Table 1
Candidate transcription factors that regulate molecular pathways activated in HCC.

SAGE library	Transcription factor	Molecular processes	*P*-value
HCC–HCV	NF-κB	Antigen presentation Antigen processing Defense response Immune response	0.004
	SREBF1	Cholesterol biosynthesis Lipid biosynthesis β-Glucoside transport Negative regulation of lipoprotein metabolism	0.05
	SP1	Electron transport; drug metabolism Oxygen and reactive oxygen species metabolism Cell-substrate junction assembly; wound healing	0.05
	IRF1	Immune response Antigen presentation; antigen processing Defense response; positive regulation of cell	0.05
HCC–HBV	HNF4-α	Lipid transport Fatty acid metabolism Smooth muscle cell proliferation	0.002
	HNF1	Acute-phase response; lipid transport Negative regulation of lipid catabolism β-Glucoside transport Negative regulation of lipoprotein metabolism	0.01
	SP1	Zinc ion homeostasis; response to biotic stimulus Nitric oxide mediated signal transduction Copper ion homeostasis; fatty acid biosynthesis	0.01
	c-Jun	Progesterone catabolism; progesterone metabolism Regulation of lipid metabolism; Prostaglandin metabolism	0.03
	C/EBP-α	Lipid transport; negative regulation of lipid catabolism Negative regulation of lipoprotein metabolism β-Glucoside transport Positive regulation of interleukin-8 biosynthesis	0.03
	SREBF1	Lipid biosynthesis; fatty acid biosynthesis Fatty acid metabolism Negative regulation of lipid catabolism Negative regulation of lipoprotein metabolism	0.03
	c-Myc	Fatty acid biosynthesis; fatty acid metabolism Fatty acid desaturation; Activation of pro-apoptotic gene products Release of cytochrome c from mitochondria	0.03
	USF1	Fatty acid metabolism Smooth muscle cell proliferation	0.03
	PPAR-α	Fatty acid metabolism Smooth muscle cell proliferation	0.03
	COUP-TFI	Lipid transport Smooth muscle cell proliferation	0.03
	C/EBP-β	Acute-phase response Regulation of interleukin-6 biosynthesis Fat cell differentiation Inflammatory response	0.03

These findings were evaluated by other pathway analysis software, Ingenuity Pathways Analysis (IPA). We applied the signaling network analysis to the transcripts up-regulated in the HCC libraries ($P < 0.005$). We found that the top signaling network activated in HCC–HBV contained several pathways involved in ERK/MAPK signaling, PPAR signaling, linoleic acid metabolism, and fatty acid metabolism (Supplemental Fig. 2A). Similarly, pathways involved in interferon signaling, NF-κB signaling, antigen presentation, PPAR signaling, linoleic acid metabolism, and fatty acid metabolism were included in the top signaling network activated in HCC–HCV (Supplemental Fig. 2B). Consistent with the results of transcription factor analysis by MetaCore™, pathway analysis indicated that *SREBF1* participates in the lipogenesis pathway in both HCC–HBV and HCC–HCV (blue nodes in Supplemental Fig. 2A and B). *SREBF1*, a major regulator of the lipogenesis pathway, binds to sterol regulatory elements on the genome [30], but less is known about its role in

HCC [31]. We therefore focused on the role of *SREBF1* signaling in HCC.

3.3. Validation of SAGE and signaling network analysis

We performed real-time RT-PCR analysis of *SREBF1* and three representative target genes (*SCD*, *FADS1*, and *FASN*) [20] on 44 samples not used for SAGE. We found that the levels of *SREBF1*, *SCD*, and *FASN* mRNAs were higher in HCC tissues and CLD tissues compared with normal liver, and that these differences were statistically significant (Fig. 1A). We further compared the expression of *SREBF1*, *FADS1*, and *FASN* between HCC and non-cancerous liver tissues, and identified the overexpression of *SREBF1* in HCC with statistical significance (Supplemental Fig. 3). Scatter plot analysis showed that the expression levels of *SREBF1* were correlated with those of *FADS1* ($R = 0.57$, $P < 0.0001$), *SCD* ($R = 0.82$, $P < 0.0001$), and *FASN* ($R = 0.74$, $P < 0.0001$) (Fig. 1B).

Since the mammalian genome encodes two *SREBF1* isoforms, *SREBF1a* and *SREBF1c* [22], we performed semi-quantitative RT-PCR with isoform specific primers to determine which of these isoforms was up-regulated in HCC. We found that *SREBF1c* mRNA, but not *SREBF1a* mRNA, was up-regulated in HCC compared with adjacent non-cancerous liver and normal liver tissues (Supplemental Fig. 4A).

3.4. Functional assay of the lipogenesis pathway in cell lines

Although genome-wide expression profiling showed that the lipogenesis pathway was activated in HCC possibly through up-regulation of *SREBF1*, it was not clear that this pathway played a role in HCC growth. To investigate the role of lipogenesis in HCC cell proliferation, we transfected two short interfering (si)-RNAs (*SREBF1*-1 and *SREBF1*-2) targeting *SREBF1* into the HuH7 and Hep3B cells. These cell lines have no chromosome amplification or deletion on 17p11, on which *SREBF1* is located [32]. Transfection of the si-RNA constructs for *SREBF1*-1 or *SREBF1*-2 decreased expression of *SREBF1* 90% and 70%, respectively, and the expression of both *SCD* and *FADS1* 70% and 60%, respectively (Fig. 2A). Because differences in *SREBF1c* and *SREBF1a* sequence alignments are very small, we could not design si-RNAs specifically targeting *SREBF1c*. We therefore checked the effect of si-RNAs on the expression of the *SREBF1* isoforms. We found that the expression of *SREBF1c* was relatively more suppressed than that of *SREBF1a* (Supplemental Fig. 4B), which may have been associated with the higher expression of *SREBF1a* than *SREBF1c* in cultured cell lines [25].

We found that the growth of these transfected cells was significantly inhibited at 72 h compared with mock transfected cells (Fig. 2B and Supplemental Fig.5A). Examination of anchorage independent cell growth showed strong suppression by deactivation of the lipogenesis pathway (Fig. 2C). Because insulin-like growth factor (IGF) is known to induce cancer cell proliferation through activation of PI3-kinase signaling followed by *SREBF1* induction, we investigated the effect of *SREBF1* knockdown on IGF2 mediated cell proliferation. Interestingly, *SREBF1* knockdown abrogated the IGF2 dependent cell proliferation (Supplemental Fig. 5B). Moreover, both the TUNEL assay and annexin V staining showed that transfection of *SREBF1* si-RNAs increased apoptosis compared with mock transfected cells (Fig. 2D and E).

We further investigated the role of *SREBF1* overexpression on cell growth *in vitro*. We transiently transfected control pCMV7 plasmids or pCMV7-*SREBF1c* plasmids (Fig. 3A), and cell proliferation was enhanced in *SREBF1* overexpressing cells compared with the control in both HuH7 and Hep3B cells evaluated by focus assay (Fig. 3B and supplemental Fig. 6). Furthermore, overexpression of *SREBF1* intensified the phosphorylation of GSK-3β, one of the major kinase phosphorylated by the activation of IGF signaling, in a dose-dependent manner (Fig. 3C).

3.5. SREBF1 Expression and prognosis

Since the above results indicated that *SREBF1* signaling may play an important role on tumor cell growth, we investigated the relationship between *SREBF1* expression and mortality in 54 HCC patients by IHC. When we examined the expression of *SREBF1* in HCC tissues and adjacent non-cancerous liver tissues, we identified the increase of the cytoplasmic *SREBF1* staining in a subset of HCC (Fig. 4A). We evaluated the expression of *SREBF1* in HCC and classified 4, 30, and 20 HCCs as *SREBF1*-negative, *SREBF1*-low, and *SREBF1*-high HCC, respectively (Fig. 4B and Supplemental Fig. 1). We could not detect any differences of clinico-pathological characteristics between *SREBF1*-high HCC and *SREBF1*-low/-negative HCC including histological steatosis (Supplemental Table4). Since the seven of these HCC samples were also used for real-time RT-PCR analysis, we investigated the relation of *SREBF1* RNA and protein expression (Fig. 4C). *SREBF1*RNA expression was significantly higher in *SREBF1*-high HCC than in *SREBF1*-low/-negative HCC with statistical significance ($P = 0.03$). Then we examined the cell proliferation of these HCC samples by PCNA staining. Notably, PCNA indexes were significantly higher in *SREBF1*-high HCC than *SREBF1*-low/-negative HCC with statistical significance ($P < 0.001$) (Fig. 4D). We further investigated the relationship between *SREBF1*

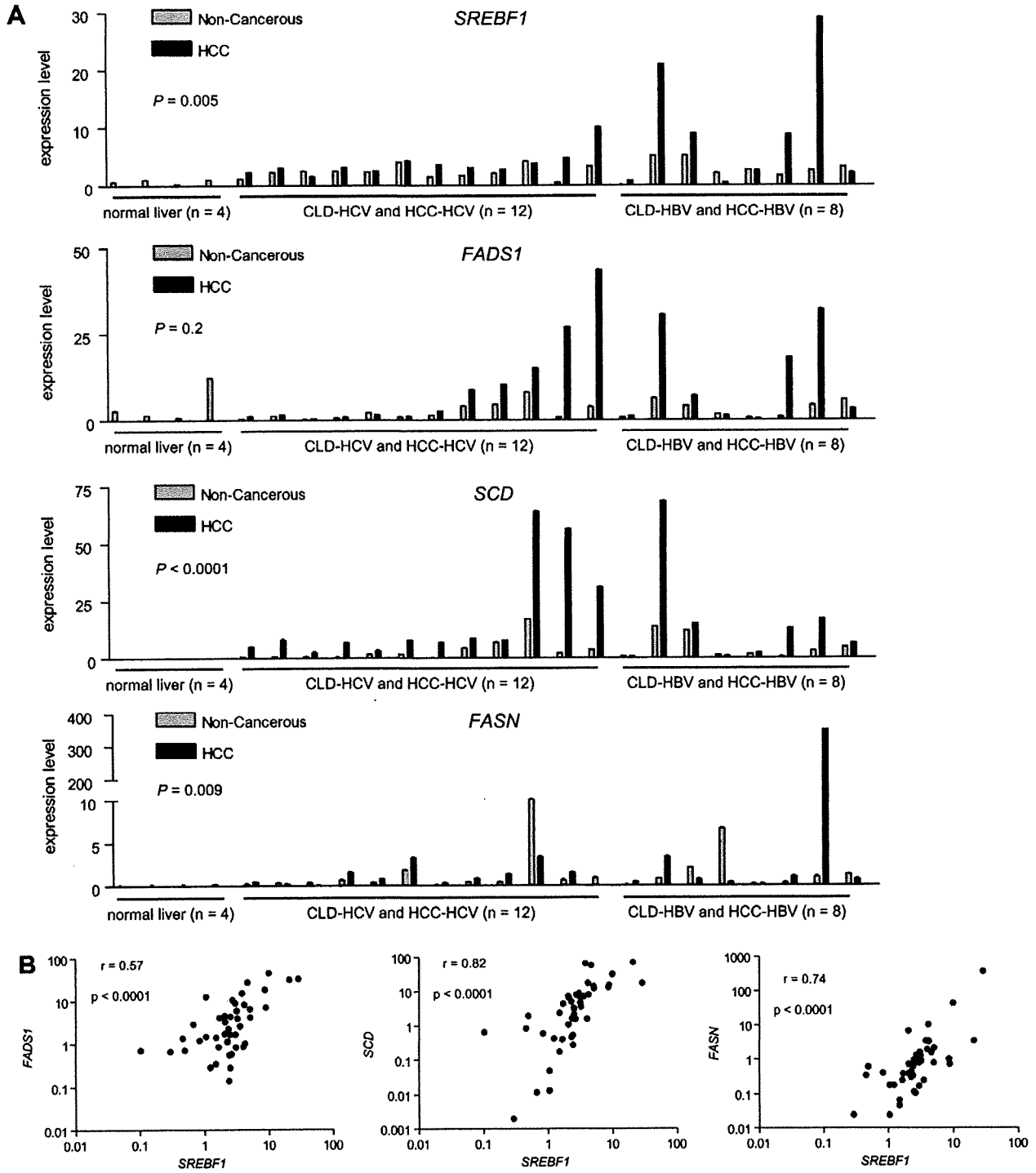

Fig. 1. (A) Real-time quantitative RT-PCR analysis. RNA was isolated from 44 tissue samples: 20 HCC, 20 corresponding CLD, and four normal liver samples. Differential expression of each gene among normal liver tissues, CLD tissues, and HCC tissues was examined by Kruskal–Wallis tests. (B) Scatter plot analysis. Gene expression levels of *FADS1*, *SCD* and *FASN* were well-correlated with those of *SREBF1*, as shown by Spearman's correlation coefficients.

protein expression and prognosis. Kaplan–Meier survival analysis showed a significant relationship between poor survival and high *SREBF1* protein expression ($P = 0.04$; Fig. 4E). Univariate Cox regression analysis showed a correlation between high *SREBF1* protein expression and high risk of mortality with statistical

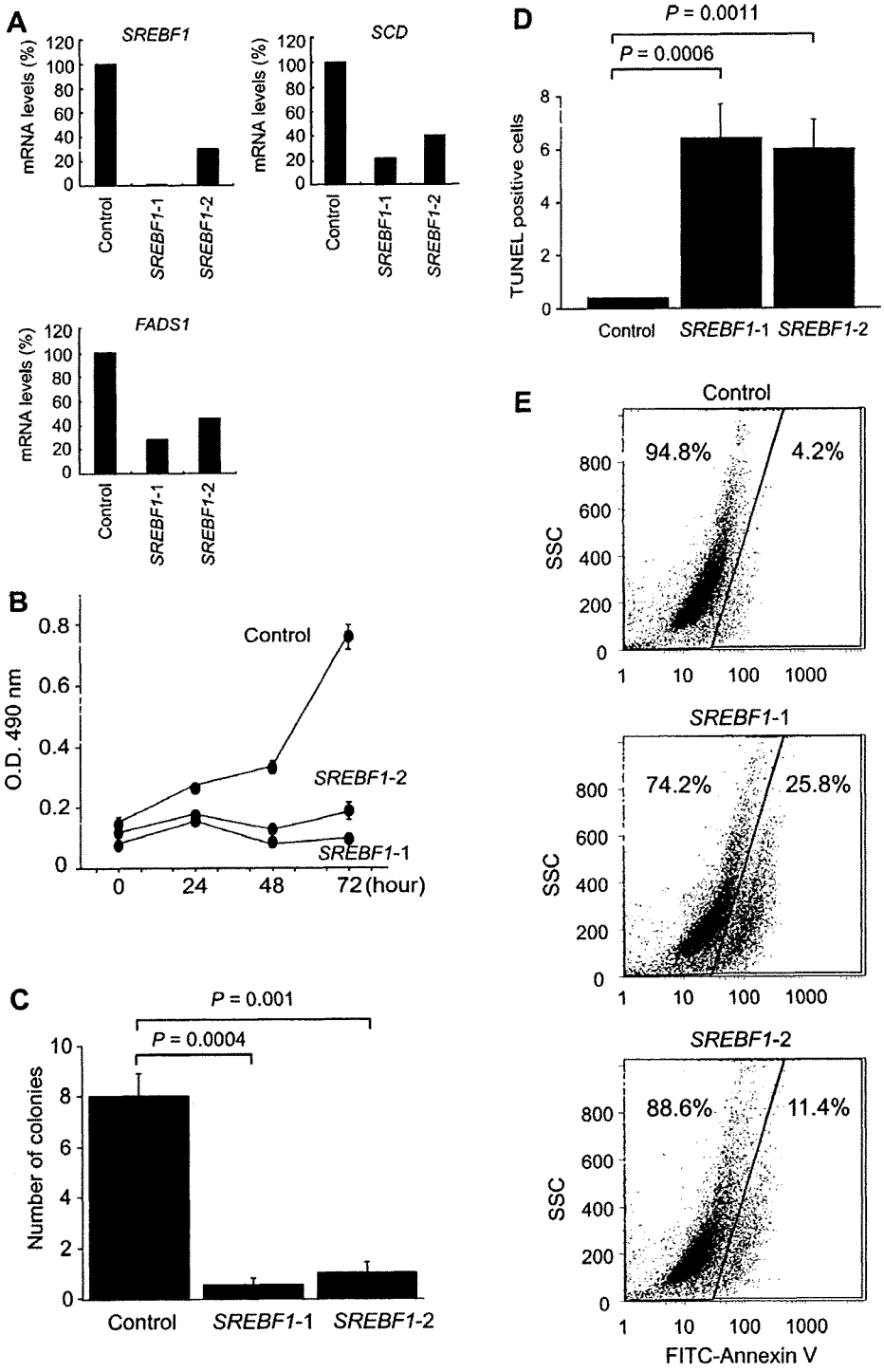

Fig. 2. (A) Effect of RNA interference targeting *SREBF1* in HuH7 cells. Expression levels of *SREBF1* mRNA were reduced by si-RNAs targeting different exons in *SREBF1*. Transcripts of *FADS1* and *SCD* were also down-regulated, showing transcriptional deactivation of the lipogenesis pathway. (B) Cell proliferation assay. Deactivation of the lipogenesis pathway severely reduced cell growth in HuH7 cells. (C) Soft agar assay. Deactivation of the lipogenesis pathway inhibited anchorage independent cell growth in HuH7 cells. (D) TUNEL assay. Deactivation of the lipogenesis pathway significantly increased the number of TUNEL-positive cells in HuH7 cells. (E) Annexin V staining evaluated by flow cytometer. Deactivation of the lipogenesis pathway significantly increased the number of annexin V positive cells in HuH7 cells.

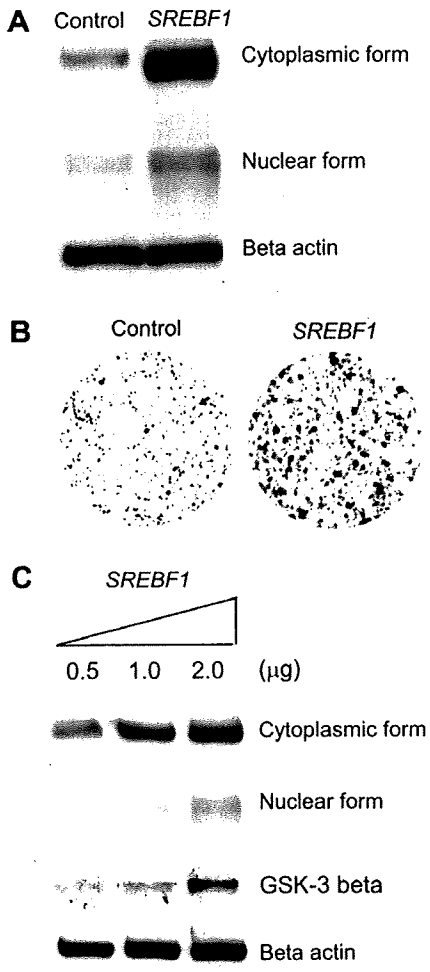

Fig. 3. (A) Western blot analysis of *SREBF1* protein expression in HuH7 cells transfected with control pCMV7 plasmids or pCMV7-*SREBF1c* plasmids. Both cytoplasmic and nuclear forms of *SREBF1* protein expression were increased by pCMV7-*SREBF1c* overexpression. (B) Focus assay of HuH7 cells transfected with control pCMV7 plasmids or pCMV7-*SREBF1c* plasmids. (C) Western blot analysis of *SREBF1* and phospho-GSK-3β protein expression in HuH7 cells transfected with indicated amounts of pCMV7-*SREBF1c* plasmids.

significance (HR, 3.7; 95% CI, 1.0–13.7; $P = 0.05$; Table 2).

4. Discussion

Using large-scale gene expression profiling, we have shown that the lipogenesis pathway is transcriptionally activated in HCC. Our SAGE profiles will be available on our homepage (http://www.intmedkanazawa.jp/) and will be submitted to the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/).

We found that the levels of expression of *FADS1*, *SCD*, and *FASN* were each correlated with those of *SREBF1*, suggesting that *SREBF1* is one of the main factors involved in the activation of lipogenesis in HCC. Activation of growth signaling pathways, such as the PI 3-kinase and mitogen-activated protein kinase pathways, has been shown to induce up-regulation of *SREBF1* in prostate and breast cancer cells [33,34]. We have observed induction of *SREBF1* protein expression by IGF2 in HuH7 cells (data not shown). Furthermore, we have identified that *SREBF1* overexpression results in the activation of cell proliferation and PI 3-kinase signaling, whereas expression inhibition of *SREBF1* abrogated the IGF2 induced cell proliferation. Although detailed mechanisms should be clarified in future, our results suggest that *SREBF1* is a key component of PI 3-kinase signaling in HCC.

SREBF1 is induced by alcohol [35], insulin, and fat [30,36], and plays a central role in the mechanism of hepatic steatosis [37]. Interestingly, these *SREBF1* inducers are risk factors for HCC [12,13,38,14]. Strikingly, two recent studies have shown that HBV and HCV infection may also induce hepatic steatosis through activation of *SREBF1* [39,40]. Furthermore, a recent report revealed the activation of *SREBF1* signaling in cancer by hypoxia [41]. Thus, these pathologic conditions such as chronic viral hepatitis, alcohol abuse, obesity, diabetes, and local hypoxia may up-regulate the expression of *SREBF1*, which, in turn, may contribute to an increased risk of hepatocarcinogenesis. Transgenic mice overexpressing *SREBF1* in the liver exhibited hepatic steatosis and hepatomegaly, suggesting the role of *SREBF1* on lipid metabolism and cell proliferation. However, it should be noted that no transgenic mice overexpressing *SREBF1* have been reported to have the risk of HCC development thus far. Interestingly, a recent report indicated that HCV core transgenic mice known to develop HCC showed coordinated activation of lipogenic pathway genes and *SREBF1* [42]. Although further studies are clearly required, we speculate that the activation of *SREBF1* may contribute to promote the development of HCC in already-initiated hepatocytes but not in normal hepatocytes.

Recently, Yahagi et al. reported the activation of lipogenic enzyme related genes in HCC [31]. In that paper, the authors suggested that *SREBF1* expression was not correlated with the expression of other lipogenic genes by Northern blotting, inconsistent with our current data. One possible explanation of these discrepancies might be the different methods for quantitation of mRNA, and we believe that real-time RT-PCR method used in our study would be more accurate. In addition, we evaluated the expression of *SREBF1* and lipogenic genes using more samples (a total of 44 liver and HCC tissues) than Yahagi et al did (10 HCC tissues). Furthermore, a recent paper indicated the coordinated activation of *SREBF1* and lipogenic genes in HCC

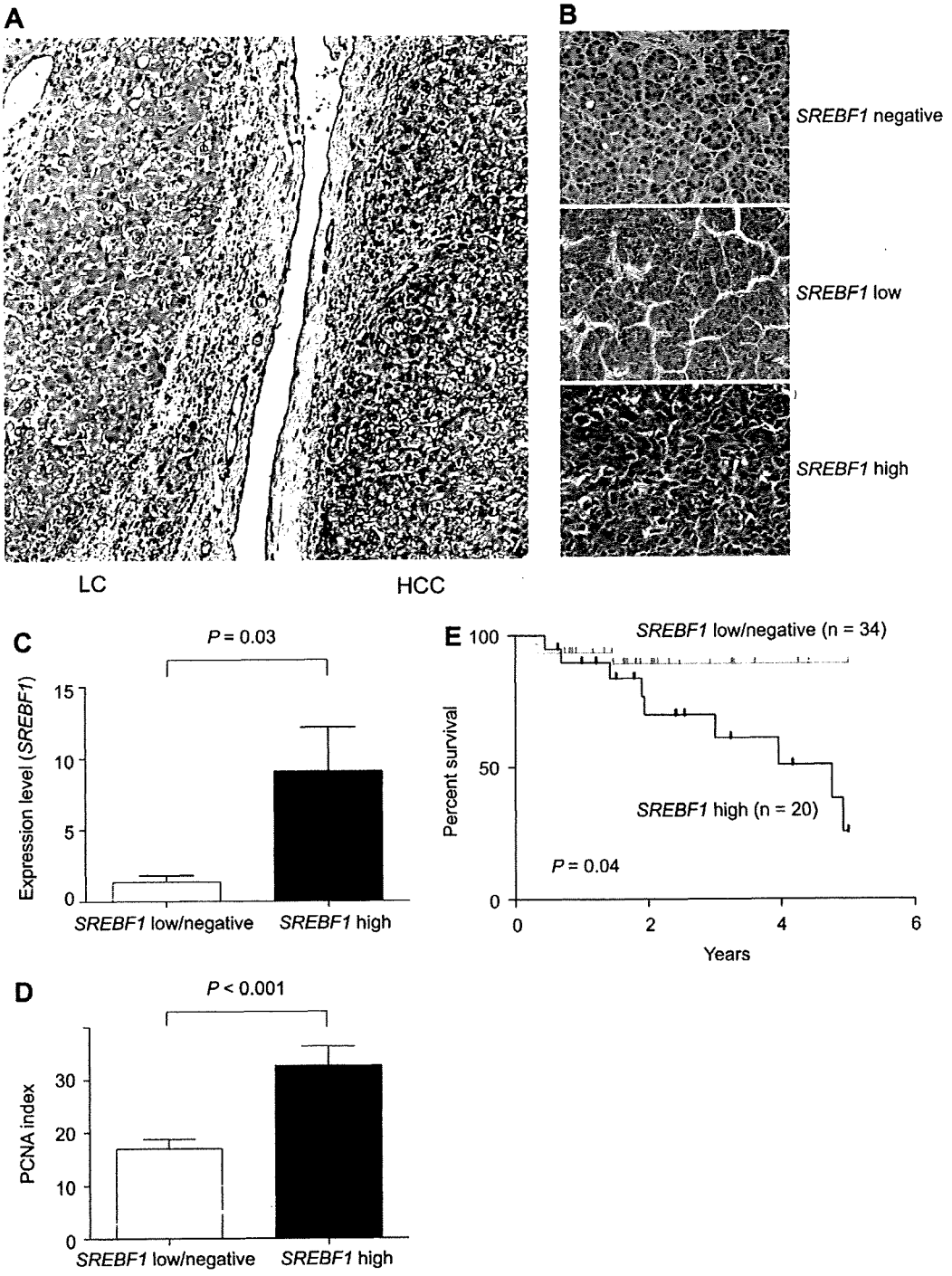

Fig. 4. (A) A photomicrograph of an HCC with adjacent non-cancerous cirrhotic liver stained with anti-*SREBF1* antibodies. (B) Representative photomicrographs of *SREBF1*-negative-, *SREBF1*-low-, and *SREBF1*-high-HCC tissues stained with anti- *SREBF1* antibodies. (C) *SREBF1* gene expression by real-time RT-PCR according to protein expression status assessed by IHC. *SREBF1* was highly expressed in *SREBF1*-high HCC ($P = 0.03$). (D) SREBF1 expression and cell proliferation in HCC. PCNA indexes in *SREBF1*-high HCC were higher than those in *SREBF1*-low/-negative HCC with statistical significance ($P < 0.001$). (E) Kaplan–Meier plots of 54 HCC patients analyzed by immunohistochemistry. The differences between *SREBF1*-high and -low/-negative HCC were analyzed by log-rank test.

developed in the liver of HCV core transgenic mice [42], strongly support our data. Although further studies using large numbers of HCC tissues may be required, these data suggest that the lipogenic gene activation seems to be mediated, at least in part, by *SREBF1* expression in HCC.

Table 2
Univariate Cox regression analysis of survival relative to *SREBF1* protein expression and clinicopathological parameters.

Variables (*n*)	HR (95% CI)	*P*-value
SREBF1 and mortality (n = 54)		
Tumor size		
<3 cm (*n* = 37)	1	
⩾3 cm (*n* = 17)	2.2 (0.6–8.3)	0.2
pTNM stage		
I, II (*n* = 45)	1	
III, IV (*n* = 9)	2.0 (0.4–9.4)	0.4
Serum AFP		
<20 ng/ml (*n* = 35)	1	
⩾20 ng/ml (n = 19)	1.5 (0.4–5.4)	0.5
SREBF1		
Low (*n* = 34)	1	
High (*n* = 20)	3.7 (1.0–13.7)	0.05

Because the majority of our HCC patients analyzed had Child–Pugh class A scores and about 70% had tumors less than 3 cm in diameter, all were expected to have a good prognosis. Indeed, patient survival in this cohort was not segregated by tumor size or pTNM stage (Table 2). Although the sample size was relatively small, we found that enhanced expression of *SREBF1* was a prognostic factor for mortality in HCC possibly due to the highly proliferative nature. Activation of lipogenesis pathways, as shown by overexpression of *FASN*, has been found to correlate with high mortality in breast, prostate, and lung cancer [43], suggesting that activation of lipogenesis may be a fundamental characteristic of cancer with poor prognosis. Thus, *SREBF1* expression may be a good biomarker for HCC classification, a finding that should be validated in a large scale cohort. Because deactivation of the lipogenesis pathway by inhibition of *SREBF1* gene expression could inhibit HCC cell growth *in vitro*, *SREBF1* may be a good target for pharmaceutical intervention in these tumors.

In conclusion, our genome-wide gene expression profiling analyses found that the lipogenesis pathway was activated in a subset of HCC. *SREBF1*, which activates the lipogenesis pathway, may be a good biomarker for HCC prognosis and may be a good target for therapeutic intervention.

Acknowledgements

We are grateful to the members of The Liver Disease Center at Kanazawa University Hospital for providing data of human liver tissue samples. We would also like to thank Dr. Hitoshi Shimano for providing invaluable reagents.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jhep.2008.07.036.

References

[1] El-Serag HB, Mason AC. Rising incidence of hepatocellular carcinoma in the United States. N Engl J Med 1999;340:745–750.
[2] Bosch FX, Ribes J, Diaz M, Cleries R. Primary liver cancer: worldwide incidence and trends. Gastroenterology 2004;127:S5–S16.
[3] Wang XW, Hussain SP, Huo TI, Wu CG, Forgues M, Hofseth LJ, et al. Molecular pathogenesis of human hepatocellular carcinoma. Toxicology 2002;181-182:43–47.
[4] Yamashita T, Kaneko S, Hashimoto S, Sato T, Nagai S, Toyoda N, et al. Serial analysis of gene expression in chronic hepatitis C and hepatocellular carcinoma. Biochem Biophys Res Commun 2001;282:647–654.
[5] Shirota Y, Kaneko S, Honda M, Kawai HF, Kobayashi K. Identification of differentially expressed genes in hepatocellular carcinoma with cDNA microarrays. Hepatology 2001;33:832–840.
[6] Okabe H, Satoh S, Kato T, Kitahara O, Yanagawa R, Yamaoka Y, et al. Genome-wide analysis of gene expression in human hepatocellular carcinomas using cDNA microarray: identification of genes involved in viral carcinogenesis and tumor progression. Cancer Res 2001;61:2129–2137.
[7] Xu XR, Huang J, Xu ZG, Qian BZ, Zhu ZD, Yan Q, et al. Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver. Proc Natl Acad Sci USA 2001;98:15089–15094.
[8] Iizuka N, Oka M, Yamada-Okabe H, Mori N, Tamesa T, Okada T, et al. Comparison of gene expression profiles between hepatitis B virus- and hepatitis C virus-infected hepatocellular carcinoma by oligonucleotide microarray data on the basis of a supervised learning method. Cancer Res 2002;62:3939–3944.
[9] Thorgeirsson SS, Grisham JW Molecular pathogenesis of human hepatocellular carcinoma. Nat Genet 2002;31:339–346.
[10] Lee JS, Thorgeirsson SS. Genome-scale profiling of gene expression in hepatocellular carcinoma: classification, survival prediction, and identification of therapeutic targets. Gastroenterology 2004;127:S51–S55.
[11] Suriawinata A, Xu R. An update on the molecular genetics of hepatocellular carcinoma. Semin Liver Dis 2004;24:77–88.
[12] El-Serag HB, Tran T, Everhart JE. Diabetes increases the risk of chronic liver disease and hepatocellular carcinoma. Gastroenterology 2004;126:460–468.
[13] Hassan MM, Hwang LY, Hatten CJ, Swaim M, Li D, Abbruzzese JL, et al. Risk factors for hepatocellular carcinoma: synergism of alcohol with viral hepatitis and diabetes mellitus. Hepatology 2002;36:1206–1213.
[14] Ohata K, Hamasaki K, Toriyama K, Matsumoto K, Saeki A, Yanagi K, et al. Hepatic steatosis is a risk factor for hepatocellular carcinoma in patients with chronic hepatitis C virus infection. Cancer 2003;97:3036–3043.
[15] Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of US adults. N Engl J Med 2003;348:1625–1638.
[16] Walsh MJ, Vanags DM, Clouston AD, Richardson MM, Purdie DM, Jonsson JR, et al. Steatosis and liver cell apoptosis in chronic hepatitis C: a mechanism for increased liver injury. Hepatology 2004;39:1230–1238.

[17] Powell EE, Jonsson JR, Clouston AD. Steatosis: co-factor in other liver diseases. Hepatology 2005;42:5–13.
[18] Velculescu VE, Zhang L, Vogelstein B, Kinzler KW Serial analysis of gene expression. Science 1995;270:484–487.
[19] Yamashita T, Hashimoto S, Kaneko S, Nagai S, Toyoda N, Suzuki T, et al. Comprehensive gene expression profile of a normal human liver. Biochem Biophys Res Commun 2000;269:110–116.
[20] Desmet VJ, Gerber M, Hoofnagle JH, Manns M, Scheuer PJ. Classification of chronic hepatitis: diagnosis, grading and staging. Hepatology 1994;19:1513–1520.
[21] Polyak K, Xia Y, Zweier JL, Kinzler KW, Vogelstein B. A model for p53-induced apoptosis. Nature 1997;389:300–305.
[22] Yokoyama C, Wang X, Briggs MR, Admon A, Wu J, Hua X, et al. SREBP-1, a basic-helix–loop-helix–leucine zipper protein that controls transcription of the low density lipoprotein receptor gene. Cell 1993;75:187–197.
[23] Wang HC, Chang WT, Chang WW, Wu HC, Huang W, Lei HY, et al. Hepatitis B virus pre-S2 mutant upregulates cyclin A expression and induces nodular proliferation of hepatocytes. Hepatology 2005;41:761–770.
[24] Takeba Y, Kumai T, Matsumoto N, Nakaya S, Tsuzuki Y, Yanagida Y, et al. Irinotecan activates p53 with its active metabolite, resulting in human hepatocellular carcinoma apoptosis. J Pharmacol Sci 2007;104:232–242.
[25] Closset J, Van de Stadt J, Delhaye M, El Nakadi I, Lambilliotte JP, Gelin M. Hepatocellular carcinoma: surgical treatment and prognostic variables in 56 patients. Hepatogastroenterology 1999;46:2914–2918.
[26] Arsura M, Cavin LG, Calvisi DF, Thorgeirsson SS, Eferl R, Ricci R, et al. Nuclear factor-kappaB and liver carcinogenesis. Cancer Lett 2005;229:157–169.
[27] Calvisi DF, Thorgeirsson SS. Molecular mechanisms of hepatocarcinogenesis in transgenic mouse models of liver cancer. Toxicol Pathol 2005;33:181–184.
[28] Eferl R, Ricci R, Kenner L, Zenz R, David JP, Rath M, et al. Liver tumor development. c-Jun antagonizes the proapoptotic activity of p53. Cell 2003;112:181–192.
[29] Xu L, Hui L, Wang S, Gong J, Jin Y, Wang Y, et al. Expression profiling suggested a regulatory role of liver-enriched transcription factors in human hepatocellular carcinoma. Cancer Res 2001;61:3176–3181.
[30] Horton JD, Goldstein JL, Brown MS. SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. J Clin Invest 2002;109:1125–1131.
[31] Yahagi N, Shimano H, Hasegawa K, Ohashi K, Matsuzaka T, Najima Y, et al. Co-ordinate activation of lipogenic enzymes in hepatocellular carcinoma. Eur J Cancer 2005;41:1316–1322.
[32] Kawaguchi K, Honda M, Yamashita T, Shirota Y, Kaneko S. Differential gene alteration among hepatoma cell lines demonstrated by cDNA microarray-based comparative genomic hybridization. Biochem Biophys Res Commun 2005;329:370–380.
[33] Van de Sande T, De Schrijver E, Heyns W, Verhoeven G, Swinnen JV Role of the phosphatidylinositol 3′-kinase/PTEN/Akt kinase pathway in the overexpression of fatty acid synthase in LNCaP prostate cancer cells. Cancer Res 2002;62:642–646.
[34] Yang YA, Han WF, Morin PJ, Chrest FJ, Pizer ES. Activation of fatty acid synthesis during neoplastic transformation: role of mitogen-activated protein kinase and phosphatidylinositol 3-kinase. Exp Cell Res 2002;279:80–90.
[35] You M, Fischer M, Deeg MA, Crabb DW Ethanol induces fatty acid synthesis pathways by activation of sterol regulatory element-binding protein (SREBP). J Biol Chem 2002;277:29342–29347.
[36] Muller-Wieland D, Kotzka J. SREBP-1: gene regulatory key to syndrome X? Ann NY Acad Sci 2002;967:19–27.
[37] Sekiya M, Yahagi N, Matsuzaka T, Najima Y, Nakakuki M, Nagai R, et al. Polyunsaturated fatty acids ameliorate hepatic steatosis in obese mice by SREBP-1 suppression. Hepatology 2003;38:1529–1539.
[38] Marrero JA, Fontana RJ, Su GL, Conjeevaram HS, Emick DM, Lok AS. NAFLD may be a common underlying liver disease in patients with hepatocellular carcinoma in the United States. Hepatology 2002;36:1349–1354.
[39] Kim KH, Shin HJ, Kim K, Choi HM, Rhee SH, Moon HB, et al. Hepatitis B virus X protein induces hepatic steatosis via transcriptional activation of SREBP1 and PPARgamma. Gastroenterology 2007;132:1955–1967
[40] Waris G, Felmlee DJ, Negro F, Siddiqui A. Hepatitis C virus induces proteolytic cleavage of sterol regulatory element binding proteins and stimulates their phosphorylation via oxidative stress. J Virol 2007;81:8122–8130.
[41] Furuta E, Pai SK, Zhan R, Bandyopadhyay S, Watabe M, Mo YY, et al. Fatty acid synthase gene is up-regulated by hypoxia via activation of Akt and sterol regulatory element binding protein-1. Cancer Res 2008;68:1003–1011.
[42] Tanaka N, Moriya K, Kiyosawa K, Koike K, Gonzalez FJ, Aoyama T. PPARalpha activation is essential for HCV core protein-induced hepatic steatosis and hepatocellular carcinoma in mice. J Clin Invest 2008;118:683–694.
[43] Kuhajda FP Fatty acid synthase and cancer: new application of an old pathway. Cancer Res 2006;66:5977–5980.

Genomics xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Comprehensive gene expression analysis of 5′-end of mRNA identified novel intronic transcripts associated with hepatocellular carcinoma

Yuji Hodo [a], Shin-ichi Hashimoto [b], Masao Honda [a], Taro Yamashita [a], Yutaka Suzuki [c], Sumio Sugano [c], Shuichi Kaneko [a,*], Kouji Matsushima [b]

[a] Department of Gastroenterology, Kanazawa University Graduate School of Medical Science, 13-1 Takara-Machi, Kanazawa, Ishikawa 920-8641, Japan
[b] Department of Molecular Preventive Medicine, School of Medicine, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[c] Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa, Chiba 277-8562, Japan

ARTICLE INFO

Article history:
Received 1 June 2009
Accepted 14 January 2010
Available online xxxx

Keywords:
5′-end serial analysis of gene expression
Transcriptional start site
Acyl-coenzyme A oxidase 2
Intron
Hepatocellular carcinoma

ABSTRACT

To elucidate the molecular feature of human hepatocellular carcinoma (HCC), we performed 5′-end serial analysis of gene expression (5′SAGE), which allows genome-wide identification of transcription start sites in addition to quantification of mRNA transcripts. Three 5′SAGE libraries were generated from normal human liver (NL), non-B, non-C HCC tumor (T), and background non-tumor tissues (NT). We obtained 226,834 tags from these libraries and mapped them to the genomic sequences of a total of 8,410 genes using RefSeq database. We identified several novel transcripts specifically expressed in HCC including those mapped to the intronic regions. Among them, we confirmed the transcripts initiated from the introns of a gene encoding acyl-coenzyme A oxidase 2 (*ACOX2*). The expression of these transcript variants were up-regulated in HCC and showed a different pattern compared with that of ordinary *ACOX2* mRNA. The present results indicate that the transcription initiation of a subset of genes may be distinctively altered in HCC, which may suggest the utility of intronic RNAs as surrogate tumor markers.

Introduction

Hepatocellular carcinoma (HCC) is the fifth most common cancer worldwide and the third most common cause of cancer mortality. HCC usually develops in patients with virus-induced (e.g., hepatitis B virus (HBV) and hepatitis C virus (HCV)) chronic inflammatory liver disease [1]; however, non-B, non-C HCC has been reported in patients negative for both HBV and HCV [2]. HCC development is a multistep process involving changes in host gene expression, some of which are correlated with the appearance and progression of a tumor. Multiple studies linking hepatitis viruses and chemical carcinogens with hepatocarcinogesis have provided insights into tumorigenesis [1,3]. Nevertheless, the genetic events that lead to HCC development remain unknown, and the molecular pathogenesis of HCC in most patients is still unclear. Therefore, elucidation of the genetic changes specific to the pathogenesis of non-B, non-C HCC may be useful to reveal the molecular features of HCCs irrelevant to viral infection.

Gene expression profiling, either by cDNA microarray [4] or serial analysis of gene expression (SAGE) [5], is a powerful molecular technique that allows analysis of the expression of thousands of genes. In particular, SAGE enables the rapid, quantitative, and simultaneous monitoring of the expression of tens of thousands of genes in various tissues [6,7]. Although numerous studies using cDNA microarrays and SAGE have been performed to clarify the genomic and molecular alterations associated with HCC [6,8–10], most expression data have been derived from the 3′-end region of mRNA. Recent advances in molecular biology have enabled genome-wide analysis of the 5′-end region of mRNA that revealed the variation in transcriptional start sites [11,12] and the presence of a large number of non-coding RNAs [13]. These approaches might be useful for identifying the unique and undefined genes associated with HCC not identified by the analysis of the 3′-end region of mRNA. SAGE based on the 5′-end (5′SAGE), a recently developed technique, allows for a comprehensive analysis of the transcriptional start site and quantitative gene expression [14]. This article is to elucidate the molecular carcinogenesis of non-B, non-C HCCs, while those heterogeneous entities are supposed not to share the same etiology, by using 5′SAGE.

Results

Annotation of the 5′SAGE tags to the human genome

We characterized a total of 226,834 tags from three unique 5′SAGE libraries (75,268 tags from the normal liver (NL) library, 75,573 tags from the non-tumor tissue (NT) library, and 75,993 tags from the tumor (T) library) and compared them against the human genome

Abbreviations: 5′SAGE, 5′-end serial analysis of gene expression; HCC, hepatocellular carcinoma; ACOX2, acyl-coenzyme A oxidase 2.

* Corresponding author. Fax: +81 76 234 4250
E-mail address: skaneko@m-kanazawa.jp (S. Kaneko).

0888-7543/$ – see front matter © 2010 Published by Elsevier Inc.
doi:10.1016/j.ygeno.2010.01.004

sequence. A total of 211,818 tags matched genomic sequences, representing 104,820 different tags in the three libraries (Table 1). About 60–65% of these tags mapped to a single locus in the genome in each library. Then, we mapped these single-matched tags to the well-annotated genes using RefSeq database (www.ncbi.nlm.nih.gov/RefSeq/, reference sequence database developed by NCBI). A total of 45,601 tags from the NL library, 39,858 from the NT library, and 41,265 from the T library were successfully mapped to 8410 unique genes (4397 genes detected in the NL library, 5194 genes in the NT library, and 6304 genes in the T library).

Gene expression profiling of non-B, non-C HCC

Abundantly expressed transcripts in the NL library and their corresponding expression in the NT and T libraries are shown in Table 2. The most abundant transcript in all three libraries was encoded by the *albumin* gene. Transcripts encoding apolipoproteins were also abundantly expressed in each library, suggesting the preservation of hepatocytic gene expression patterns in HCC. Of note, the expression of *haptoglobin* (NL: 631, NT: 329, T: 57) and metallothionein 1G (NL: 392, NT: 169, T: 2) was decreased in the NT library and more in T library compared with NL library. Furthermore, the expression of metallothionein 2A (NL: 1027, NT: 872, T: 19), *metallothionein 1X* (NL: 547, NT: 644, T: 11), and metallothionein 1E (NL: 275, NT: 340, T: 2) was decreased almost fifty-fold or more in the T library compared with the NL and NT libraries. In contrast, the expression of ribosomal protein S29 (NL: 372, NT: 1011, T: 1768) was increased in the NT library and more in T library compared with NL library. Thus, transcripts associated with a certain liver function including xenobiotic metabolism might be suppressed whereas those associated with protein synthesis might be expressed in non-B, non-C HCC, similar to that observed in HCV-HCC [15].

We then investigated the characteristics of gene expression patterns in non-B, non C HCC. Two hundred fifty-four and 172 genes were up- or down-regulated in the T library more than five-fold compared with the NL library (data not shown). The top 10 genes are listed in Table 3a, and we identified several novel genes not yet reported to be differentially expressed in non-B, non-C HCC. Representative novel gene expression changes identified by 5′SAGE were validated by semi-quantitative reverse transcriptase-polymerase chain reaction (RT-PCR) analysis (Supplemental Fig. 1). RT-PCR results showed that the expression of *galectin 4* (*LGALS4*), *X antigen family, member 1 (XAGE 1)*, *retinol dehydrogenase 11* (*RDH11*), *dehydrogenase/reductase member 10* (*DHRS10*), *transmembrane 14A* (*TMEM14A*), *stimulated by retinoic acid 13 homolog* (*STRA13*), and *dual specificity phosphatase 23* (*DUSP23*) was increased, whereas the expression of *C-type lectin superfamily 4 member G* (*CLEC4G*) was decreased in HCC tissues compared with the non-tumor tissues.

To further characterize the gene expression patterns of non-B, non-C HCC comprehensively, we compared the Gene Ontology process of three types of HCCs (i.e., non-B, non-C HCC; HBV-HCC; HCV-HCC) based on our previously described data [16]. The pathway analysis using MetaCore™ software showed that the immune related and cell adhesion related pathways were up-regulated in HCV-HCC with statistically significance, and the insulin signaling and angiogenesis related pathways were up-regulated in HBV-HCC with statistically significance, confirming our previous results [16]. Interestingly, genes associated with progesterone signaling were up-regulated in non-B, non-C HCC, while genes associated with proteolysis in the cell cycle, apoptosis and the ESR1-nuclear pathway were up-regulated in all types of HCC (Supplemental Fig. 2).

Dynamic alteration of transcription initiation in HCC

Although various transcriptome analyses have discovered considerable gene expression changes in cancer, it is still unclear if transcription is differentially initiated and/or terminated in HCC compared with the non-cancerous liver. We therefore explored the characteristics of transcription initiation and/or termination in HCC using 5′SAGE and 3′SAGE data. Markedly, we observed relevant differences between 5′SAGE and 3′SAGE data derived from the same HCC sample (Tables 3a and b). For example, a gene encoding *coagulation factor XIII, B polypeptide* was 13-fold up-regulated at transcription start sites (5′SAGE) but two-fold down-regulated at transcription termination sites (3′SAGE). On the other hand, a gene encoding *adenylate cyclase 1* was 50-fold down-regulated at transcriptional termination sites (3′SAGE) but showed no difference at transcriptional start sites (5′SAGE). These data suggest the dramatic alteration of all process of transcription in HCC, and the transcripts initiated at certain sites might be specifically associated with and involved in HCC pathogenesis, which could be a novel marker for HCC diagnosis.

Identification of novel intronic transcripts in HCC

Recent lines of evidence suggest that the majority of sequences of eukaryotic genomes may be transcribed, not only from known transcription start sites but also from intergenic regions and introns [17,18]. Introns are recognized as a significant source of functional non-coding RNAs (ncRNAs) including microRNAs (miRNAs) [18]. Moreover, a recent report implied the role of some large intronic RNAs in the pathogenesis of several types of malignancies [19]. Thus, analysis of transcripts originating from introns might be valuable for elucidating the genetic traits of HCC. We therefore focused on the transcriptional start sites potentially initiated from the intron and deregulated in HCC using 5′SAGE data. We identified that 97% of 5′ SAGE tags annotated by the RefSeq database matched the sequences in the exons, while 3% matched those in the introns (1257 in the N library, 1225 in the NT library, and 1261 in the T library) (Table 4a). To identify the possible promoter regions located in the intron, we clustered the different SAGE tags to a certain genomic region if these tags positioned within 500 bp intervals (Supplemental Fig. 3), as described previously [12].

Table 1
Experimental matching of 5′SAGE tags to genome.

	Normal liver	Non-tumor	Tumor	Total
All tags	75,268	75,573	75,993	226,834
Tags mapped to genome (%)				
1 locus/genome	51,076 (71.2)	47,200 (68.0)	48,503 (68.5)	146,779 (69.3)
Multiple loci/genome	20,608 (28.8)	22,142 (32.0)	22,289 (31.5)	65,039 (30.7)
Total tags	71,684 (100)	69,342 (100)	70,792 (100)	211,818 (100)
Unique tags mapped to genome (%)				
1 locus/genome	20,736 (65.5)	20,487 (60.2)	23,753 (60.7)	64,976 (62.0)
Multiple loci/genome	10,914 (34.5)	13,548 (39.8)	15,382 (39.3)	39,844 (38.0)
Total tags	31,650 (100)	34,035 (100)	39,135 (100)	104,820 (100)
Total tags to RefSeq	45,601	39,858	41,265	126,724
Unique gene	4397	5194	6304	8410

5′SAGE indicates 5′-end serial analysis of gene expression.

Please cite this article as: Y. Hodo, et al., Comprehensive gene expression analysis of 5′-end of mRNA identified novel intronic transcripts associated with hepatocellular carcinoma, Genomics (2010), doi:10.1016/j.ygeno.2010.01.004

Table 2
The highly expressed genes in the NL library and corresponding expression in the NT and T libraries (top 50 from NL library).

Tag count			Ratio		Gene
NL	NT	T	NT/NL	T/NL	
3731	1716	2328	0.460	0.624	Albumin
2484	2146	2042	0.864	0.822	Apolipoprotein C-I
1955	1603	1079	0.820	0.552	Apolipoprotein A-II
1653	1050	828	0.635	0.501	Apolipoprotein A-I
1252	1908	1203	1.524	0.961	Transthyretin (prealbumin, amyloidosis type I)
1233	724	220	0.587	0.178	Serpin peptidase inhibitor, clade A, member 1
1027	872	19	0.849	0.019	Metallothionein 2A
755	1144	762	1.515	1.009	Ferritin, light polypeptide
713	632	680	0.886	0.954	Alpha-1-microglobulin/bikunin precursor
635	524	1336	0.825	2.104	Apolipoprotein E
631	329	57	0.521	0.090	Haptoglobin
600	228	212	0.380	0.353	Fibrinogen gamma chain
549	395	302	0.719	0.550	Apolipoprotein C-III
547	644	11	1.177	0.020	Metallothionein 1X
479	257	290	0.537	0.605	Tumor protein, translationally-controlled 1
463	217	53	0.469	0.114	Serpin peptidase inhibitor, clade A, member 3
393	204	206	0.519	0.524	Ribosomal protein L26
392	169	2	0.431	0.005	Metallothionein 1G
372	1011	1768	2.718	4.753	Ribosomal protein S29
306	163	223	0.533	0.729	Ribosomal protein S27
279	135	159	0.484	0.570	Ribosomal protein S16
275	340	2	1.236	0.007	Metallothionein 1E
269	170	246	0.632	0.914	Ribosomal protein S23
260	142	92	0.546	0.354	Fibrinogen beta chain
260	200	195	0.769	0.750	Aldolase B, fructose-bisphosphate
255	228	286	0.894	1.122	Ribosomal protein S12
248	162	198	0.653	0.798	Ribosomal protein S14
246	175	70	0.711	0.285	Interferon induced transmembrane protein 3
239	198	273	0.828	1.142	Ribosomal protein L31
229	264	0	1.153	0.004	Hepcidin antimicrobial peptide
228	149	156	0.654	0.684	Ribosomal protein S20
222	191	117	0.860	0.527	Ubiquitin B
216	218	352	1.009	1.630	Ribosomal protein L41
210	150	155	0.714	0.738	Ribosomal protein, large, P1
201	110	90	0.547	0.448	Ribosomal protein, large, P2
198	102	64	0.515	0.323	Fibrinogen alpha chain
196	143	408	0.730	2.082	Ribosomal protein L37
192	123	56	0.641	0.292	Ribosomal protein L37a
191	208	346	1.089	1.812	Ribosomal protein L30
174	109	76	0.626	0.437	Ribosomal protein L35
169	208	3	1.231	0.018	Cytochrome P450, family 2, subfamily E, polypeptide 1
167	105	300	0.629	1.796	Apolipoprotein H (beta-2-glycoprotein I)
162	106	33	0.654	0.204	Serum amyloid A4, constitutive
159	85	157	0.535	0.987	Ribosomal protein L34 (RPL34)
159	113	229	0.711	1.440	Transferrin
155	84	135	0.542	0.871	Ribosomal protein S11
152	125	101	0.822	0.664	Ribosomal protein S13
147	84	1	0.571	0.007	Nicotinamide *N*-methyltransferase
147	180	35	1.224	0.238	Hemopexin
146	89	121	0.610	0.829	Alpha-2-HS-glycoprotein

To avoid division by 0, a tag value of 1 for any tag that was not detectable was used.
NL, normal liver; NT, non-tumor; T, tumor.

More than 2 tags were detected in the intronic regions of the 164 genes in the NL, 168 genes in the NT, and 157 genes in the T library, suggesting that these regions might be potential intronic promoter regions (Table 4a). The biological process of these intron-origin transcripts using Human Protein Reference Database (http://www.hprd.org/) showed that these were related to basic cellular functions such as signal transduction, transport, and regulation of the nucleobase and nucleotide, suggesting that these intronic transcripts may play a fundamental role in the liver (data not shown). Among these genes, 12 were differentially expressed between the NL and T libraries more than four-fold (Table 4b). Interestingly, intronic transcripts (determined by 5′SAGE) of genes encoding *SAMD3*, *ACOX2*, *HGD*, *CYP3A5*, *KNG1* and *AGXT* were increased, while their 3′ transcripts (determined by 3′SAGE) were decreased in HCC. In contrast, both 5′ intronic transcripts and 3′ transcripts encoding *HFM1*, *SERPINA1*, *SUPT3H*, *A2M* and *LR8* were similarly decreased in HCC. Taken together, these data imply that the canonical- and intronic-promoter activities of a subset of genes including *SAMD3*, *ACOX2*, *HGD*, *CYP3A5*, *KNG1* and *AGXT* might be differently regulated in HCC.

Table 3a
Differently expressed genes in HCC (top 10 from 5′SAGE).

5′SAGE	3′SAGE	5′/3′	Gene
T/NL	T/NL	Ratio	
Up-regulated gene			
19	6	3.17	P antigen family, member 2 (prostate associated)
18	10	1.8	Lectin, galactoside-binding, soluble, 4
16	3	5.33	Choline phosphotransferase 1
14	2	7	X antien family, member 1
14	2	7	Dehydrogenase/reductase (SDR family) member 4
14	2	7	Sterol-C5-desaturase-like
13	0.5	26	Coagulation factor XIII, B polypeptide
13	2.33	5.58	Retinol dehydrogenase 11 (all-trans and 9-cis)
13	0.5	26	Transmembrane protein 14A
12	1.33	9.02	Dual specificity phosphatase 23
Down-regulated gene			
0.00436	0.0137	0.318	Hepcidin antimicrobial peptide
0.0051	ND		Metallothionein 1G
0.0068	0.04	0.17	Nicotinamide *N*-methyltransferase
0.00727	ND		Metallothionein 1E (functional)
0.0098	0.0526	0.186	C-reactive protein, pentraxin-related
0.0145	ND		Metallothionein 1 M
0.0152	ND		Phospholipase A2, group IIA (platelets, synovial fluid)
0.0178	0.111	0.16	Cytochrome P450, family 2, subfamily E, polypeptide 1
0.0185	0.192	0.096	Metallothionein 2A
0.0201	ND		Metallothionein 1X

3′SAGE, 3′-end serial analysis of gene expression; 5′SAGE, 5′-end serial analysis of gene expression; HCC, hepatocellular carcinoma; NL, normal liver; T, tumor.

Table 3b
Differently expressed genes in HCC (top 10 from 3′SAGE).

5′SAGE	3′SAGE	5′/3′	Gene
T/NL	T/NL	Ratio	
Up-regulated gene			
ND	15		Leukocyte immunoglobulin-like receptor, subfamily B, member 1
ND	12		Fibroblast growth factor 5
1	11	0.909	Adenosine deaminase, tRNA-specific 1
5	11	0.454	px19-like protein
4.4	11	0.4	APC11 anaphase promoting complex subunit 11 homolog
ND	10.3		Chromosome 21 open reading frame 77
ND	10		von Willebrand factor
2.333	10	0.233	ATX1 antioxidant protein 1 homolog (yeast)
18	10	1.8	Lectin, galactoside-binding, soluble, 4
ND	9.5		Solute carrier family 26 (sulfate transporter), member 2
Down-regulated gene			
0.5	0.012	41.7	ELL associated factor 1
0.5	0.0137	36.5	TGF beta-inducible nuclear protein 1
0.000436	0.0137	0.032	Hepcidin antimicrobial peptide
1	0.0179	55.9	Basic, immunoglobulin-like variable motif containing
ND	0.0182		DNA fragmentation factor, 45 kDa, alpha polypeptide
1	0.0185	54.1	GRIP1 associated protein 1
ND	0.0189		Nuclear factor of activated T-cells 5, tonicity-responsive
1	0.0204	49	Adenylate cyclase 1
0.333	0.0312	10.7	Dihydroorotate dehydrogenase
0.738	0.0312	23.7	Ribosomal protein, large, P1

3′SAGE, 3′-end serial analysis of gene expression; 5′SAGE, 5′-end serial analysis of gene expression; HCC, hepatocellular carcinoma; NL, normal liver; T, tumor.

Please cite this article as: Y. Hodo, et al., Comprehensive gene expression analysis of 5′-end of mRNA identified novel intronic transcripts associated with hepatocellular carcinoma, Genomics (2010), doi:10.1016/j.ygeno.2010.01.004

Table 4a
Number of 5′SAGE tags mapped to intronic region.

	NL	NT	T
Tag mapped to intron	1287	1253	1292
Total promoter region	952	981	1020
(tag number = 1)	788	813	863
(tag number ≧2)	164	168	157

ACOX2 as a novel intronic gene deregulated in HCC

A subset of genes listed above may be transcribed from intronic regions specifically in HCC. Among these genes, we focused on the regulation of *ACOX2*, which is reported to be potentially involved in peroxisomal beta-oxidation and hepatocarcinogenesis [20]. The intron-origin expression of *ACOX2* increased six-fold in HCC compared with the NT by 5′SAGE, while the expression based on the 3′end was almost similar between HCC and NT lesions (Table 4b). Close examination of 5′SAGE data identified two potential intron-origin transcripts of *ACOX2* (Supplemental Fig. 4). The first (intronic-*ACOX2-1*) was initiated upstream of the tenth exon, whereas the second (intronic-*ACOX2-2*) was initiated upstream of the twelfth exon of *ACOX2* (Supplemental Fig. 4). The sequence of the intronic part was unique, and the remaining part of the sequence was shared with the canonical transcripts of *ACOX2*.

The expression of canonical *ACOX2* and the two types of intron-origin transcripts was investigated in NL, NT, and T tissues by RT-PCR (Fig. 1A). Although canonical *ACOX2* expression was decreased in T than in NL, the intron-origin transcript, particularly intronic-*ACOX2-1*, was increased in T. Intronic-*ACOX2-2* transcripts also showed a modest increase. We further evaluated the alteration of these transcripts in 19 HBV-HCCs, 20 HCV-HCCs, and 4 non-B, non-C HCCs, and their background liver tissues by canonical *ACOX2* and intronic-*ACOX2* specific real-time detection (RTD)-PCR. Although the expression of canonical *ACOX2* was decreased, the expression of intronic-*ACOX2* was significantly increased (Fig. 1B). Importantly, the gene expression ratios of intronic-to canonical *ACOX2* increased more in moderately differentiated HCCs (mHCC) than in well-differentiated HCCs (wHCC), suggesting the involvement of intronic-*ACOX2* expression on HCC progression.

Table 4b
Differentially expressed intronic promoter regions in HCC.

5′SAGE T/NL	3′SAGE T/NL	5′/3′ Ratio	Gene
Up-regulated			
9	1	9.00	Sterile alpha motif domain containing 3 (SAMD3)
6	0.89	6.74	Acyl-Coenzyme A oxidase 2, branched chain (ACOX2)
6	0.62	9.68	Homogentisate 1,2-dioxygenase (homogentisate oxidase) (HGD)
6	0.009	666.67	Cytochrome P450, family 3, subfamily A, polypeptide 5 (CYP3A5)
5	0.64	7.81	Kininogen 1 (KNG1)
4	0.36	11.11	Alanine-glyoxylate aminotransferase (AGXT)
4	1	4.00	Crystallin, alpha A (CRYAA)
Down-regulated			
0.13	1	0.13	HFM1, ATP-dependent DNA helicase homolog (*S. cerevisiae*) (HFM1)
0.25	0.51	0.49	Serpin peptidase inhibitor, clade A member 1 (SERPINA1)
0.25	1	0.25	Suppressor of Ty 3 Homolog (*S. cerevisiae*) (SUPT3H)
0.25	0.2	1.25	Alpha-2-macroglobulin (A2M)
0.25	0.083	3.13	LR8 protein (LR8)

3′SAGE, 3′-end serial analysis of gene expression; 5′SAGE, 5′-end serial analysis of gene expression; HCC, hepatocellular carcinoma; NL, normal liver; NT, non-tumor; T, tumor.

Discussion

This is the first comprehensive transcriptional analysis of tissue lesions of non-B, non-C HCC, background liver and NL using the 5′ SAGE method. Approximately 6.7% of our 5′SAGE tags showed no matching within the human genome, possibly due to the presence of a single nucleotide polymorphism (SNP) in the human genome. Out of the complete matched tags in the genome, 70% were assigned to unique positions and 30% to two or more loci. The tags with multiple matches with genomic loci were largely retrotransposon elements, repetitive sequences, and pseudogenes.

In this study, the analysis of non-B, non-C HCC enabled us to evaluate direct molecular changes associated with HCC without any bias of gene induction by virus infection. The gene expression profile based on our 5′SAGE tags revealed that albumin and apolipoproteins were highly expressed in NL, indicating the massive production of plasma proteins in NL; these results are similar to those of our previous study using 3′SAGE [6]. Other genes such as *aldolase B*, *antitrypsin*, and *haptoglobin* were also highly expressed in NL, in both the 5′SAGE and 3′SAGE libraries (Table 2) [6]. Comparison of the expression profiles among NL, background NT and T identified several differentially expressed transcripts in T. *Galectin-4* was up-regulated and *hepcidin*, *NNMT*, *CYP2E1*, and *metallothionein* were down-regulated in HCC in accordance with previous findings (Table 3a) [8,9,21]. Moreover, *CLEC4G*, which was predominantly expressed in the sinusoidal endothelial cells of the liver, was down-regulated in HCC. In addition, we first found that *P antigen family, member 2 (PAGE2)* and *XAGE1* were up-regulated in HCC (Table 3a, Supplemental Fig. 1). These genes were members of cancer-testis antigen that include MAGE-family genes. MAGE-family members were originally found to be up-regulated in HCV-related HCC, and reported to be useful as molecular markers and as possible target molecules for immunotherapy in human HCC [22]. In this study, we identified that these members of genes were also up-regulated in non B, non-C HCC. Thus, these genes may be useful as molecular markers and therapeutic targets for the treatment of a certain type of human HCC.

There existed some discrepancy between 5′SAGE and 3′SAGE results, even though they were derived from the same sample. Technical issues such as amplication error, difference of restriction enzyme, and annotation error have been described previously [14]. It is possible that 3′ transcripts might be more stable than 5′ transcripts by binding of ribosomal proteins during translation. Another possibility is the diversity of the transcriptional start and/or termination sites. One of the advantages of 5′SAGE analysis is the potential to determine the transcriptional start sites in each gene. Indeed, a recent study indicated the importance of an insulin splice variant in the pathogenesis of insulinomas [23]. Considering the diversity of 5′ends of genes, it is more appropriate to perform 5′SAGE in combination with 3′SAGE when determining the frequency of gene expression and identifying novel transcript variants.

Here, we were able to identify at least 12 intron-origin transcripts that were differentially expressed in HCC compared with the background liver or NL. These transcripts could not be identified by the 3′SAGE approach. We also performed detailed expression analysis of *ACOX2* that was involved in the beta-oxidation of peroxisome. We were able to clone the intron-origin *ACOX2* RNAs (intronic-*ACOX2-1*, *2*) for the first time and found that intronic-*ACOX2-1* was significantly overexpressed in T compared with NT and NL. The ratio of intronic-*ACOX2-1* and canonical *ACOX2* (relative intronic-*ACOX2*) was progressively up-regulated from NL via the background liver to HCC. Importantly, the expression of relative intronic-*ACOX2* was more up-regulated in moderately differentiated HCC than in well-differentiated HCC. The intronic difference in expression might be due to a polymorphism, since the 5′SAGE library for NL and T were from different people. The mechanisms of stepwise increase of intronic-*ACOX2* in the process of hepatocarinogenesis should be clarified in future.

Please cite this article as: Y. Hodo, et al., Comprehensive gene expression analysis of 5′-end of mRNA identified novel intronic transcripts associated with hepatocellular carcinoma, Genomics (2010), doi:10.1016/j.ygeno.2010.01.004

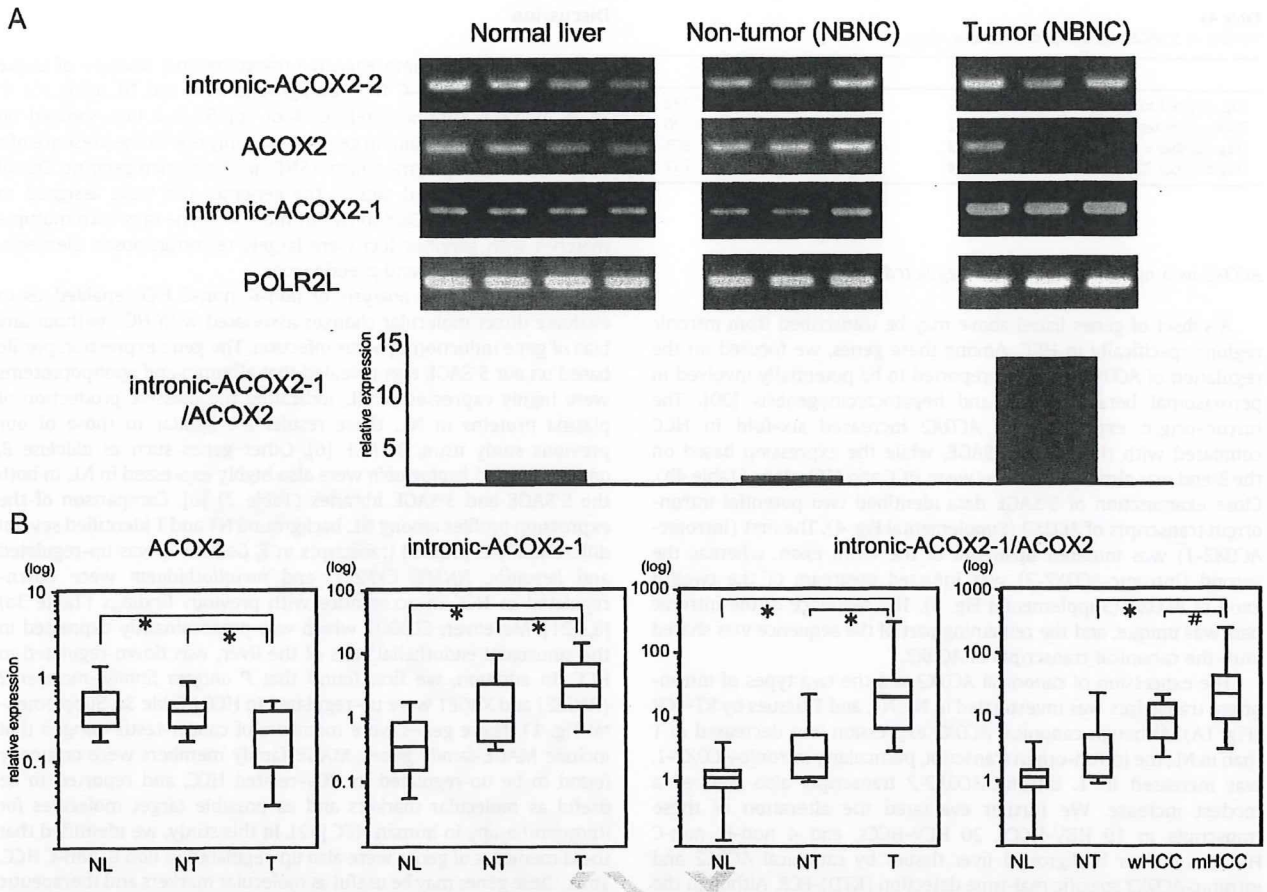

Fig. 1. (A) RT-PCR results of *ACOX2* and *ACOX2* intronic RNAs in independent NL, NT (non-B, non-C), and T (non-B, non-C) samples. RT-PCR was performed in triplicate for each sample-primer set from cDNA. The PCR products were semi-quantitatively analyzed with ImageJ software and calculated as levels relative to *polymerase (RNA) II (DNA directed) polypeptide L* (*POLR2L*). The bar graph indicates the expression ratio of intronic-ACOX2-1 to canonical ACOX2. The expression pattern of intron 1 was different from that of canonical *ACOX2*. (B) RTD-PCR analysis of *ACOX2* and *ACOX2* intronic RNAs in NL, T (HBV-related, HCV-related, and non-B, non-C), and NT tissues. Quantitative RTD-PCR was performed in duplicate for each sample-primer set from cDNA. Each sample was normalized relative to *POLR2L*. All HCC tissues were pathologically diagnosed as well differentiated HCC (wHCC) or moderately differentiated HCC (mHCC). Kruskal–Wallis tests and Mann–Whitney *U* tests were used for statistical analysis. ACOX2, acyl-Coenzyme A oxidase 2; HCC, hepatocellular carcinoma; NL, normal liver; NT, non-tumor; RT-PCR, reverse transcriptase-polymerase chain reaction; RTD-PCR, real-time detection-PCR; T, tumor. $*P<0.01$, $\#P<0.05$.

ACOX2 is a rate-limiting enzyme of branched-chain acyl-CoA oxidase involved in the degradation of long branched fatty acid and bile acid intermediates in peroxisomes. *ACOX2* expression was associated with the differentiation state of hepatocytes and was repressed under the undifferentiated phase of human hepatoma cell lines [24]. A decreased *ACOX2* expression was also reported in prostate cancer [25]. Here, the expression of canonical *ACOX2* was decreased, while that of intronic-*ACOX2-1* was increased in HCC. The deduced amino acid of intronic-*ACOX2-1* encodes the C-terminal (from 386 to 681 amino acids) of canonical *ACOX2*, lacking the active sites for FAD binding and a fatty acid as the substrate, suggesting that the protein may be functionally departed [26]. The biological role of the increased intronic-*ACOX2-1* was not clear, but it might be reflected by the activation of peroxisome proliferators-activated receptor (PPAR). It is reported that mice lacking ACOX1, another rate-limiting enzyme in peroxisomal straight-chain fatty acid oxidation, developed steatosis and HCC characterized by increased mRNA and protein expression of genes regulated by PPARα [27]. The importance of PPARα activation in HCC development has been recently reported using HCV core protein transgenic mice [28]. Moreover, the overexpression of alpha-methylacyl-CoA racemase, an enzyme for branched-chain fatty acid beta-oxidation, is reported to be a reliable diagnostic marker of prostate cancer and is associated with the decreased expression of ACOX2 [25]. Therefore, the expression of intronic-*ACOX2-1* might open the door for further investigations of their potential clinical use, e.g., serving as diagnostic markers of HCC, although the functional relevance of this gene should be further clarified.

In conclusion, we report the first comprehensive transcriptional analysis of non-B, non-C HCC, NT background liver, and NL tissue, based on 5′SAGE. This study offers new insights into the transcriptional changes that occur during HCC development as well as the molecular mechanism of carcinogenesis in the liver. The results suggest the presence of unique intron-origin RNAs that are useful as diagnostic markers and may be used as new therapeutic targets.

Material and methods

Samples

Samples were obtained from a 56-year-old man who had undergone surgical hepatic resection for the treatment of solitary HCC. Serological tests for hepatitis B surface (HBs) antigen and anti-HCV antibodies were negative. Tumor (T) and non-tumor (NT) tissue samples were separately obtained from the tumorous parts (diagnosed as moderately differentiated HCC) and non-tumorous parts

Please cite this article as: Y. Hodo, et al., Comprehensive gene expression analysis of 5′-end of mRNA identified novel intronic transcripts associated with hepatocellular carcinoma, Genomics (2010), doi:10.1016/j.ygeno.2010.01.004

(diagnosed as mild chronic hepatitis: F1A1) of the resected tissue. We also obtained five normal liver (NL) tissue samples from five patients who had undergone surgical hepatic resection because of metastatic liver cancer. None of the patients was seropositive for both HBs antigen and anti-HCV antibodies. Neither heavy alcohol consumption nor the intake of chemical agents was observed before surgical resection. All laboratory values related to hepatic function were within the normal range. All procedures and risks were explained verbally and provided in a written consent form.

We additionally used independent four NL tissue samples, 19 HBV-HCCs, 20 HCV-HCCs and 4 non-B, non-C HCCs, and their background liver tissue samples for reverse transcriptase-polymerase chain reaction (RT-PCR) and real-time detection (RTD)-PCR (Supplemental Table 1). Four non-B, non-C HCCs were histologically diagnosed as moderately differentiated HCCs, and the adjacent non-cancerous liver tissues were diagnosed as a normal liver, a chronic hepatitis, a pre-cirrhotic liver and a cryptogenic liver cirrhosis, respectively. None of the patients was seropositive for HBs antigen, anti-HBs antibodies, anti-hepatitis B core (HBc) antibodies and anti-HCV antibodies. Neither heavy alcohol consumption nor the intake of chemical agents was observed. Histological grading of the tumor was evaluated by two independent pathologists as described previously [16].

Generation of the 5′ SAGE library

5′SAGE libraries were generated as previously described [14]. Five to ten micrograms of poly(A)+RNA was treated with bacterial alkaline phosphatase (BAP; TaKaRa, Otsu, Japan). Poly(A)+RNA was extracted twice with phenol: chloroform (1:1), ethanol precipitated, and then treated with tobacco acid pyrophosphatase (TAP). Two to four micrograms of the BAP-TAP-treated poly(A)+RNA was divided into two aliquots and an RNA linker containing recognition sites for *Eco*RI/*Mme*I was ligated using RNA ligase (TaKaRa): one aliquot was ligated to a 5′-oligo 1 (5′-GGA UUU GCU GGU GCA GUA CAA CGA AUU CCG AC-3′) linker, and the other aliquot was ligated to a 5′-oligo 2 (5′-CUG CUC GAA UGC AAG CUU CUG AAU UCC GAC-3′) linker. After removing unligated 5′-oligo, cDNA was synthesized using RNaseH-free reverse-transcriptase (Superscript II, Invitrogen, Carlsbad, CA, USA) at 12 °C for 1 h and 42 °C for the next hour, using 10 pmol of dT adapter-primer (5′-GCG GCT GAA GAC GGC CTA TGT GGC CTT TTT TTT TTT TTT TTT-3′). After first-strand synthesis, RNA was degraded in 15 mM NaOH at 65 °C for 1 h. cDNA was amplified in a volume of 100 μl by PCR with 16 pmol of 5′ (5′ [biotin]-GGA TTT GCT GGT GCA GTA CAA-3′ or 5′[biotin]-CTG CTC GAA TGC AAG CTT CTG-3′) and 3′ (5′-GCG GCT GAA GAC GGC CTA TGT-3′) PCR primers. cDNA was amplified using 10 cycles at 94 °C for 1 min, 58 °C for 1 min, and 72 °C for 2 min. PCR products were digested with the *Mme*I type IIS restriction endonuclease (NEB, Pickering, Ontario, Canada). The digested 5′-terminal cDNA fragments were bound to streptavidin-coated magnetic beads (Dynal, Oslo, Norway). cDNA fragments that bound to the beads were directly ligated together in a reaction mixture containing T4 DNA ligase in a supplied buffer for 2.5 h at 16 °C. The ditags were amplified by PCR using the following primers: 5′ GGA TTT GCT GGT GCA GTA CA 3′ and 5′ CTG CTC GAA TGC AAG CTT CT 3′. The PCR products were analyzed by polyacrylamide gel electrophoresis (PAGE) and digested with *Eco*RI. The region of the gel containing the ditags was excised and the fragments were self-ligated to produce long concatamers that were then cloned into the *Eco*RI site of pZero 1.0 (Invitrogen). Colonies were screened by PCR using the M13 forward and reverse primers. PCR products containing inserts of more than 600 bp were sequenced with Big Dye terminator ver.3 and analyzed using a 3730 ABI automated DNA sequencer (Applied Biosystems, Foster City, CA, USA). All electrophoretograms were reanalyzed by visual inspection to check for ambiguous bases and to correct misreads. In this study, we obtained 19–20 bp tag information.

Association of the 5′SAGE tags with their corresponding genes

We attempted to align our 5′tags with the human genome (NCBI build 36, available from http://www.genome.ucsc.edu/) using the alignment program ALPS (http://www.alps.gi.k.u-tokyo.ac.jp/). Only tags that matched in sense orientation were considered in our analysis. The RefSeq database was searched for transcripts corresponding to the regions adjacent to the alignment location of each 5′tag.

RT-PCR

Total RNA was extracted using a ToTally RNA extraction kit (Ambion, Inc., Austin, TX, USA). Total RNA (500 ng) was reverse-transcribed in a 100-μl reaction solution containing 240 U of Moloney murine leukemia virus reverse transcriptase (Promega, Madison, WI, USA), 80 U of RNase inhibitor (Promega), 4.6 mM $MgCl_2$, 6.6 mM DTT, 1 mM dNTPs, and 2 mM random hexamer (Promega), at 42 °C for 1 h. PCR was performed in a 20-μl volume containing 0.5 U of AmpliTaq DNA polymerase (Applied Biosystems), 16.6 mM $(NH_4)_2SO_4$, 67 mM Tris–HCl, 6.7 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 1 mM dNTPs, and 1.5 μM sense and antisense primers, using an ABI 9600 thermal cycler (Applied Biosystems). The amplification protocol included 28–30 cycles of 95 °C for 45 s, 58 °C for 1 min, and 72 °C for 1 min. Primer sequences are shown in Supplemental Table 2. RT-PCR was performed in triplicate for each sample-primer set. Each sample was normalized relative to *polymerase (RNA) II (DNA directed) polypeptide L (POLR2L)*. *POLR2L* is a housekeeping gene that showed relatively stable gene expression in various tissues [29]. The PCR products were semi-quantitatively analyzed with ImageJ software (http://rsb.info.nih.gov/ij/).

RTD-PCR

Intron-origin transcript expression was quantified using TaqMan Universal Master Mix (Applied Biosystems). The samples were amplified using an ABI PRISM 7900HT Sequence Detection System (Applied Biosystems). Using the standard curve methods, quantitative PCR was performed in duplicate for each sample-primer set. Each sample was normalized relative to *POLR2L*. The assay IDs used were Hs00185873_m1 for *ACOX2* and Hs00360764_m1 for *POLR2L*. The specific primers and probe sequence of intronic-*ACOX2-1* were 5′-TTCATAAAGTTGTGAGCAGAGGAAA-3′ (forward), 5′-TGCACCACTTACT-GAGCATCTACTC-3′ (reverse), and 5′-ACTTCTTACCTCAGAGCTG-3′ (probe).

Analysis of pathway network

MetaCore™ software (GeneGo Inc., St. Joseph, MI) was used to investigate the molecular pathway networks of non-B, non-C HCC, HBV-HCC and HCV-HCC. All genes up-regulated more than five-fold in all HCC libraries subjected to Enrichment analysis in GO process networks by default settings ($p<0.05$).

Statistical analysis

Kruskal–Wallis tests were used to compare the expression among normal liver, non-cancerous tissues, and HCC tissues. Mann–Whitney *U* tests were also used to evaluate the statistical significance of ACOX2 gene expression levels between two groups. All statistical analyses were performed using R (http://www.r-project.org/).

Acknowledgments

The authors would like to thank Mr. Shungo Deshimaru and Ms. Keiko Harukawa for technical assistance.

Please cite this article as: Y. Hodo, et al., Comprehensive gene expression analysis of 5′-end of mRNA identified novel intronic transcripts associated with hepatocellular carcinoma, Genomics (2010), doi:10.1016/j.ygeno.2010.01.004

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2010.01.004.

References

[1] H.B. El-Serag, K.L. Rudolph, Hepatocellular carcinoma: epidemiology and molecular carcinogenesis, Gastroenterology 132 (2007) 2557–2576.

[2] Y. Yokoi, S. Suzuki, S. Baba, K. Inaba, H. Konno, S. Nakamura, Clinicopathological features of hepatocellular carcinomas (HCCs) arising in patients without chronic viral infection or alcohol abuse: a retrospective study of patients undergoing hepatic resection, J. Gastroenterol. 40 (2005) 274–282.

[3] R.N. Aravalli, C.J. Steer, E.N. Cressman, Molecular mechanisms of hepatocellular carcinoma, Hepatology 48 (2008) 2047–2063.

[4] D.J. Duggan, M. Bittner, Y. Chen, P. Meltzer, J.M. Trent, Expression profiling using cDNA microarrays, Nat. Genet. 21 (1999) 10–14.

[5] V.E. Velculescu, L. Zhang, B. Vogelstein, K.W. Kinzler, Serial analysis of gene expression, Science 270 (1995) 484–487.

[6] T. Yamashita, S. Hashimoto, S. Kaneko, S. Nagai, N. Toyoda, T. Suzuki, K. Kobayashi, K. Matsushima, Comprehensive gene expression profile of a normal human liver, Biochem. Biophys. Res. Commun. 269 (2000) 110–116.

[7] S. Hashimoto, S. Nagai, J. Sese, T. Suzuki, A. Obata, T. Sato, N. Toyoda, H.Y. Dong, M. Kurachi, T. Nagahata, K. Shizuno, S. Morishita, K. Matsushima, Gene expression profile in human leukocytes, Blood 101 (2003) 3509–3513.

[8] H. Okabe, S. Satoh, T. Kato, O. Kitahara, R. Yanagawa, Y. Yamaoka, T. Tsunoda, Y. Furukawa, Y. Nakamura, Genome-wide analysis of gene expression in human hepatocellular carcinomas using cDNA microarray: identification of genes involved in viral carcinogenesis and tumor progression, Cancer. Res. 61 (2001) 2129–2137.

[9] Y. Shirota, S. Kaneko, M. Honda, H.F. Kawai, K. Kobayashi, Identification of differentially expressed genes in hepatocellular carcinoma with cDNA microarrays, Hepatology 33 (2001) 832–840.

[10] T. Yamashita, M. Honda, S. Kaneko, Application of serial analysis of gene expression in cancer research, Curr. Pharm. Biotechnol. 9 (2008) 375–382.

[11] Y. Suzuki, H. Taira, T. Tsunoda, J. Mizushima-Sugano, J. Sese, H. Hata, T. Ota, T. Isogai, T. Tanaka, S. Morishita, K. Okubo, Y. Sakaki, Y. Nakamura, A. Suyama, S. Sugano, Diverse transcriptional initiation revealed by fine, large-scale mapping of mRNA start sites, EMBO Rep. 2 (2001) 388–393.

[12] K. Kimura, A. Wakamatsu, Y. Suzuki, T. Ota, T. Nishikawa, R. Yamashita, J. Yamamoto, M. Sekine, K. Tsuritani, H. Wakaguri, S. Ishii, T. Sugiyama, K. Saito, Y. Isono, R. Irie, N. Kushida, T. Yoneyama, R. Otsuka, K. Kanda, T. Yokoi, H. Kondo, M. Wagatsuma, K. Murakawa, S. Ishida, T. Ishibashi, A. Takahashi-Fujii, T. Tanase, K. Nagai, H. Kikuchi, K. Nakai, T. Isogai, S. Sugano, Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes, Genome. Res. 16 (2006) 55–65.

[13] T. Shiraki, S. Kondo, S. Katayama, K. Waki, T. Kasukawa, H. Kawaji, R. Kodzius, A. Watahiki, M. Nakamura, T. Arakawa, S. Fukuda, D. Sasaki, A. Podhajska, M. Harbers, J. Kawai, P. Carninci, Y. Hayashizaki, Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 15776–15781.

[14] S. Hashimoto, Y. Suzuki, Y. Kasai, K. Morohoshi, T. Yamada, J. Sese, S. Morishita, S. Sugano, K. Matsushima, 5′-end SAGE for the analysis of transcriptional start sites, Nat. Biotechnol. 22 (2004) 1146–1149.

[15] T. Yamashita, S. Kaneko, S. Hashimoto, T. Sato, S. Nagai, N. Toyoda, T. Suzuki, K. Kobayashi, K. Matsushima, Serial analysis of gene expression in chronic hepatitis C and hepatocellular carcinoma, Biochem. Biophys. Res. Commun. 282 (2001) 647–654.

[16] T. Yamashita, M. Honda, H. Takatori, R. Nishino, H. Minato, H. Takamura, T. Ohta, S. Kaneko, Activation of lipogenic pathway correlates with cell proliferation and poor prognosis in hepatocellular carcinoma, J. Hepatol. 50 (2009) 100–110.

[17] J.S. Mattick, Introns: evolution and function, Curr. Opin. Genet. Dev. 4 (1994) 823–831.

[18] J.S. Mattick, I.V. Makunin, Non-coding RNA, Hum. Mol. Genet. 15 (Spec No 1) (2006) R17–29.

[19] R. Louro, A.S. Smirnova, S. Verjovski-Almeida, Long intronic noncoding RNA transcription: expression noise or expression choice? Genomics 93 (2009) 291–298.

[20] S. Yu, S. Rao, J.K. Reddy, Peroxisome proliferator-activated receptors, fatty acid oxidation, steatohepatitis and hepatocarcinogenesis, Curr. Mol. Med. 3 (2003) 561–572.

[21] N. Kondoh, T. Wakatsuki, A. Ryo, A. Hada, T. Aihara, S. Horiuchi, N. Goseki, O. Matsubara, K. Takenaka, M. Shichita, K. Tanaka, M. Shuda, M. Yamamoto, Identification and characterization of genes associated with human hepatocellular carcinogenesis, Cancer. Res. 59 (1999) 4990–4996.

[22] Y. Kobayashi, T. Higashi, K. Nouso, H. Nakatsukasa, M. Ishizaki, T. Kaneyoshi, N. Toshikuni, K. Kariyama, E. Nakayama, T. Tsuji, Expression of MAGE, GAGE and BAGE genes in human liver diseases: utility as molecular markers for hepatocellular carcinoma, J. Hepatol. 32 (2000) 612–617.

[23] A.H. Minn, M. Kayton, D. Lorang, S.C. Hoffmann, D.M. Harlan, S.K. Libutti, A. Shalev, Insulinomas and expression of an insulin splice variant, Lancet 363 (2004) 363–367.

[24] H. Stier, H.D. Fahimi, P.P. Van Veldhoven, G.P. Mannaerts, A. Volkl, E. Baumgart, Maturation of peroxisomes in differentiating human hepatoblastoma cells (HepG2): possible involvement of the peroxisome proliferator-activated receptor alpha (PPAR alpha), Differentiation 64 (1998) 55–66.

[25] S. Zha, S. Ferdinandusse, J.L. Hicks, S. Denis, T.A. Dunn, R.J. Wanders, J. Luo, A.M. De Marzo, W.B. Isaacs, Peroxisomal branched chain fatty acid beta-oxidation pathway is upregulated in prostate cancer, Prostate 63 (2005) 316–323.

[26] K. Tokuoka, Y. Nakajima, K. Hirotsu, I. Miyahara, Y. Nishina, K. Shiga, H. Tamaoki, C. Setoyama, H. Tojo, R. Miura, Three-dimensional structure of rat-liver acyl-CoA oxidase in complex with a fatty acid: insights into substrate-recognition and reactivity toward molecular oxygen, J. Biochem. 139 (2006) 789–795.

[27] K. Meyer, Y. Jia, W.Q. Cao, P. Kashireddy, M.S. Rao, Expression of peroxisome proliferator-activated receptor alpha, and PPARalpha regulated genes in spontaneously developed hepatocellular carcinomas in fatty acyl-CoA oxidase null mice, Int. J. Oncol. 21 (2002) 1175–1180.

[28] N. Tanaka, K. Moriya, K. Kiyosawa, K. Koike, F.J. Gonzalez, T. Aoyama, PPARalpha activation is essential for HCV core protein-induced hepatic steatosis and hepatocellular carcinoma in mice, J. Clin. Invest. 118 (2008) 683–694.

[29] C. Rubie, K. Kempf, J. Hans, T. Su, B. Tilton, T. Georg, B. Brittner, B. Ludwig, M. Schilling, Housekeeping gene variability in normal and cancerous colorectal, pancreatic, esophageal, gastric and hepatic tissues, Mol. Cell. Probes. 19 (2005) 101–109.

Please cite this article as: Y. Hodo, et al., Comprehensive gene expression analysis of 5′-end of mRNA identified novel intronic transcripts associated with hepatocellular carcinoma, Genomics (2010), doi:10.1016/j.ygeno.2010.01.004

Liver INTERNATIONAL
OFFICIAL JOURNAL OF THE INTERNATIONAL ASSOCIATION FOR THE STUDY OF THE LIVER

Liver International ISSN 1478-3223

CLINICAL STUDIES

dUTP pyrophosphatase expression correlates with a poor prognosis in hepatocellular carcinoma

Hajime Takatori[1], Taro Yamashita[1], Masao Honda[1], Ryuhei Nishino[1], Kuniaki Arai[1], Tatsuya Yamashita[1], Hiroyuki Takamura[2], Tetsuo Ohta[2], Yoh Zen[3] and Shuichi Kaneko[1]

1 Department of Gastroenterology, Kanazawa University Graduate School of Medical Science, Ishikawa, Japan
2 Department of Gastroenterologic Surgery, Kanazawa University Graduate School of Medical Science, Ishikawa, Japan
3 Pathology Section, Kanazawa University Hospital, Ishikawa, Japan

Keywords
dUTP pyrophosphatase – hepatocellular carcinoma – prognosis – serial analysis of gene expression

Abbreviations
5-FU, 5-fluorouracil; dUTPase, dUTP pyrophosphatase; HCC, hepatocellular carcinoma; IHC, immunohistochemistry; qRT-PCR, quantitative reverse transcription-polymerase chain reaction; SAGE, serial analysis of gene expression.

Correspondence
Masao Honda, MD, Department of Gastroenterology, Kanazawa University Graduate School of Medical Science, 13-1 Takara-Machi, Kanazawa, Ishikawa 920-8641, Japan
Tel: +81 76 265 2233
Fax: +81 76 234 4250
e-mail: mhonda@m-kanazawa.jp

Received 13 August 2009
Accepted 26 October 2009

DOI:10.1111/j.1478-3231.2009.02177.x

Abstract

Background: Hepatocellular carcinoma (HCC) is a malignancy with a poor prognosis, partly owing to the lack of biomarkers that support its classification in line with its malignant nature. To discover a novel molecular marker that is related to the efficacy of treatment for HCC and its biological nature, we performed serial analysis of gene expression (SAGE) in HCC, normal liver and cirrhotic liver tissues. *Methods:* Gene expression profiles of HCC tissues and non-cancerous liver tissues were obtained by SAGE. Suppression of the target gene by RNA interference was used to evaluate its role in HCC *in vitro*. The relation of the identified marker and prognosis was statistically examined in surgically resected HCC patients. *Results:* We identified significant over-expression of *DUT*, which encodes dUTP pyrophosphatase (dUTPase), in HCC tissue, and this was confirmed in about two-thirds of the HCC samples by reverse-transcription polymerase chain reaction ($n = 20$). Suppression of dUTPase expression using short interfering RNAs inhibited cell proliferation and sensitized HuH7 cells to 5-fluorouracil treatment. Nuclear dUTPase expression was observed in 36.6% of surgically resected HCC samples ($n = 82$) evaluated by immunohistochemistry, and its expression was significantly correlated with the histological grades ($P = 0.0099$). Notably, nuclear dUTPase expression correlated with a poor prognosis with statistical significance (HR, 2.47; 95% CI, 1.08–5.66; $P = 0.032$). *Conclusion:* Taken together, these results suggest that nuclear dUTPase may be a good biomarker for predicting prognosis in HCC patients after surgical resection. Development of novel dUTPase inhibitors may facilitate the eradication of HCC.

Hepatocellular carcinoma (HCC) is the fifth most common malignancy and the third leading cause of cancer-related death worldwide (1). Several risk factors are responsible for HCC development, including alcoholism, aflatoxin and genetic diseases such as haemochromatosis and α-1 antitrypsin deficiency; however, the major risk factor is chronic hepatitis owing to hepatitis B virus (HBV) or hepatitis C virus (HCV) infection (2–4). Several treatment options are currently available for HCC management, which include liver transplantation, surgical resection, percutaneous ethanol injection, radio-frequency ablation, transcatheter arterial chemoembolization and systemic or local chemotherapy, and optimal treatment is determined based on tumour stage and liver function (5, 6). However, more than 80% of HCC cases develop advanced HCC after initial treatment (7).

Various chemotherapeutic drugs have been investigated for their antitumour activity in advanced HCC. For example, 5-fluorouracil (5-FU), a thymidylate synthase inhibitor, was the first reported drug studied for the treatment of advanced HCC; however, a median survival rate of 3–5 months has discouraged the further use of 5-FU as a single chemotherapeutic agent (8, 9). Interferon-α (IFN-α) has been reported to have antitumour activity against advanced HCC, and recent reports have suggested the efficacy of a combination of 5-FU/IFN-α for advanced HCC treatment (10–12), although convincing evidence for improved survival rate remains lacking. A recent study has indicated that 16% of advanced HCC patients responded positively to 5-FU/IFN-α treatment with clear and significant survival benefits compared with stable or progressive disease

patients (13). Thus, drug sensitivity appears to be one of the major determinants of the prognosis of advanced HCC patients treated with chemotherapy. Therefore, a hallmark of successful treatment would be the identification of useful biomarkers for determining the survival benefits offered by each treatment strategy.

In this study, we investigated the gene expression profiles of HCCs using serial analysis of gene expression (SAGE) to identify novel molecular markers or targets for the treatment of HCC (14–18). Here, we identified the upregulation of the *DUT* gene that encodes dUTP pyrophosphatase (dUTPase) in HCC. Markedly, HCC with a high nuclear dUTPase expression correlated with a poorly differentiated morphology and a poor prognosis. *DUT* gene knockdown not only suppressed cell proliferation but also sensitized HuH7 cells to low-dose 5-FU.

Materials and methods

Samples

All HCC tissues, adjacent non-cancerous liver tissues and normal liver tissues were obtained from 110 patients undergoing a hepatectomy between 1997 and 2006 in Kanazawa University Hospital, Kanazawa, Japan. Five normal liver tissue samples were obtained from patients undergoing surgical resection of the liver for the treatment of metastatic colon cancer. These samples were snap-frozen in liquid nitrogen immediately after resection. One hundred and five HCC and surrounding non-cancerous liver samples were obtained from patients undergoing surgical resection of the liver for HCC treatment, and part of these samples were used for the recent study (19). Three HCC and adjacent non-cancerous liver tissue samples were snap-frozen in liquid nitrogen and later used for SAGE. Twenty HCC tissues and their corresponding non-cancerous liver tissues were also snap-frozen and later used for real-time reverse transcription-polymerase chain reaction (RT-PCR) analysis, as described previously (19). Eighty-two additional HCC samples were formalin-fixed, paraffin-embedded and used for immunohistochemistry (IHC). HCC and adjacent non-cancerous liver tissues were histologically characterized, as reported elsewhere (19).

All strategies used for gene expression analysis as well as tissue acquisition processes were approved by the Ethics Committee and the Institutional Review Board of Kanazawa University Hospital. All procedures and risks were explained verbally to each patient, who then provided written informed consent.

Serial analysis of gene expression

Total RNA was purified from each homogenized tissue sample using a ToTally RNA extraction kit (Ambion Inc., Austin, TX, USA), and polyadenylated RNA was isolated using a MicroPoly (A) Pure kit (Ambion). A total of 2.5 μg of mRNA per sample was analysed by SAGE (20, 21). SAGE libraries were randomly sequenced at the Genomic Research Center (Shimadzu-Biotechnology, Kyoto, Japan), and the sequence files were analysed with SAGE 2000 software. The size of each SAGE library was normalized to 300 000 transcripts per library, and the abundance of transcripts was compared with SAGE 2000 software. Monte Carlo simulation was used for selecting genes whose expression levels were significantly different between the two libraries (22). Each SAGE tag was annotated using a gene-mapping website SAGE Genie database (http://cgap.nci.nih.gov/SAGE/) and the Source database (http://smd.stanford.edu/cgi-bin/source/sourceSearch), as described previously (23).

Quantitative reverse transcription-polymerase chain reaction

A 1 μg aliquot of each total RNA was reverse-transcribed using SuperScript II reverse-transcriptase (Invitrogen, Carlsbad, CA, USA). Real-time RT-PCR analysis was performed using the ABI PRISM 7700 sequence detection system (Applied Biosystems, Foster City, CA, USA). Using the standard curve method, quantitative PCR was performed in duplicate for each sample–primer set. Each sample was normalized relative to β actin. The assay IDs used were Hs00798995_s1 for dUTPase and Hs99999903_m1 for β actin.

RNA interference targeting *DUT*

Small interfering RNAs (siRNAs) targeting *DUT* or control (scrambled sequence) were synthesized by Dharmacon (Dharmacon Research Inc., Lafayette, CO, USA). The target sequences of *DUT* are 5′-AAGUUGU GAAAACGGACAUUC-3′ (DUT1) and 5′-CGGACAUU CAGAUAGCGCUTT-3′ (DUT2). Lipofectamine 2000™ reagent (Invitrogen) was used for transfection according to the manufacturer's instructions.

Cell proliferation assay, soft agar assay and matrigel invasion assay

Cell proliferation assays were performed using a Cell Titer96 Aqueous kit in quintuplicate (Promega, Madison, WI, USA). For the soft agar assay, 1×10^4 cells were suspended in 2 ml of 0.36% agar with growth medium and added in each well of a six-well plate containing a base layer of 0.72% agar. The plates were incubated at 37 °C in a 5% CO_2 incubator for 2 weeks. Matrigel invasion assays were performed using BD BioCoat™ Matrigel Matrix Cell Culture Inserts and Control Inserts (BD Biosciences, San Jose, CA, USA), as described in the manufacturer's instruction. 5-FU was obtained from Kyowa Kirin (Kyowa Kirin, Tokyo, Japan). All experiments were repeated at least twice.

Immunohistochemistry

Mouse monoclonal anti-dUTPase antibody M01 (Abnova Corporation, Taipei, Taiwan) and mouse antiproliferating

cell nuclear antigen (PCNA) monoclonal antibody PC10 (Calbiochem, San Diego, CA, USA) were used to evaluate the immunoreactivity of HCC and adjacent non-cancerous liver samples using a Dako EnVision+™ kit (Dako, Carpinteria, CA, USA), according to the manufacturer's instruction. Immunoreactivity was evaluated by determining the percentage of cells expressing dUTPase in the examined fields, graded as low (0–50%) or high (> 50%). The PCNA index was evaluated as described previously (19).

Statistical analysis

Student's *t*-test was used to determine the statistical significance of the differences in cell viability between the two groups. The Mann–Whitney *U*-test was used for the analysis of gene expression between chronic liver disease (CLD) and HCC tissues. The χ^2-test was used to evaluate the correlation between clinicopathological characteristics and dUTPase expression status. Univariate and multivariate Cox proportional hazards regression analysis was used to evaluate the association of dUTPase expression and clinicopathological parameters with patient outcome. All statistical analyses were performed using SPSS software (SPSS software package; SPSS Inc., Chicago, IL, USA) and GRAPHPAD PRISM software (GraphPad Software Inc., La Jolla, CA, USA).

Results

Gene expression profiling identified the overexpression of *DUT* in hepatocellular carcinoma

To overcome the considerable individual variability of transcriptomic characteristics, we constructed a SAGE library of normal human liver using RNAs derived from five normal liver tissues. In addition, we constructed two SAGE libraries derived from three HCC tissues or corresponding non-cancerous liver tissues from patients who developed HCC with a history of chronic hepatitis C. We detected a total of 226 267 tags corresponding to 45 746 unique tags from these SAGE libraries (supporting information Table S1). After excluding the tags detected only once in each library, we selected 15 333 reliable unique transcripts expressed in at least one of the SAGE libraries to avoid contamination of tags derived from sequence errors. Then, we annotated these transcripts using SAGE Genie database and the Source database to identify the potential subcellular localization of transcripts categorized into eight groups in each SAGE library.

The number of nuclear component-related transcripts was increased in the HCC library compared with the normal liver and non-cancerous liver libraries, whereas the other cellular component-related transcripts did not show this tendency (supporting information Fig. S1). Because nuclear component-related genes may closely correlate with cancer cell proliferation and chemosensitivity (24), we further investigated the expression of nuclear component-related tags in each library, and identified 10 transcripts associated with nucleotide/nucleoside metabolism that are overexpressed in HCC (Table 1). Using Monte Carlo simulation, we evaluated the significance of differentially expressed transcripts in HCC and corresponding CLD libraries or in HCC and normal liver libraries. We identified a *DUT* gene encoding dUTPase (dUTPase) whose expression was significantly altered ($P = 0.01$). We also identified a *TS* gene encoding thymidylate synthase in the list, but the difference did not reach statistical significance.

dUTPase is a phosphatase known to maintain a dUMP pool by catalysing the hydrolysis of dUTP to dUMP, and thus provides a substrate of thymidylate synthase. Its role in HCC is unknown; therefore, we examined *DUT* expression in 20 independent HCC and corresponding non-cancerous liver tissues, and identified significant overexpression of *DUT* in HCC tissue ($P = 0.0015$) (Fig. 1A). Moreover, we detected more than a two-fold increase in *DUT* expression in 70% of HBV-related and HCV-related HCC cases (14 of 20 HCCs) compared with the non-cancerous liver tissues (Fig. 1B). We further examined the expression of *DUT* in 238 HCC tissues compared with the non-cancerous liver tissues using publicly available microarray data (GSE5975) (Fig. S2). Consistent with the SAGE data, *DUT* was overexpressed more than two-fold in 121 of 238 HCC tissues (median: 2.03), whereas *TS* was overexpressed more than two-fold in 54 of 238 HCC tissues (median: 1.41) compared with the non-cancerous liver tissues.

Pivotal role of dUTP pyrophosphatase expression in cell proliferation in hepatocellular carcinoma cell lines

In general, cancer gene signatures discovered by comparison between tumour and non-tumour tissues are more likely to reflect the differences in the control of cell proliferation and growth (25). Accordingly, we investigated the function of dUTPase in cell proliferation in HuH7 cells by *DUT* gene knockdown. *DUT* expression was decreased by 60–70% following the transfection of the siRNA constructs specifically targeting *DUT* 48 h after transfection (DUT1 in Fig. 2A and DUT2 in Fig. S3A), and cell growth was significantly inhibited compared with the control 72 h after transfection (Fig. 2B and Fig. S3B). Anchorage-independent cell growth was also significantly impaired by *DUT* gene knockdown 14 days after transfection (Fig. 2C). Furthermore, *DUT* gene knockdown decreased the numbers of both migrating and invading cells 72 h after transfection (Fig. 2D and E).

dUTPase is known to be associated with thymidylate synthesis (26), and thus we evaluated the effects of 5-FU, a thymidylate synthase inhibitor, on dUTPase expression in HCC cell lines *in vitro*. When we treated HuH7 cells with low-dose 5-FU (0.25 mg/ml), we could not detect any growth-inhibitory effects (Fig. 2F). Based on this condition, we evaluated the effect of *DUT* gene knockdown on 5-FU sensitivity 72 h after transfection.

Table 1. Genes associated with nucleic acid metabolism overexpressed in hepatocellular carcinoma

Tag sequence	Normal liver	Non-cancerous liver	HCC	Fold*	Gene	*P*-value†
CAGCTCCGCT	0	2	11	5.5	dUTP pyrophosphatase	0.010
AAAGGATAAT	0	0	3	> 3	General transcription factor II H, polypeptide 2	0.127
ACGGTCCAGG	0	0	3	> 3	Cytidine deaminase	0.127
ATGTAGAGTG	0	0	3	> 3	Thymidylate synthase	0.127
TGGGGATTAC	1	0	3	> 3	Zinc ribbon domain containing, 1	0.127
CACCCTGTAC	2	2	6	3	Solute carrier family 29	0.147
GAACGCCTAA	1	1	3	3	Dihydropyrimidinase-like 2	0.308
GCGCTGGTAC	0	1	3	3	2′-5′-oligoadenylate synthetase 3	0.308
CTTAGTGCAA	0	2	4	2	3′-phosphoadenosine 5′-phosphosulphate synthase 2	0.335
TTGTTACATC	0	2	3	1.5	Phosphoribosyl pyrophosphatase synthetase-associated protein 1	0.506

*Fold increase was calculated by dividing the number of tags in HCC by that of tags in non-cancerous liver. To avoid division by 0, a tag value of 1 was used for any tag that was not detectable in one sample.

†Statistical significance of differentially expressed genes between two groups (HCC and non-cancerous liver libraries) was calculated using Monte Carlo simulation.

HCC, hepatocellular carcinoma.

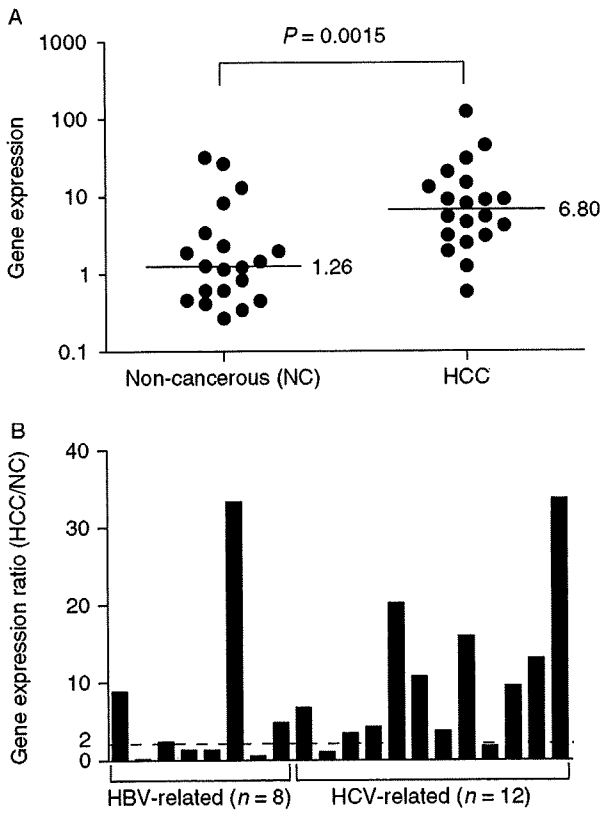

Fig. 1. (A) Quantitative reverse transcription-polymerase chain reaction analysis of *DUT* expression in hepatocellular carcinoma (HCC) and corresponding non-cancerous liver tissues. *DUT* was significantly activated in HCC tissues compared with non-cancerous liver tissues ($P = 0.0015$). A median value in each group is indicated. (B) *DUT* gene expression ratios of HCC and corresponding non-cancerous liver tissues. Fourteen of 20 HCC tissues expressed *DUT* more than two-fold compared with the background non-cancerous liver tissues. HBV, hepatitis B virus; HCV, hepatitis C virus.

Interestingly, *DUT* gene knockdown not only suppressed cell proliferation but also sensitized HuH7 cells to low-dose 5-FU (Fig. 2F and Fig. S3B). These data suggest that dUTPase overexpression in HCC tissues may be associated with enhanced cell proliferation and 5-FU resistance.

Intense dUTP pyrophosphatase expression is correlated with a poor prognosis in hepatocellular carcinoma patients

To characterize the clinicopathological characteristics of dUTPase expression in HCC, we performed IHC using an additional independent HCC cohort. Accordingly, we explored the dUTPase expression in HCC using 82 formalin-fixed paraffin-embedded HCC specimens. All HCC tissues were surgically resected at the Liver Disease Center of Kanazawa University Hospital with full clinical information, and their immunoreactivity to anti-dUTPase antibodies was evaluated by IHC. We noticed that anti-dUTPase antibodies reacted to both nuclear (red arrows) and cytoplasmic (blue arrows) isoforms of dUTPase, as described previously (26) (Fig. 3A and B). We therefore evaluated the nuclear and cytoplasmic expression of dUTPase separately. We stratified HCC tissues and evaluated the dUTPase expression status based on the percentages of dUTPase-positive cells. The frequency of nuclear or cytoplasmic dUTPase-positive cells was highly variable in each HCC tissue, and we defined HCCs with nuclear or cytoplasmic dUTPase expressed in ≥50% of tumour cells as nuclear or cytoplasmic dUTPase-high HCC (Fig. 3C). Nuclear dUTPase overexpression was detected in 36.6% (30 of 82), whereas cytoplasmic dUTPase overexpression was detected in 67.1% (55 of 82) of HCC tissues compared with the corresponding non-cancerous liver tissues

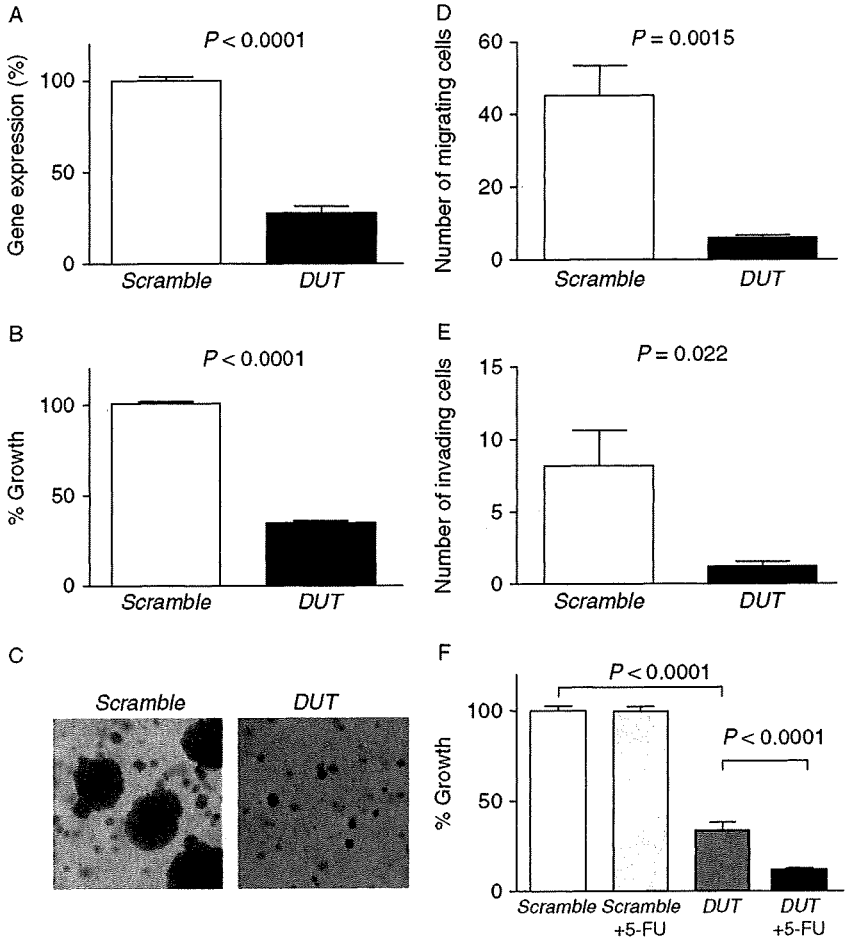

Fig. 2. (A) Transfection of small interfering RNAs targeting *DUT* (DUT1) decreased *DUT* expression compared with the control (scrambled sequence). Gene expression was evaluated in triplicate 72 h after transfection (mean ± SD). (B) *DUT* gene knockdown significantly suppressed cell proliferation ($P < 0.0001$). Cell viability was evaluated in triplicate 72 h after transfection (mean ± SD). (C) Soft agar assay. *DUT* gene knockdown suppressed anchorage-independent cell growth. (D and E) Matrigel invasion assay. *DUT* gene knockdown decreased the numbers of both migrating and invading cells. Experiments were performed in triplicate (mean ± SD). (F) *DUT* gene knockdown sensitized HuH7 cells to low-dose 5-fluorouracil (5-FU) (0.25 μg/ml), which had no effect on the cell proliferation in the control (mean ± SD).

(Table 2). In general, non-cancerous hepatocytes rarely expressed nuclear dUTPase (Fig. 3A).

We investigated the clinicopathological characteristics of nuclear or cytoplasmic dUTPase in low /high HCC cases (Table 2). The expression status of nuclear dUTPase showed no correlation with age, gender, virus, presence of cirrhosis, α-fetoprotein value, tumour size and TNM stages. However, nuclear dUTPase expression was significantly correlated with the histological grades of HCC ($P = 0.0099$), and high frequencies of nuclear dUTPase-positive cells were associated with poorly differentiated cell morphology in the HCC tissue. In contrast, cytoplasmic dUTPase expression was not correlated with the histological grades of HCC ($P = 0.077$). We examined the cell proliferation of these HCC samples by PCNA staining, and PCNA indexes were significantly higher in nuclear dUTPase high HCC than low HCC with statistical significance ($P = 0.01$) (Fig. S4).

We further investigated the prognostic significance of dUTPase expression in HCC. Strikingly, high nuclear dUTPase expression in HCC tissue correlated with a poor survival outcome compared with low nuclear dUTPase expression ($P = 0.0036$), whereas high cytoplasmic dUTPase expression had little effects when evaluated by recurrence-free survival (Fig. 3D). Furthermore, univariate Cox regression analysis showed a significant correlation between high nuclear dUTPase expression and a high risk of mortality (HR, 2.47; 95% CI, 1.08–5.66; $P = 0.032$; Table 3). By multivariate Cox regression analysis, TNM stage (HR, 2.75; 95% CI, 1.11–6.79; $P = 0.027$) and nuclear dUTPase (HR, 2.61; 95% CI, 1.13–6.05; $P = 0.024$) were independent prognostic factors associated with a high risk of mortality, and other clinicopathological features did not add independent prognostic information. These data indicate a significant correlation between the malignant potential of