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

Title: **Comprehensive gene expression analysis of 5'-end of mRNA identified novel intronic transcripts associated with hepatocellular carcinoma**

Authors: Yuji Hodo[1], Shin-ichi Hashimoto[2], Masao Honda[1], Taro Yamashita[1], Yutaka Suzuki[3], Sumio Sugano[3], Shuichi Kaneko[1] and Kouji Matsushima[2]

Affiliations: [1] Department of Gastroentelorogy, Kanazawa University Graduate School of Medical Science, 13-1 Takara-Machi, Kanazawa, Ishikawa 920-8641, Japan

[2] Department of Molecular Preventive Medicine, School of Medicine, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

[3] Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa, Chiba 277-8562, Japan

Keywords: 5'-end serial analysis of gene expression, transcriptional start site, acyl-coenzyme A oxidase 2, intron, hepatocellular carcinoma

Corresponding author: Shuichi Kaneko, M.D., Ph.D., Department of Gastroentelorogy, Kanazawa University Graduate School of Medical Science, 13-1 Takara-Machi, Kanazawa, Ishikawa 920-8641, Japan

Tel: +81-76-265-2235

Fax: +81-76-234-4250

Email: skaneko@m-kanazawa.jp

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

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.

1. 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 techinque 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.

2. Results

2.1 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 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 8,410 unique genes (4,397 genes detected in the NL library, 5,194 genes in the NT library, and 6,304 genes in the T library).

2.2 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 metallotionein 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 angiogensis 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).

2.3 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 (Table 3a & b). For example, a gene encoding *coagulation factor XIII, B polypeptide* was thirteen-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 fifty-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.

2.4 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].

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.

2.5 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.

3. 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.

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.

4. 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 (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 EcoRI/MmeI 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 MmeI 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 EcoRI. 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 EcoRI 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'-TGCACCACTTACTGAGCATCTACTC-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 epression levels between two groups. All statistical analyses were performed using R (http://www.r-project.org/).

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

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.

Figure legends

Figure 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 indicates 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; and T, tumor.

*: $P < 0.01$, #: $P < 0.05$

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	4,397	5,194	6,304	8,410

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

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		
NL	NT	T	NT/NL	T/NL	Gene
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 indicates normal liver; NT, non-tumor; and T, tumor.

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

5’SAGE T/NL	3’SAGE T/NL	5’/3’ ratio	Gene
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 famliy, 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 1M
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 indicates 3’-end serial analysis of gene expression; 5’SAGE, 5’-end serial analysis of gene expression; HCC, hepatocellular carcinoma; NL, normal liver; and T, tumor.

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

5'SAGE T/NL	3'SAGE T/NL	5'/3' ratio	Gene
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, 45kDa, 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 indicates 3'-end serial analysis of gene expression; 5'SAGE, 5'-end serial analysis of gene expression; HCC, hepatocellular carcinoma; NL, normal liver; and T, tumor.

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

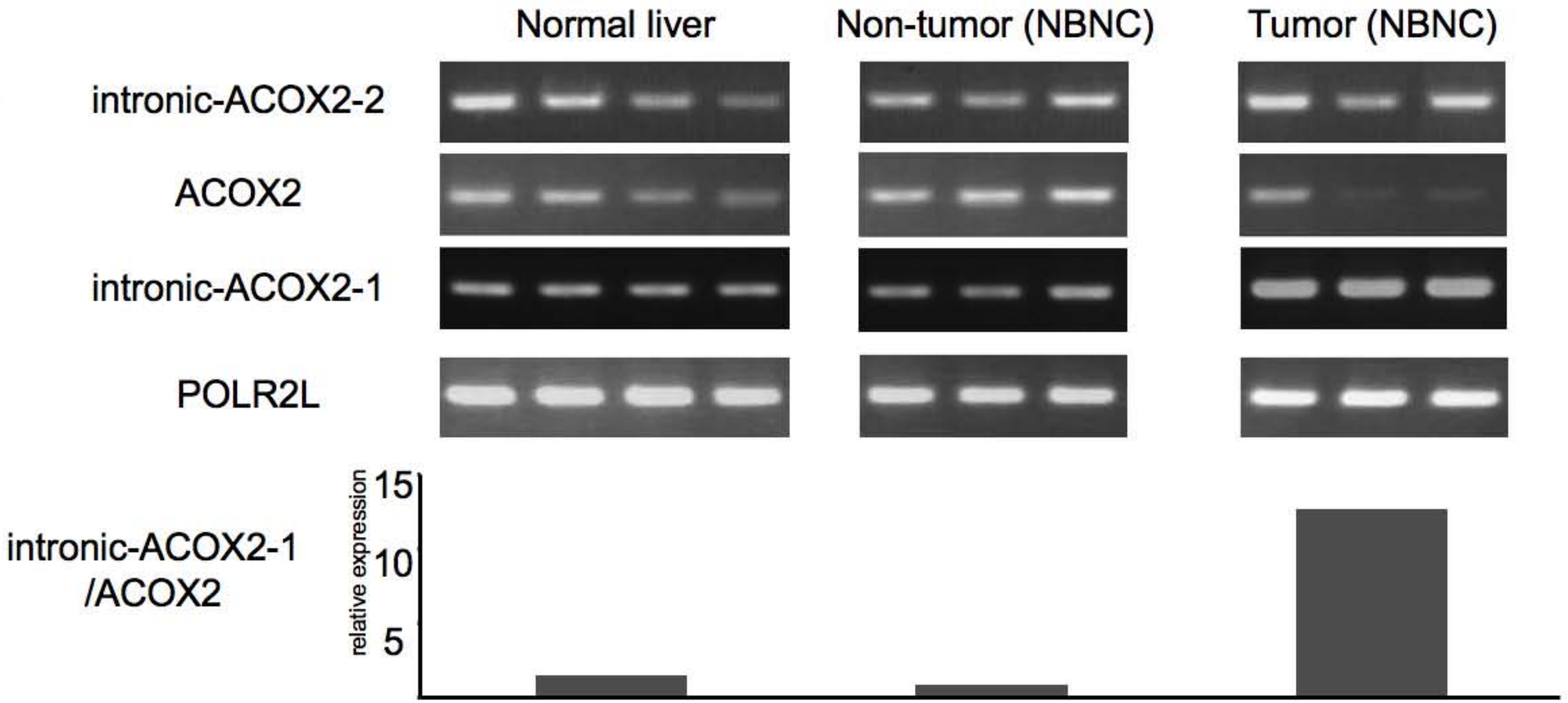
Table 4b. Differentially expressed intronic promoter regions in HCC

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)

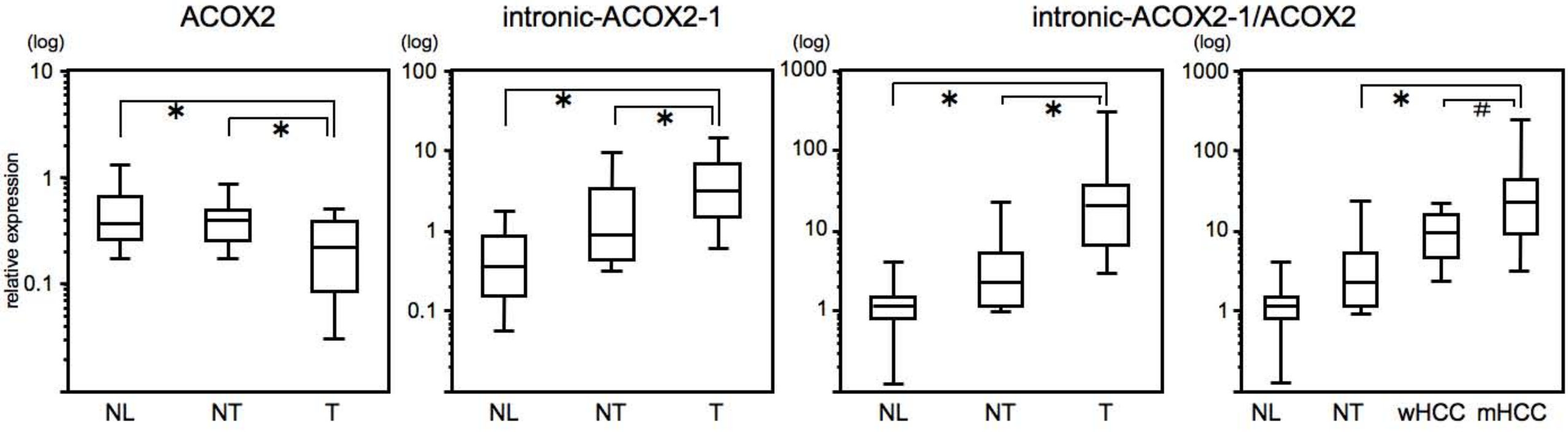
3’SAGE indicates 3’-end serial analysis of gene expression; 5’SAGE, 5’-end serial analysis of gene expression; HCC, hepatocellular carcinoma; NL, normal liver; NT, non-tumor; and T, tumor.

Figure 1

(A)


Normal liver
Non-tumor (NBNC)
Tumor (NBNC)
intronic-ACOX2-2
ACOX2
intronic-ACOX2-1
POLR2L
intronic-ACOX2-1 /ACOX2
relative expression
15
10
5

(B)


ACOX2
intronic-ACOX2-1
intronic-ACOX2-1/ACOX2
(log)
relative expression
10
1
0.1
100
1000
*
#
NL
NT
T
wHCC
mHCC

Supplemental Figure Captions

Supplemental Figure 1

RT-PCR analysis and 5′SAGE tag abundance. RT-PCR of representative genes differentially expressed in the NT and T libraries are shown with the tag number in each 5′ SAGE library. At least three samples showed similar differential expression patterns, as identified by 5′SAGE.

CLEC4G indicates C-type lectin superfamily 4 member G; DHRS10, dehydrogenase/reductase member 10; DUSP23, dual specificity phosphatase 23; LGALS4, galectin 4; NT, non-tumor; POLR2L, polymerase (RNA) II (DNA directed) polypeptide L; RDH11, retinol dehydrogenase 11; STRA13, stimulated by retinoic acid 13 homolog;T, tumor; TMEM14A, transmembrane 14A; and XAGE 1, X antigen family, member 1.

Supplemental Figure 2

Using MetaCore^{TM} software, we compared the Gene Ontology process networks of the up-regulated gene in three types of hepatocellular carcinoma (HCC) (HBV-, HCV-,

non-B non-C). Ten statiscally significant and similar networks are shown in the log scale.

Supplemental Figure 3

Scheme of clustering for the 5’-end serial analysis of gene expression (5’SAGE) tags mapped to the intronic region. Blue boxes and lines indicate the RefSeq sequences. We clustered the 5′SAGE tags by binning them at >500-bp intervals and assuming putative intronic promoter regions.

Supplemental Figure 4

Identification of a novel transcript transcribed from the intronic region. The boxes and lines represent exons and introns, respectively. Two potential intron-origin transcripts of ACOX2 were identified. 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*. The sequence of the intronic part was unique, and the remaining part of the sequence was shared with the canonical transcripts of *ACOX2*.

Supplemental Table 1. Clinical features of 43 HCC samples for RTD-PCR

Age (years)	61.9±1.4
Sex (men:women)	36:7
HBsAg (+) : (-) : (-) HCVAb (-) (+) (-)	19:20:4
Histology (well: moderately)	7:36
AFP (ng/μl)	363±177
PIVKA-□(mAU/ml)	1760±817

HCC indicates hepatocellular carcinoma; and RTD-PCR, real-time detection-polymerase chain reaction.

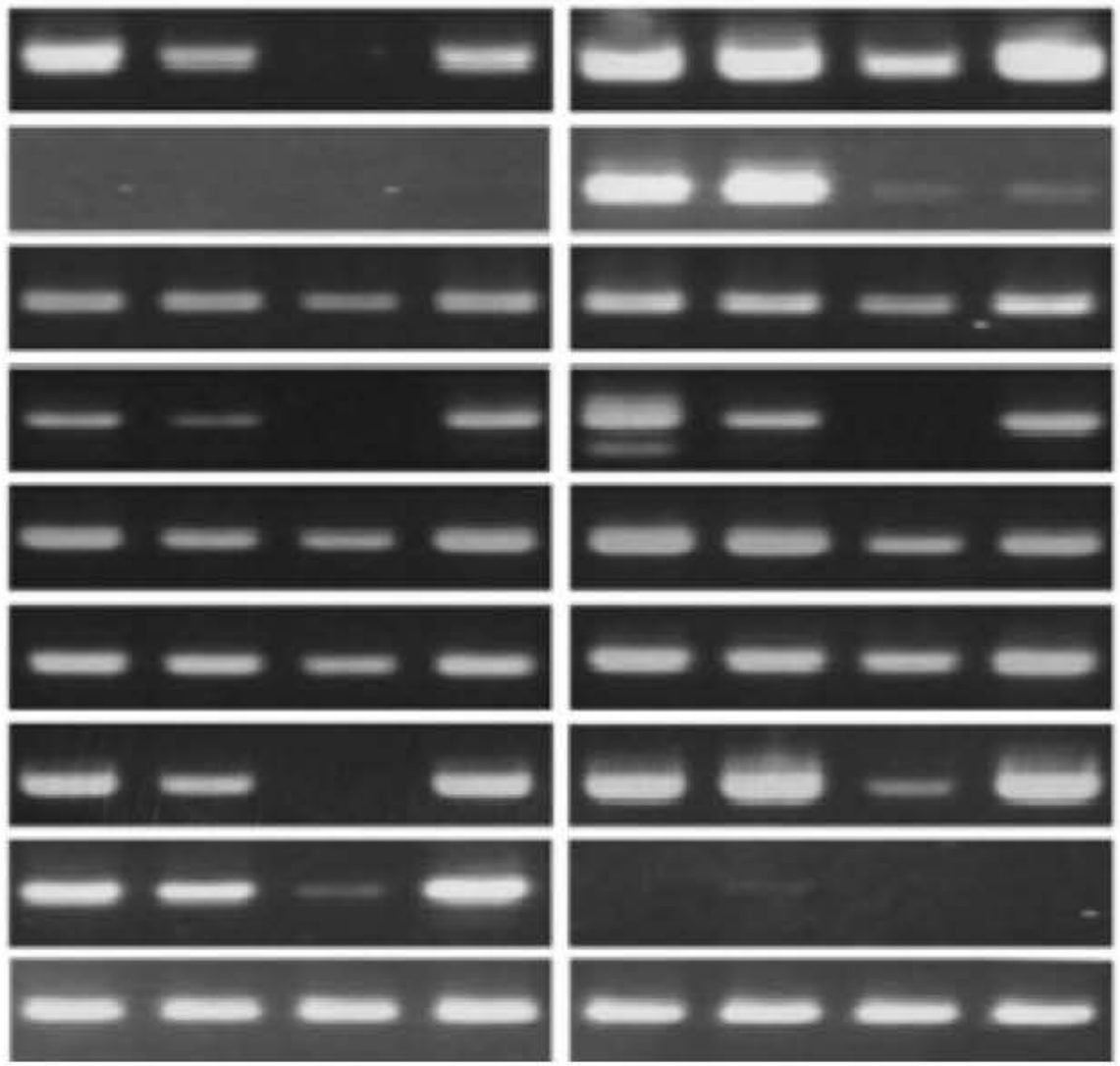
Supplemental Table 2. Primer sequences for RT-PCR

Gene	Forward	Reverse
LGALS4	AGATGGTCACCCACCTGCAA	CACAGCGAATGGACAGATCAA
XAGE1	TCTGGATTCTTTCTCCGCTACTG	TGCGTTGTTTCAGCTTGTCTTC
RDH11	ATGGGTGTCTGTCCAAGCTC	GTTGGTTGGGAGGAACTGAA
DHRS10	CCGAGTGGTTATCTGCGACAA	ACTCGGACACCATATGGACTTTC
TMEM14A	CCTGCTGTAATGGGCAGAACA	ATAGCACTGGCAGCAATACAATTAA
STRA13	AGCCCCTGGTCCACAGAAG	TATCGCATGGCTGGAAAGGT
DUSP23	CAACTTCTCCTGGGTGCTTC	TTCGTTCGCTGGTAGAACTG
CLEC4G	GCGCACCTGGTGATCGTT	CTATGTTGGGCCAAGTGTTTCTG
POLR2L	GGAGAAGTGACCACGCTGAAA	TTGCCCGGCCTAAATCTG
ACOX2	CTTTGCACTGACCAATGCTG	GATGGTGCTGGTTGCTTCTT
intronic-ACOX2-	TCAGTAAGTGGTGCACAGGG	TGCGTCAGGGTCTGTAAATG
intronic-ACOX2-	AATTTAGTGGTTTGCAGGACA	TGCGTCAGGGTCTGTAAATG

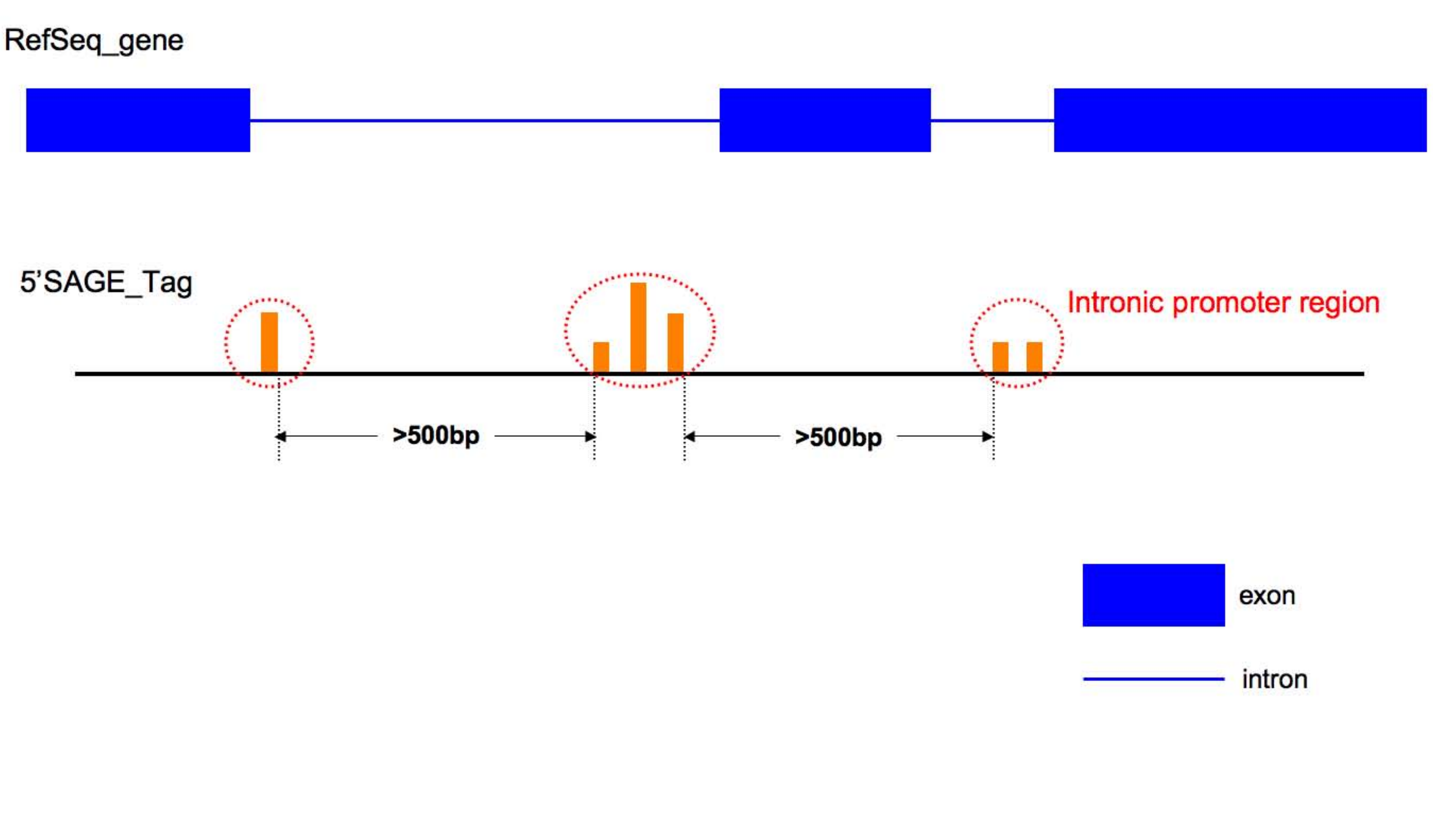
ACOX2 indicates acyl-coenzyme A oxidase 2; CLEC4G, C-type lectin superfamily 4 member G; DHRS10, dehydrogenase/reductase member 10; DUSP23, dual specificity phosphatase 23; LGALS4, galectin 4; POLR2L, polymerase (RNA) II (DNA directed) polypeptide L; RDH11, retinol dehydrogenase 11; RT-PCR, reverse transcriptase-polymerase chain reaction; STRA13, stimulated by retinoic acid 13 homolog; TMEM14A, transmembrane 14A; and XAGE 1, X antigen family, member 1.

Supplemental Figure 1

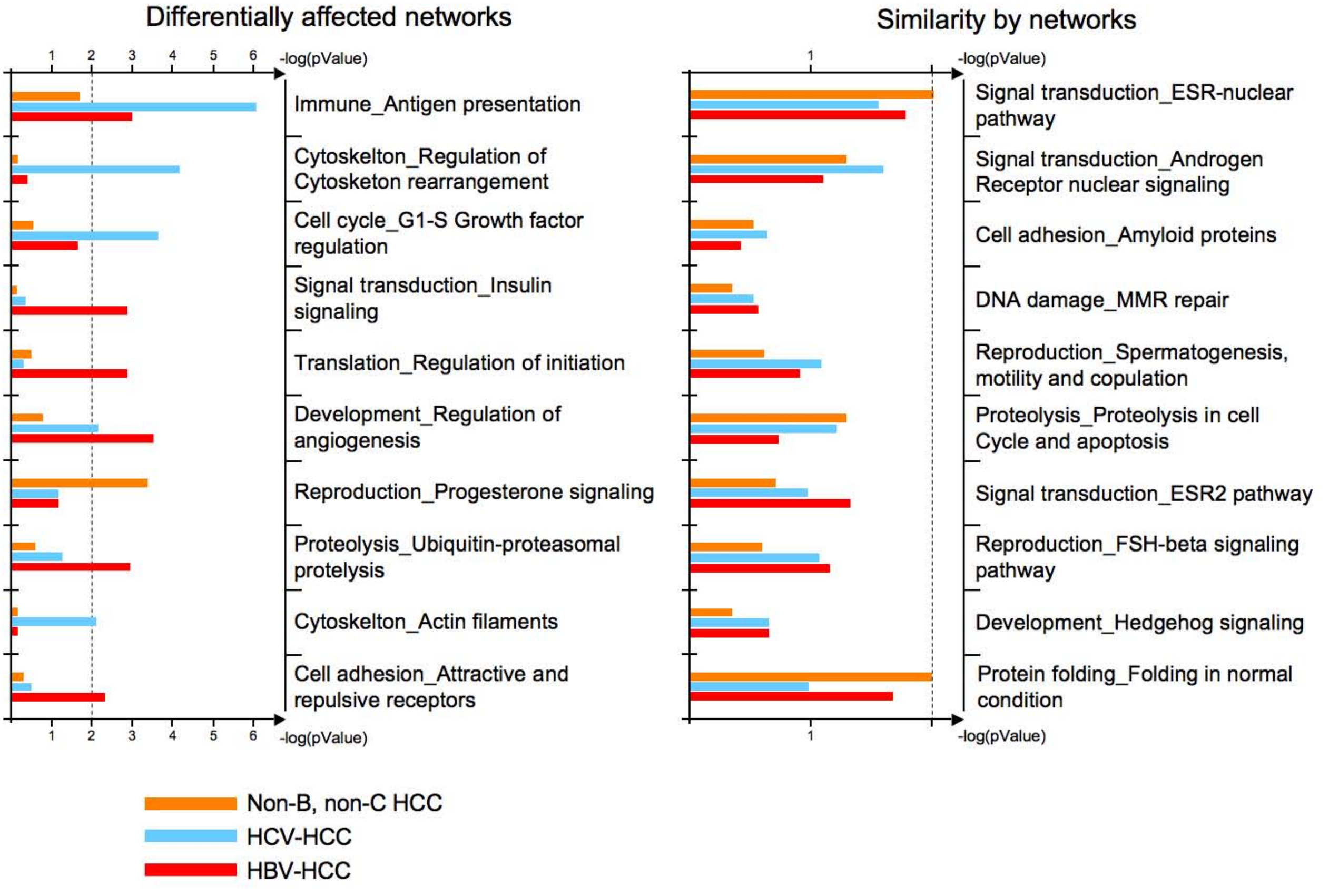
Gene	Tag number NT	Tag number T
LGALS4	2	18
XAGE1	0	14
RDH11	2	13
DHRS10	5	13
TMEM14A	1	14
STRA13	2	12
DUSP23	0	10
CLEC4G	5	0
POLR2L	29	29



Supplemental Figure 3



Supplemental Figure 2



Supplemental Figure 4