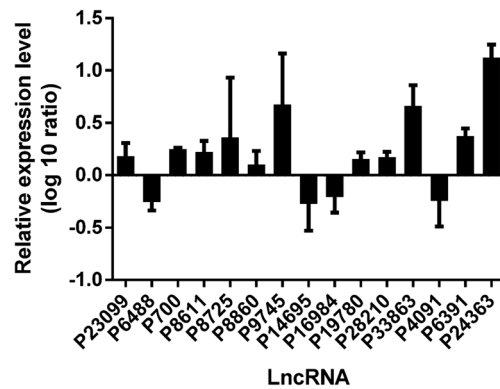
hundred twenty-one downregulated LncRNAs and 76 mRNAs were included in the downregulated CNC network to compose the network node. These 197 network nodes made 394 associated network pairs of co-expression LncRNAs and mRNAs (Fig. 5b). In both of the CNC network, most of the pairs were presented as positive correlation.

### LncRNAs-associated protein-coding genes are more likely to function in cellular and physiological processes

As mentioned previously, LncRNAs can regulate the expression of their overlapping or adjacent protein-coding genes in multiple mechanisms and are often transcribed together with their associated protein-coding genes. The function of the LncRNAs may be reflected by the function of their associated protein-coding genes in some extent. Therefore, through gene ontology (GO) term enrichment analysis using the protein-coding genes which is associated with differentially expressed LncRNAs, we can obtain some hints about the function of the LncRNAs. According to CNC network, we submitted a list of 110 protein-coding genes associated with upregulated LncRNAs and 75 protein-coding genes associated with downregulated LncRNAs in P3 group to MAS 3.0 system (<http://bioinfo.capitalbio.com/mas3/>). The most relevant GO terms were enriched in cellular process and physiological process (Fig. 6a, b). It is indicated that LncRNAs were likely to perform their function in HCC by regulating cellular process- and physiological process-related genes.

### Validation of the candidate LncRNAs by qPCR

To study the role of LncRNA in the recurrence/metastasis of HCC, we performed qPCR analysis in the cell lines of MHCC97H and MHCC97L which have exactly the same genetic background but have different abilities of invasion. We randomly selected 15 out of the significantly differentially expressed LncRNAs in P3 and validated their expression levels in cell lines (MHCC97H and MHCC97L). In those LncRNAs, nine of which were upregulated (P23099, P8860, P14695, P28210, P4091, P6391, P24363, P8725, and P9745) and six of which were downregulated (P6488, P700, P8611, P16984, P19780, and P33863). As shown in Fig. 7, compared with cell line MHCC97L (low metastatic cell line), 11 LncRNAs were upregulated and 4 LncRNAs were downregulated in the cell line MHCC97H. However, only eight of the selected LncRNAs are consistent with the microarray data.



**Fig. 7** The validation of 15 differentially expressed LncRNAs in the cell line MHCC97H and MHCC97L by qPCR analysis. The relative expression levels of LncRNAs in MHCC97H were calculated by dividing the expression of corresponding LncRNAs in MHCC97L, then transformed to log 10. The relative expression level, which is higher than 0, indicates upregulated in MHCC97H cell line; while the relative expression level, which is less than 0, indicates downregulated in MHCC97H cell line

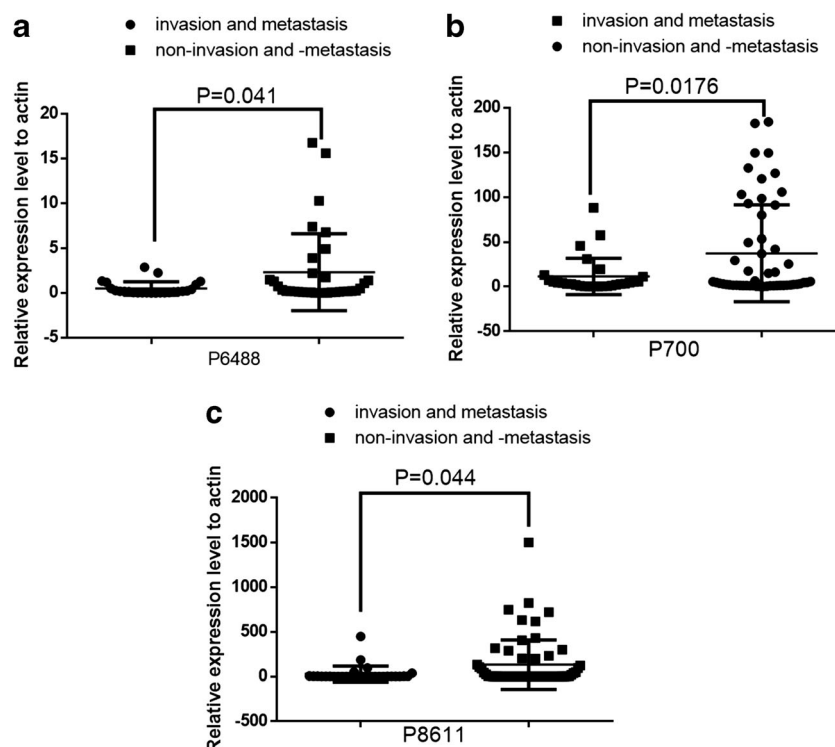
The expression levels of the LncRNA P24363 (probe number) in MHCC97H was almost 11 than that of P24363 in MHCC97L, respectively. P24363, which was upregulated in MHCC97H cell line and T1, is 372 bp in length with four exons and spanning from 121554101 to 121642550 on Homo sapiens chromosome 9 (alternate assembly CHM1\_1.1) with toll-like receptor 4 isoform D (TLR-4) at the 3' side with a distance of 991,120 bp and with bone morphogenetic protein/retinoic acid-inducible neural-specific protein precursor (BRINP1) at the 5' side with a distance of 461,243 bp. The distance seems too long for P24363 to regulate the expression of TLR-4 or BRINP1 in the way of trans-regulation.

To further validate the prognostic potential of the identified LncRNAs, we analyzed the expression levels of LncRNAs (P6488, P700, P8611), which has been verified at cell line level, by qPCR in 59 clinical tumor samples (25 of these samples had out-of-liver invasion and metastasis, but 34 of these samples had no invasion and metastasis). Here, we find that LncRNAs P6488, P700, and P8611 were significantly downregulated in the samples with invasion and metastasis out of the liver (Fig. 8); these results are well consistent with the microarray data. Therefore, the LncRNA P6488, P700, and P8611 might be a potential interesting prognostic biomarker for the invasion and metastasis of HCC.

### Discussion

In recent years, increasing evidences have demonstrated that the LncRNAs played important roles in complicated diseases such as cancer [29]. In solid cancers, the dysregulation of LncRNAs might be a useful information to predict the progression of diseases [30] and prognosis [31]. Since the discovering of the first LncRNA HULC in HCC, which was

**Fig. 8** The expression levels of LncRNAs (P6488, P700, P8611) in 59 clinical tumor samples. The clinical samples were divided into two groups: samples had out-of-liver invasion and metastasis (invasion and metastasis group),  $n=25$ ; samples without invasion and metastasis (none invasion and metastasis group),  $n=34$ . The expression levels of LncRNA were analyzed by qPCR



upregulated in HCC and knockdown of which could alter the expression of 5 genes [11], several LncRNAs have been identified to involve in the development and progression of HCC, such as MALT1 [32], TUC388 [33], Dreh [34], LET [35], and H19 [36]. Here, we have reported the overall LncRNAs expression profiles of primary HCC tissues with different recurrence/metastasis potentials, as well as the normal liver tissues; meanwhile, we also systematically analyzed the characteristic of LncRNA profiles associated with the recurrence/metastasis behavior of HCC by comparing the differentially expressed LncRNAs between groups T1 and T2. Such information will facilitate other researchers to explore the function of LncRNAs in HCC development, invasion, and metastasis. Similar works have been reported by Zhu et al. [37] but with different focus; in their analysis, they applied three pairs of HCC tumor tissues and adjacent nontumor tissues to uncover the LncRNAs that involved in the oncogenesis of HCC; but here, we focus more on the LncRNAs that are involved in the invasion and metastasis of HCC.

Compared with protein-coding genes, LncRNAs are generally expressed at low levels and are most likely to show tissue-specific expression pattern [38, 39]. To compare the overall expression profiles of LncRNAs and mRNAs in HCC invasion and metastasis, we employed microarrays containing both LncRNA and mRNA probes to detect the expression of LncRNAs and mRNAs simultaneously. Consisted with previous reports, we found that about 70 % LncRNAs were expressed above background in HCC tissues, while more than 80 % mRNAs were expressed above the background

(Table 2). In our study, there are more differentially expressed mRNAs than LncRNAs.

Hereditary susceptibility is one of the main etiologic factors of HCC. Recent studies have reported that there are susceptible genes to HCC in some chromosomes, such as Maelstrom located on chromosome 1 could promote the HCC metastasis via Akt/GSK-3 $\beta$ /snail signaling pathway [40] and STAT4 located on chromosome 2 is a risk factor for HBV-related HCC [41]. Additionally, Jiang et al. identified a susceptibility locus at Xq22.1 for HBV-related HCC [42]. Here, we found that the significantly differentially expressed LncRNAs were not equally distributed on all chromosomes (Fig. 2). Compared with the other chromosomes, chromosomes 1 and 2 had higher percentage of dysregulated LncRNAs and protein-coding mRNAs. Therefore, the chromosomes 1 and 2 might exist susceptible LncRNAs to the tumorigenesis, invasion, and metastasis of HCC.

In this study, we identified 1780 and 2720 significantly differentially expressed LncRNAs by comparing the T1 and T2 groups with NT group, respectively, while only 1170 differentially expressed LncRNAs by comparing the expression between T1 and T2 (Table 3). Since the function of LncRNAs cannot be inferred from the sequences or structures as we do for protein-coding genes or miRNAs, the functions of LncRNAs have to be predicted via genomic association analysis with protein-coding genes for the reason that LncRNAs often regulate the expression of their neighborhood or overlapped protein-coding genes [43]. Based on the genomic relationship between the LncRNAs and the protein-coding genes,

we have classified the LncRNAs as sense, antisense, intronic, bidirectional, and intergenic [9] and analyzed the genomic context of differentially expressed LncRNAs. We found that all of the five categories could be examined in the differentially expressed LncRNAs (Table 5 and Fig. 4), and this is well consistent with the report by Zhu et al. [37]. Strikingly, we found that nearly one third of them were long intergenic non-coding RNAs (LincRNAs) in all comparison. It indicated that LincRNAs were more abundant than other classes of LncRNAs in HCC, and they might be the major functional LncRNAs during HCC progression. To further define the function of LncRNAs in the tumorigenesis and development of HCC, we performed GO enrichment analysis by using protein-coding genes which are associated with differentially expressed LncRNAs. We found that the protein-coding genes that are associated with LncRNAs were inclined to be enriched in cellular process, physiological process, and metabolism (Fig. 6). The metastasis of tumor is resulting from a complicated systematic change of the cancer cells, including epithelial to mesenchymal transition (EMT) in which cancer cells become mesenchymal and motile, invades into the surrounding matrix and migrates away from the primary tumor site, seeding in the nearby tissue or secondary organs and overgrowth to form tumor again [44]. Changes in the cytoskeleton, signaling pathway, and energy metabolism are needed to complete the metastasis. It is not a surprise in our study that the LncRNAs-associated protein-coding genes were enriched in cellular process, physiological process, and metabolism.

Overall, we studied the characteristic of LncRNA expression profiles which are related to the recurrence/metastasis of HCC through using microarray assay and provided valuable information for further study the roles of LncRNAs in HCC, as well as the underlying molecular mechanisms.

**Acknowledgments** This work is supported by the Key Clinical Specialty Discipline Construction Program of Fujian, P. R. C.; the Key Project of National Science and Technology of China (Grant no. 2012ZX10002010-001-006 and Grant no. 2012ZX10002016-013), the National Natural Science Foundation of China (Grant no. 31201008), the Key Project of Fujian Province (Grant no. 2013YZ0002-3), the Scientific Foundation of Fuzhou Health Department (Grant no. 2013-S-wq18, and Grant no. 2013-S-wp1), the Mengchao Hepatobiliary Hospital of Fujian Medical University (Grant no. QDZJ-2014-004), the Science and Technology Bureau of Fuzhou City (Grant no. 2014-S-139-1), the Fujian Provincial Health and Family Planning Commission (Grant no. 2014-2-41), and the Research Development Special Fund of Indirectly Affiliated Hospital of Fujian Medical University (Grant no. FZS13004Y).

## References

- Jemal A et al. Global cancer statistics. *CA: Cancer J Clin.* 2011;2: 69–90.
- Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA: Cancer J Clin.* 2012;1:10–29.
- Allemani C et al. Global surveillance of cancer survival 1995–2009: analysis of individual data for 25 676 887 patients from 279 population-based registries in 67 countries (CONCORD-2). *Lancet.* 2014. doi:10.1016/S0140-6736(14)62038-9.
- Rahbari NN, Mehrabi A, Mollberg NM, Müller SA, Koch M, Büchler MW, et al. Hepatocellular carcinoma: current management and perspectives for the future. *Ann Surg.* 2011;3:453–69.
- Ng KM, Yan T, Black D, Chu FC, Morris DL. Prognostic determinants for survival after resection/ablation of a large hepatocellular carcinoma. *HPB (Oxford).* 2009;4:311–20.
- Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. *Mol Cell.* 2011;6:904–14.
- Huang J-l, Zheng L, Hu Y-W. Characteristics of long noncoding RNA and its relation to hepatocellular carcinoma. *Carcinogenesis.* 2014;3:507–14.
- Ernst C, Morton CC. Identification and function of long non-coding RNA. *Front Cell Neurosci.* 2013. doi:10.3389/fncel.2013.00168.
- Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell.* 2009;4:629–41.
- Quagliata L et al. LncRNA HOTTIP/HOXA13 expression is associated with disease progression and predicts outcome in hepatocellular carcinoma patients. *Hepatology.* 2013;3:911–23.
- Panzitt K et al. Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as noncoding RNA. *Gastroenterology.* 2007;1:330–42.
- Wang J et al. CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. *Nucleic Acids Res.* 2010;16:5366–83.
- Yang F et al. Long noncoding RNA high expression in hepatocellular carcinoma facilitates tumor growth through enhancer of zeste homolog 2 in humans. *Hepatology.* 2011;5:1679–89.
- Yuan J-H et al. A long noncoding RNA activated by TGF- $\beta$  promotes the invasion-metastasis cascade in hepatocellular carcinoma. *Cancer Cell.* 2014;5:666–81.
- Wang F, Yuan J-H, Wang S-B, Yang F, Yuan S-X, Ye C, et al. Oncofetal long noncoding RNA PVT1 promotes proliferation and stem cell-like property of hepatocellular carcinoma cells by stabilizing NOP2. *Hepatology.* 2014;4:1278–90.
- Yu W et al. Tumor suppressor long non-coding RNA, MT1DP is negatively regulated by YAP and Runx2 to inhibit FoxA1 in liver cancer cells. *Cell Signal.* 2014;12:2961–8.
- Ørom UA et al. Long noncoding RNAs with enhancer-like function in human cells. *Cell.* 2010;1:46–58.
- Barabási A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet.* 2004;2:101–13.
- Dinger ME et al. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* 2008;9:1433–45.
- Mercer TR et al. Long noncoding RNAs in neuronal-glia fate specification and oligodendrocyte lineage maturation. *BMC Neurosci.* 2010. doi:10.1186/1471-2202-11-14.
- Carninci P et al. The transcriptional landscape of the mammalian genome. *Science.* 2005;5740:1559–63.
- Wang X et al. Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature.* 2008;7200:126–30.
- Faghihi MA, Wahlestedt C. Regulatory roles of natural antisense transcripts. *Nat Rev Mol Cell Biol.* 2009;9:637–43.
- Louro R et al. Conserved tissue expression signatures of intronic noncoding RNAs transcribed from human and mouse loci. *Genomics.* 2008;1:18–25.
- Louro R, Smimova AS, Verjovski-Almeida S. Long intronic non-coding RNA transcription: expression noise or expression choice? *Genomics.* 2009;4:291–8.
- Trinklein ND et al. An abundance of bidirectional promoters in the human genome. *Genome Res.* 2004;1:62–6.

27. Morris KV et al. Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. *PLoS Genet.* 2008;11:e1000258.
28. Liu J et al. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis. *Plant Cell Online.* 2012;11:4333–45.
29. Gutschner T, Diederichs S. The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biol.* 2012;6:703–19.
30. Prensner JR et al. Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat Biotechnol.* 2011;8:742–9.
31. Gupta RA et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature.* 2010;7291:1071–6.
32. Gutschner T, Hämmerle M, Diederichs S. MALAT1—a paradigm for long noncoding RNA function in cancer. *J Mol Med.* 2013;7:791–801.
33. Braconi C et al. Expression and functional role of a transcribed noncoding RNA with an ultraconserved element in hepatocellular carcinoma. *Proc Natl Acad Sci.* 2011;2:786–91.
34. Huang JF et al. Hepatitis B virus X protein (HBx)-related long noncoding RNA (LncRNA) down-regulated expression by HBx (Dreh) inhibits hepatocellular carcinoma metastasis by targeting the intermediate filament protein vimentin. *Hepatology.* 2013;5:1882–92.
35. Yang F et al. Repression of the long noncoding RNA-LET by histone deacetylase 3 contributes to hypoxia-mediated metastasis. *Mol Cell.* 2013;6:1083–96.
36. Tsang WP, Kwok T. Riboregulator H19 induction of MDR1-associated drug resistance in human hepatocellular carcinoma cells. *Oncogene.* 2007;33:4877–81.
37. Zhu J et al. The long noncoding RNA expression profile of hepatocellular carcinoma identified by microarray analysis. *PLoS One.* 2014;9(7):e101707.
38. Kutter C et al. Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLoS Genet.* 2012;7:e1002841.
39. Derrien T et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 2012;9:1775–89.
40. Liu L et al. Maelstrom promotes hepatocellular carcinoma metastasis by inducing epithelial-mesenchymal transition by way of Akt/GSK-3 $\beta$ /Snail signaling. *Hepatology.* 2013;2:531–43.
41. Jiang D-K et al. Genetic variants in STAT4 and HLA-DQ genes confer risk of hepatitis B virus-related hepatocellular carcinoma. *Nat Genet.* 2012;1:72–5.
42. Jiang J-H et al. An X-chromosomal association study identifies a susceptibility locus at Xq22. 1 for hepatitis B virus-related hepatocellular carcinoma. *Clin Res Hepatol Gastroenterol.* 2013;6:586–95.
43. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet.* 2009;3:155–9.
44. Sulzmaier FJ, Ramos JW. RSK isoforms in cancer cell invasion and metastasis. *Cancer Res.* 2013;20:6099–105.