The baseline characteristics and test results are listed in Table 1. The overall rate of RVR/cEVR was 43% in the model building set and 48% in the validation set. There were no significant differences in the clinical backgrounds between these two groups. Hepatitis C viral mutations, such as mutations in interferon-sensitivity determining region or core amino acid residues 70 and 91, were not included in the present analysis. The dataset of laboratory tests was based on the digitized records in this hospital. Continuous data was split into categorized data by increment of 10; For example, age was categorized into <30, 30–39, 40–49, 50–59, 60–69, and ≥70.

Statistical analysis

Based on this database, the recursive partitioning analysis algorithm referred to as CART was implemented to define meaningful subgroups of patients with respect to the possibility of achieving RVR/cEVR. The CART belongs to a family of nonparametric regression methods based on binary recursive partitioning of data. The software automatically explore the data to search for optimal split variables, builds a decision tree structure and finally classifies all subjects into particular subgroups that are homogeneous with respect to the outcome of interest.[18] During the CART analysis, first, the entire study population, and thereafter, all newly defined subgroups, were investigated at every step of the analysis to determine which variable at what cut-off point yielded the most significant division into two prognostic subgroups that were as homogeneous as possible with respect to estimates of RVR/cEVR possibilities. This algorithm uses the impurity function (Gini criterion function) for splitting.[19] A restriction was imposed on the tree construction such that terminal subgroups resulting from any given split must have at least 20 patients. The CART procedure stopped when either no additional significant variable was detected or when the sample size was below 20. The resulting final subgroups were most homogeneous with respect to the probability of achieving RVR/cEVR. For this analysis, data mining software Clementine version 12.0 (SPSS Inc, Chicago, IL) was utilized. SPSS 15.0 (SPSS Inc, Chicago, IL) was used for logistic regression analysis.

RESULTS

Factors associated with RVR/cEVR by standard statistical analysis

WE FIRST ANALYZED 72 variables by univariate and multivariate logistic regression analysis to find factors associated with RVR/cEVR (Table 2). Patients with RVR/cEVR were significantly younger than those without. Among histological findings, grade of steatosis and stage of fibrosis was significantly lower in RVR/cEVR. Among hematologic tests, hemoglobin and hematcrit was significantly higher in RVR/cEVR. Among blood chemistry tests, creatinine and low-density lipoprotein cholesterol (LDL-C) was significantly higher and gamma-glutamyltransferase (GGT), low-density-lipoprotein cholesterol (LDL-C), and blood sugar were significantly lower in RVR/cEVR. The level of HCV RNA was significantly lower in RVR/cEVR. There were no significant differences in other tests.

Multivariate logistic regression analysis was performed on age, fibrosis stage, steatosis, HCVRNA, creatinine, hemoglobin, GGT, LDL-C, and blood sugar: hematcrit was not included since it is closely associated with hemoglobin. On multivariate analysis, age, grade of steatosis, level of HCV RNA, creatinine, hemoglobin, GGT, and LDL-cholesterol remained significant whereas stage of fibrosis, hemoglobin and blood sugar were not.

The CART analysis

The CART analysis was carried out on the model building set of 269 patients using the same variables as logistic regression analysis. Figure 1 shows the resulting decision tree. The CART analysis automatically selected five predictive variables to produce a total of seven subgroups of patients. The grade of steatosis was selected as the variable of initial split with an optimal cut-off of 30%. The possibility of achieving RVR/cEVR was only 18% for patients with hepatic steatosis of 30% or more compared to 47% for patients with hepatic steatosis of less than 30%. Among patients with hepatic steatosis of less than 30%, the level of serum LDL-C, with an optimal cut-off of 100 mg/dL, was selected as the variable of second split. Patients with higher LDL-C level had the higher probability of RVR/cEVR (57% vs. 32%). Among patients with LDL-C of less than 100 mg/dL, age, with an optimal cut-off of 60, was selected as the third variable of split. Younger patients had the higher probability of RVR/cEVR (49% vs. 15%). Among patients younger than 60, the blood sugar, with an optimal cut-off of 120 mg/dL, was selected as the forth variable of split. Patients with lower blood sugar level had the higher probability of RVR/cEVR (71% vs. 31%). Among patients with hepatic steatosis of less than 30% and LDL-C of 100 mg/dL or more, age, with an optimal cut-off of 50, was selected as the third variable of split, younger being the predictor of higher RVR/cEVR probability (77% vs. 50%). Among patients older than 50,

© 2009 The Japan Society of Hepatology

Table 1 Clinical characteristics of patients

	Model set $n = 269$	Validation set $n = 131$	*P*-value
Sex (M/F)	127/142	55/76	0.325
Age (years)	57.7 ± 10.1	57.6 ± 10.0	0.932
Body weight (kg)	59.6 ± 11.0	57.5 ± 9.5	0.094
Body mass index (kg/m^2)	23.2 ± 3.1	23.3 ± 3.8	0.934
Total protein (g/dL)	7.6 ± 0.5	7.7 ± 0.6	0.558
Albumin (g/dL)	4.2 ± 0.3	4.2 ± 0.3	0.349
Globulin (g/dL)	3.4 ± 0.5	3.4 ± 0.6	0.989
Aspartate aminotransferase (IU/L)	58.1 ± 43.1	55.8 ± 37.5	0.601
Alanine aminotransferase (IU/L)	70.9 ± 49.2	66.4 ± 52.6	0.462
Gamma-glutamyltransferase (IU/L)	49.6 ± 44.0	45.2 ± 34.4	0.33
Lactate dehydrogenase (IU/L)	289.3 ± 112.3	301.5 ± 109.3	0.417
Total bilirubin (mg/dL)	0.71 ± 0.28	0.69 ± 0.23	0.317
Direct bilirubin (mg/dL)	0.23 ± 0.12	0.25 ± 0.10	0.147
Indirect bilirubin (mg/dL)	0.48 ± 0.21	0.44 ± 0.16	0.064
Alkaline phosphatase (IU/L)	290.9 ± 107.6	292.5 ± 107.6	0.917
Leucine aminopeptidase (IU/L)	64.3 ± 14.3	65.5 ± 12.3	0.543
Thymol turbidity test (KU)	7.1 ± 3.4	8.0 ± 3.7	0.062
Zinc sulfate turbidity test (KU)	15.4 ± 4.9	16.3 ± 5.4	0.188
Choline esterase (IU/L)	318.1 ± 81.7	321.1 ± 78.1	0.798
Ammonia (microg/dL)	39.7 ± 20.2	45.0 ± 15.6	0.668
Blood sugar (mg/dL)	125.9 ± 41.1	117.4 ± 47.9	0.081
Glycohemoglobin (%)	5.6 ± 1.6	5.4 ± 1.2	0.797
Total cholesterol (mg/dL)	170.8 ± 33.9	175.6 ± 36.8	0.170
Low-density-lipoprotein-cholesterol (mg/dL)	96.5 ± 25.2	100.9 ± 28.5	0.153
High-density-lipoprotein-cholesterol (mg/dL)	54.2 ± 15.9	55.2 ± 17.4	0.612
Triglyceride (mg/dL)	108.5 ± 47.8	102.8 ± 46.4	0.306
Creatinine (mg/dL)	0.72 ± 0.15	0.74 ± 0.17	0.236
Urea nitrogen (mg/dL)	14.1 ± 3.4	14.9 ± 3.9	0.123
Uric acid (mg/dL)	5.3 ± 1.2	5.2 ± 1.2	0.715
Sodium (mEq/L)	142.2 ± 2.0	142.4 ± 2.0	0.471
Potassium (mEq/L)	4.3 ± 0.3	4.3 ± 0.4	0.578
Chloride (mEq/L)	104.0 ± 2.2	104.0 ± 2.6	0.905
Calcium (mg/dL)	9.1 ± 0.4	9.2 ± 0.4	0.479
Phosphorus (mg/dL)	3.5 ± 0.5	3.5 ± 0.6	0.814
Magnesium (mg/dL)	2.2 ± 0.2	2.3 ± 0.3	0.390
Amylase (IU/L)	178.7 ± 125.8	175.1 ± 133.1	0.118
Creatine kinase (IU/L)	114.9 ± 147.6	119.3 ± 73.7	0.849
Iron (microg/dL)	104.7 ± 53.2	109 ± 37	0.726
Ferritin (ng/mL)	111.3 ± 103.3	59.7 ± 118.5	0.405
C-reactive peptide (mg/dL)	0.2 ± 1.1	0.1 ± 0.1	0.586
Immunoglobulin G (mg/dL)	1849 ± 426	1988 ± 525	0.129
Immunoglobulin M (mg/dL)	141 ± 69	205 ± 106	0.200
Immunoglobulin A (mg/dL)	323 ± 675	291 ± 81	0.784
Triiodothyronine (pg/mL)	2.3 ± 0.3	2.2 ± 0.3	0.358
Thyroxin (ng/dL)	0.9 ± 0.1	0.9 ± 0.1	0.872
Thyroid stimulating hormone (micro IU/mL)	1.8 ± 1.4	1.7 ± 0.7	0.939
White blood cell count (/microl)	5243 ± 1591	5286 ± 1101	0.843
Segmented neutrophils (%)	55.4 ± 10.8	57.0 ± 10.0	0.297
Band neutrophils (%)	1.5 ± 1.6	0.5 ± 0.6	0.250
Eosinophils (%)	2.9 ± 2.3	2.4 ± 1.4	0.127

© 2009 The Japan Society of Hepatology

Table 1 *Continued*

	Model set $n = 269$	Validation set $n = 131$	*P*-value
Basophiles (%)	0.6 ± 0.4	0.6 ± 0.3	0.727
Lymphocytes (%)	34.6 ± 9.6	34.0 ± 9.3	0.682
Monocytes (%)	6.6 ± 2.2	6.2 ± 2.6	0.149
Red blood cell count (10^4/microl)	458 ± 43	455 ± 47	0.643
Hemoglobin (g/dL)	14.4 ± 1.5	14.5 ± 1.5	0.618
Hematcrit (%)	42.7 ± 4.0	42.9 ± 4.4	0.717
Reticulocytes (%)	1.4 ± 0.4	1.4 ± 0.4	0.762
Mean corpuscular volume (fL)	93.3 ± 4.5	93.8 ± 5.41	0.466
Mean corpuscular hemoglobin concentration (pg)	31.5 ± 1.9	31.7 ± 2.3	0.583
Mean corpuscular hemoglobin concentration (g/dL)	33.8 ± 0.9	33.7 ± 1.3	0.910
Platelets (10^4/microl)	16.8 ± 5.4	16.3 ± 4.5	0.480
Prothrombin time (s)	11.7 ± 1.2	11.7 ± 0.9	0.762
Prothrombin time (activity %)	104.6 ± 14.4	102.6 ± 14.8	0.363
Prothrombin time (international normalized ratio)	1.0 ± 0.1	1.0 ± 0.1	0.387
Thrombin time (%)	97.2 ± 31.3	109 ± 31.5	0.231
Activated partial thromboplastin time (s)	29.7 ± 4.4	29.1 ± 2.7	0.260
Hepaplastin test (%)	97.8 ± 20.3	95.4 ± 19.4	0.523
Fibrinogen (%)	237 ± 44	225 ± 45	0.069
Hepatitis C virus RNA (<850/≥850 KIU/mL)	130/139	70/61	0.394
Histological grade of			
Activity (A1/A2/A3)	138/107/24	62/55/14	0.714
Fibrosis (F1/F2/F3/F4)	135/74/57/3	58/40/27/6	0.131
Steatosis (0%/1–9%/10–29%/30%≦)	89/109/37/34	49/45/21/16	0.643
Hepatitis C virus RNA negative at week 12 (yes/no)	116/153	63/68	0.349

the level of GGT, with an optimal cutoff of 40 U/L, were then selected as the fourth level of split, low levels being the predictor of higher RVR/cEVR probability (60% vs. 35%).

All five factors selected as significant variables in the CART analysis were also significantly associated with RVR/cEVR by univariate analysis (Table 2). In addition, steatosis, LDL-C, age and GGT were also independently

Table 2 Factors associated with rapid or complete early virological response by univariate and multivariate logistic regression analysis

Parameter	Category	Univariate			Multivariate		
		Odds	95% CI	*P*-value	Odds	95% CI	*P*-value
Age (years)	<50 vs. ≥50	2.65	1.51–4.65	<0.001	2.03	1.04–3.97	0.039
Fibrosis stage	F1-2 vs. F3-4	2.47	1.31–4.66	0.005	1.77	0.85–3.68	0.120
Steatosis (%)	<30 vs. ≥30	4.11	1.64–10.29	0.003	2.88	1.07–7.79	0.037
Hepatitis C virus RNA (KIU/mL)	<850 vs. ≥850	1.97	1.21–3.22	0.007	1.93	1.09–3.43	0.025
Creatinine (mg/dL)	≥0.8 vs. <0.8	3.30	1.96–5.56	<0.001	3.54	1.88–6.67	<0.001
Hemoglobin (g/dL)	≥14.5 vs. <14.5	1.76	1.08–2.87	0.023	1.38	0.74–2.57	0.320
Hematcrit (%)	≥43 vs. <43	1.75	1.07–2.84	0.003			
Gamma-glutamyltransferase (IU/L)	<40 vs. ≥40	2.06	1.26–3.37	0.004	2.45	1.32–4.56	0.005
Low-density-lipid cholesterol (mg/dL)	≥100 vs. <100	2.71	1.61–4.55	<0.001	2.21	1.21–4.06	0.010
Blood sugar (mg/dL)	<120 vs. ≥120	2.00	1.02–3.95	0.045	1.42	0.64–3.13	0.390

CI, confidence interval.

© 2009 The Japan Society of Hepatology

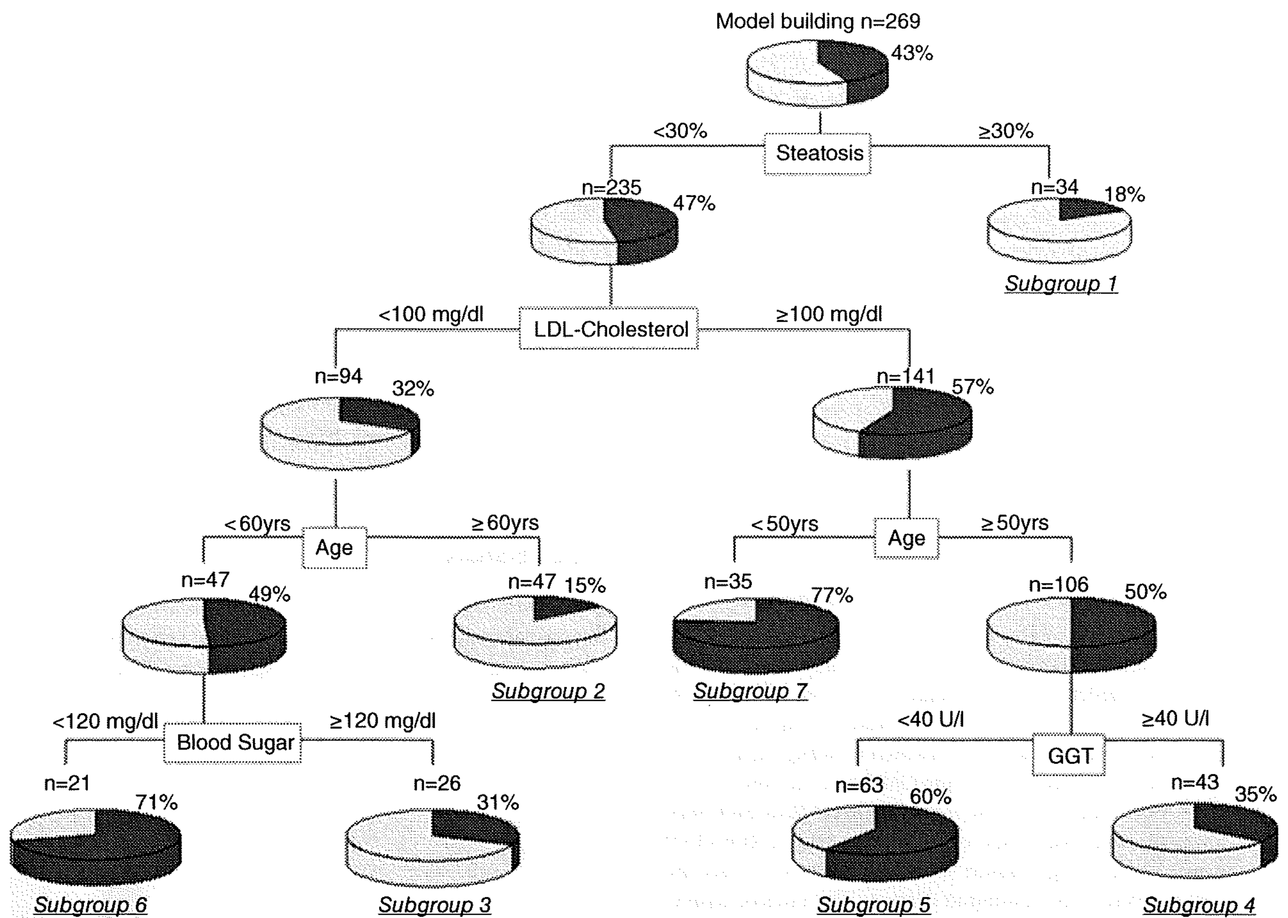


Figure 1 Classification and regression tree analysis. Boxes indicate the factors used for splitting and the cut-off value for the split. Pie charts indicate the rate of RVR/cEVR for each group of patients after splitting. Terminal subgroups of patients discriminated by the analysis are numbered from one to seven. GGT, gamma-glutamyltransferase; LDL, low-density-lipoprotein.

associated with RVR/cEVR by multivariate logistic regression analysis while blood sugar was not (Table 2). On the other hand, HCVRNA and creatinine which were significantly associated with RVR/cEVR by multivariate analysis were not selected as significant variables in CART analysis.

The probabilities of RVR/cEVR for the seven subgroups derived by this process were highly variable. The subgroup whose hepatic steatosis was less than 30%, serum LDL-C was 100 mg/dL or more and of an age less than 50 years (subgroup 7) showed the highest probability of RVR/cEVR (77%), while the subgroup whose hepatic steatosis more than 30% (subgroup 1) and the subgroup whose hepatic steatosis was less than 30% but serum LDL-C was less than 100 mg/dL and of an age greater than 60 years (subgroup 2) showed the lowest probability of RVR/cEVR (18% and 15%, respectively).

Validation of the CART analysis

The results of the CART analysis were validated with a validation dataset of 131 cases which is independent of the model building dataset. Each patient in the validation set was allocated to subgroups 1–7 using the flowchart form of the CART tree. The rates of RVR/cEVR were 20% for subgroups 1 and 2, 29% for subgroups 3, 38% for subgroup 4, 59% for subgroup 5, 71% for subgroup 6, and 85% for subgroups 7. The rates of RVR/cEVR for each subgroup of patients were closely correlated between the model building dataset and the validation dataset (Fig. 2).

© 2009 The Japan Society of Hepatology

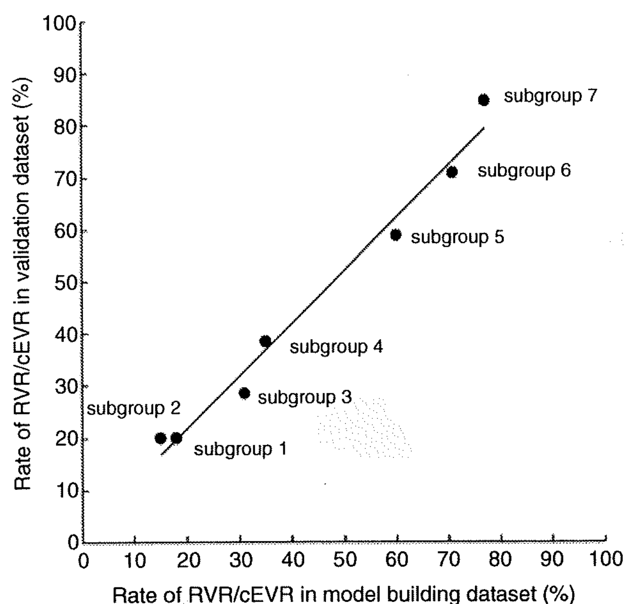


Figure 2 Validation of the classification and regression tree (CART) analysis: Subgroup stratified comparison of the rate of rapid or complete early virological response (RVR/cEVR) between the model building and validation datasets. Each patient in the validation set was allocated to subgroups 1–7 by following the flow-chart form of the CART tree and the rates of RVR/cEVR were calculated. The rate of RVR/cEVR in each subgroup was plotted. The *x*-axis represents the rate of RVR/cEVR in the model building datasets and the y-axis represents the rate of RVR/cEVR in the validation datasets. The rates of achieving RVR/cEVR in each subgroup of patients closely correlated between the model building dataset and the validation dataset ($r^2 = 0.987$).

Construction of 3 groups according to the probability of RVR/cEVR

If the seven subgroups were reconstructed into three groups according to their rate of RVR/cEVR, the rate of RVR/cEVR was 16% for low probability group (subgroup 1 and 2), 46% for intermediate probability group (subgroup 3, 4, and 5) and 75% for high probability group (subgroup 6 and 7; $P < 0.0001$).

Effect of adherence

Adherence of PEG-IFN and RBV was not included as a variable of analysis since the present study aimed to develop a pre-treatment model for the prediction of response. To analyze the possible effect of adherence on the result of CART analysis, three groups of patients divided by CART (low, intermediate and high probability group) were further stratified according to adherence of PEG-IFN and RBV. Poor adherence was defined as taking less than 80% planned dose of PEG-IFN or RBV at 12 weeks, and good adherence was defined as taking more than 80% planned dose of both PEG-IFN and RBV at 12 weeks. The result is shown in Figure 3. Among patients with good adherence, the rate of RVR/cEVR was 19% for low probability group, 52% for intermediate probability group and 77% for high probability group. Among poor adherence group, the rate of RVR/cEVR was 13% for low probability group, 41% for intermediate probability group and 73% for high probability group. Collectively, even after adjustment for adherence, 3 groups of patients divided by CART analysis still had low, intermediate and high probability of achieving RVR/cEVR, respectively.

DISCUSSION

IN THE PRESENT study, we performed the CART analysis and built a simple decision tree model for the pre-treatment prediction of response to PEG-IFN plus

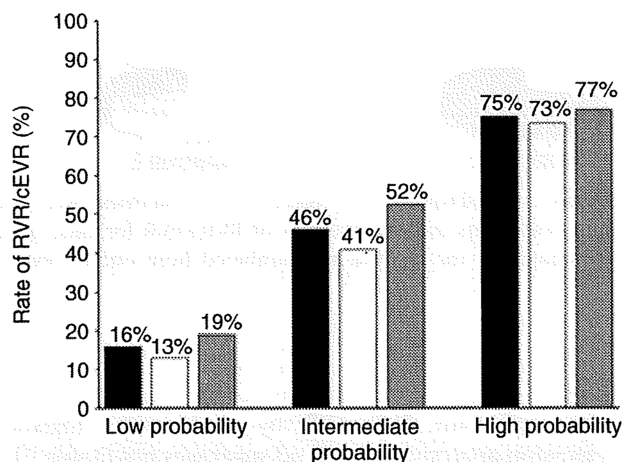


Figure 3 The rate of rapid or complete early virological response (RVR/cEVR) between the classification and regression tree (CART) groups stratified by adherence. The three groups of patients divided by CART (low, intermediate and high probability group) were further stratified according to adherence of peg-interferon (PEG-IFN) plus ribavirin (RBV). Black, white and gray boxes in the bar chart indicate total patients, patients with poor adherence (taking less than 80% planned dose of PEG-IFN or RBV at 12 weeks), and good adherence (taking more than 80% planned dose of both PEG-IFN and RBV at 12 weeks), respectively. Even after adjustment for adherence, 3 groups of patients divided by CART analysis still had low, intermediate and high probability of achieving RVR/cEVR, respectively.

© 2009 The Japan Society of Hepatology

RBV therapy. The analysis highlighted five host variables relevant to response: steatosis, LDL-C, age, blood sugar and GGT. Classification of patients based on these variables identified subgroups of patients with high probabilities of achieving RVR/cEVR among difficult to treat chronic hepatitis C patients. The reproducibility of the model was confirmed by the independent validation datasets. According to the result of the CART, patients were categorized into 3 groups: the rate of RVR/cEVR was 16% for low probability group, 46% for intermediate probability group and 75% for high probability group. The result of the CART analysis could be readily applicable to clinical practice because patients could be allocated to specific subgroups with a defined rate of response simply by following the flow-chart form. Although an early disappearance of serum HCV RNA is the prerequisite for achieving SVR, no reliable baseline predictors of response to PEG-IFN plus RBV therapy are established to date. Thus, this model may have the potential to support decisions in patient selection for PEG-IFN plus RBV therapy or to tailor treatment strategies for individual patients. Moreover, our result may provide a rationale for treating metabolic factors to improve the efficacy of antiviral therapy.

Among variables relevant to the prediction of RVR/cEVR, the grade of hepatic steatosis was selected as the variable of the first split. Previous studies suggested that steatosis induces resistance to IFN and RBV combination therapy[20,21] along with underlining metabolic factors such as insulin resistance or obesity.[21–24] In the present study, the grade of steatosis correlated positively with BMI and serum glucose level (data not shown) suggesting the etiologic role of metabolic factors. In addition, serum glucose level was selected as a predictor of RVR/cEVR at the fourth level of split. Serum GGT, which is associated with obesity,[25] insulin resistance[26] and response to IFN therapy,[27–30] was also selected as a predictor of RVR/cEVR at fourth level of splitting which may emphasize the importance of metabolic factors in therapeutic resistance. These findings raise the possibility that treatment of these metabolic factors may improve the virological response to the PEG-IFN plus RBV therapy. This hypothesis should be examined by a prospective study.

We and others have reported that steatosis, obesity and insulin resistance are associated with the progression of fibrosis,[17,31–33] which can interfere indirectly with the effect of IFN on hepatocytes. Other possible mechanisms of resistance by steatosis or metabolic factors include dysregulation of adipocytokines[34] or oxidative stress which may inhibit intracellular IFN signaling pathway.[35] Despite these findings, the precise mechanism of resistance is not established and further investigation is needed.

Another factor relevant in the prediction of RVR/cEVR was LDL-C. LDL-C was selected as the second factor for splitting by CART, and was an independent predictor of RVR/cEVR by logistic regression analysis. LDL-C recently has attracted attention as a novel predictor of response to IFN or PEG-IFN plus RBV.[30,36,37] Since *in vitro* study showed that LDL-C receptor acts as a receptor for HCV and LDL-C competitively inhibit the binding of HCV,[38] high level of serum LDL-C may inhibit HCV entry to hepatocytes and attenuate replication. LDL-C and its receptor may be a future therapeutic target.

Not all factors selected as significant variables in the CART analysis were also significantly associated with response by standard statistical analysis: blood sugar was associated with response by univariate analysis but not by multivariate logistic regression analysis On the other hand, HCVRNA and creatinine which were significantly associated with RVR/cEVR by multivariate analysis were not selected as significant variables in CART analysis. These differences may indicate both the unique feature and the limitations of the CART analysis. To note, blood sugar was significantly associated with RVR/cEVR within specialized subgroups of patients defined by the CART analysis: in subgroup of patients with steatosis <30%, LDL-C <100 mg/dL and younger than 60, which indicate the unique feature of the CART analysis that it could visualize significant predictors that specifically apply to selected patients. The limitation is that not all significant factors may be adopted in the decision tree since we applied the rule to stop CART procedure when the sample size was below 20. This rule was applied to avoid the generation of over-fit model which may lack universality. Therefore, it is possible that HCVRNA or creatinine may become a significant variable in the CART analysis if larger number of patients were included in the analysis. Stage of fibrosis was significantly associated with response to therapy by univariate analysis but not by multivariate analysis and not selected as a significant variable in the CART analysis. The possible reason is that advanced fibrosis is associated with older age as a confounding factor.

CART analyses are gaining acceptance in medical research in addition to biomedical field. Recent publications include the prediction of aggressive prostate cancer,[8] diabetic vascular complications,[19] prognosis of melanoma,[7,39] response to preoperative radiotherapy for rectal tumor,[9] prognostic groups in colorectal carcinoma,[12] and outcome after liver failure.[11] An advantage

© 2009 The Japan Society of Hepatology

of CART over traditional regression models is that it can identify prognostic subgroups that are useful in clinical practice. Because the results of CART analysis are presented as a decision tree, which is intuitive, they can be readily interpreted by medical professionals without any specific knowledge of statistics. The most important consideration is that five variables used in the decision tree were clinical parameters that are readily available by the usual work-up of patients before therapy. Especially, glucose, GGT and LDL-C are simple biochemical markers that are easily measured at a low cost. Using this model, we can rapidly develop an estimate of the response before treatment, which may facilitate clinical decision making.

In conclusion, we built a pre-treatment model for the prediction of virological response in PEG-IFN plus RBV therapy. Because this decision tree model was made up of simple host factors such as steatosis, LDL-C, age, blood sugar and GGT, it can be easily applied to clinical practice. This model may have the potential to support decisions in patient selection for PEG-IFN plus RBV therapy based on the possibility of response against a potential risk of adverse events or costs, and may provide a rationale for treating metabolic factors to improve the efficacy of antiviral therapy.

ACKNOWLEDGEMENTS

THIS STUDY WAS supported by a grant-in-aid from Ministry of Health, Labor and Welfare, Japan. There exist no conflicts of interest.

REFERENCES

1 Strader DB, Wright T, Thomas DL, Seeff LB. Diagnosis, management, and treatment of hepatitis C. *Hepatology* 2004; 39: 1147–71.
2 Fried MW, Shiffman ML, Reddy KR *et al.* Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med* 2002; 347: 975–82.
3 Manns MP, McHutchison JG, Gordon SC *et al.* Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. *Lancet* 2001; 358: 958–65.
4 Davis GL, Wong JB, McHutchison JG, Manns MP, Harvey J, Albrecht J. Early virologic response to treatment with peginterferon alfa-2b plus ribavirin in patients with chronic hepatitis C. *Hepatology* 2003; 38: 645–52.
5 Lee SS, Ferenci P. Optimizing outcomes in patients with hepatitis C virus genotype 1 or 4. *Antivir Ther* 2008; 13 (Suppl 1): 9–16.
6 Breiman LJH, Friedman RA, Olshen CJ, Stone CM. *Classification and Regression Trees.* Calif: Wadsworth, 1980.
7 Averbook BJ, Fu P, Rao JS, Mansour EG. A long-term analysis of 1018 patients with melanoma by classic Cox regression and tree-structured survival analysis at a major referral center: Implications on the future of cancer staging. *Surg* 2002; 132: 589–602.
8 Garzotto M, Beer TM, Hudson RG *et al.* Improved detection of prostate cancer using classification and regression tree analysis. *J Clin Oncol* 2005; 23: 4322–9.
9 Zlobec I, Steele R, Nigam N, Compton CC. A predictive model of rectal tumor response to preoperative radiotherapy using classification and regression tree methods. *Clin Cancer Res* 2005; 11: 5440–3.
10 Jin H, Lu Y, Harris ST *et al.* Classification algorithms for hip fracture prediction based on recursive partitioning methods. *Med Decis Making* 2004; 24: 386–98.
11 Baquerizo A, Anselmo D, Shackleton C *et al.* Phosphorus ans an early predictive factor in patients with acute liver failure. *Transplantation* 2003; 75: 2007–14.
12 Valera VA, Walter BA, Yokoyama N *et al.* Prognostic groups in colorectal carcinoma patients based on tumor cell proliferation and classification and regression tree (CART) survival analysis. *Ann Surg Oncol* 2007; 14: 34–40.
13 Martin MA, Meyricke R, O'Neill T, Roberts S. Mastectomy or breast conserving surgery? Factors affecting type of surgical treatment for breast cancer – a classification tree approach. *BMC Cancer* 2006; 6: 98.
14 LeBlanc M, Crowley J. A review of tree-based prognostic models. *Cancer Treat Res* 1995; 75: 113–24.
15 Costanza MC, Paccaud F. Binary classification of dyslipidemia from the waist-to-hip ratio and body mass index: a comparison of linear, logistic, and CART models. *BMC Med Res Methodol* 2004; 4: 7.
16 Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology* 1996; 24: 289–93.
17 Kurosaki M, Matsunaga K, Hirayama I *et al.* The presence of steatosis and elevation of alanine aminotransferase levels are associated with fibrosis progression in chronic hepatitis C with non-response to interferon therapy. *J Hepatol* 2008; 48: 736–42.
18 Segal MR, Bloch DA. A comparison of estimated proportional hazards models and regression trees. *Stat Med* 1989; 8: 539–50.
19 Miyaki K, Takei I, Watanabe K, Nakashima H, Omae K. Novel statistical classification model of type 2 diabetes mellitus patients for tailor-made prevention using data mining algorithm. *J Epidemiol* 2002; 12: 243–8.
20 Akuta N, Suzuki F, Tsubota A *et al.* Efficacy of interferon monotherapy to 394 consecutive naive cases infected with hepatitis C virus genotype 2a in Japan: therapy efficacy as consequence of tripartite interaction of viral, host and interferon treatment-related factors. *J Hepatol* 2002; 37: 831–6.

© 2009 The Japan Society of Hepatology

21 Poynard T, Ratziu V, McHutchison J *et al.* Effect of treatment with peginterferon or interferon alfa-2b and ribavirin on steatosis in patients infected with hepatitis C. *Hepatology* 2003; 38: 75–85.
22 Bressler BL, Guindi M, Tomlinson G, Heathcote J. High body mass index is an independent risk factor for nonresponse to antiviral treatment in chronic hepatitis C. *Hepatology* 2003; 38: 639–44.
23 Romero-Gomez M, Del Mar Viloria M, Andrade RJ *et al.* Insulin resistance impairs sustained response rate to peginterferon plus ribavirin in chronic hepatitis C patients. *Gastroenterology* 2005; 128: 636–41.
24 Konishi I, Horiike N, Hiasa Y *et al.* Diabetes mellitus reduces the therapeutic effectiveness of interferon-alpha2b plus ribavirin therapy in patients with chronic hepatitis C. *Hepatol Res* 2007; 37: 331–6.
25 Marchesini G, Avagnina S, Barantani EG *et al.* Aminotransferase and gamma-glutamyltranspeptidase levels in obesity are associated with insulin resistance and the metabolic syndrome. *J Endocrinol Invest* 2005; 28: 333–9.
26 Fraser A, Ebrahim S, Smith GD, Lawlor DA. A comparison of associations of alanine aminotransferase and gamma-glutamyltransferase with fasting glucose, fasting insulin, and glycated hemoglobin in women with and without diabetes. *Hepatology* 2007; 46: 158–65.
27 Mazzella G, Salzetta A, Casanova S *et al.* Treatment of chronic sporadic-type non-A, non-B hepatitis with lymphoblastoid interferon: gamma GT levels predictive for response. *Dig Dis Sci* 1994; 39: 866–70.
28 Villela-Nogueira CA, Perez RM, de Segadas Soares JA, Coelho HS. Gamma-glutamyl transferase (GGT) as an independent predictive factor of sustained virologic response in patients with hepatitis C treated with interferon-alpha and ribavirin. *J Clin Gastroenterol* 2005; 39: 728–30.
29 Berg T, Sarrazin C, Herrmann E *et al.* Prediction of treatment outcome in patients with chronic hepatitis C: significance of baseline parameters and viral dynamics during therapy. *Hepatology* 2003; 37: 600–9.
30 Akuta N, Suzuki F, Kawamura Y *et al.* Predictive factors of early and sustained responses to peginterferon plus ribavirin combination therapy in Japanese patients infected with hepatitis C virus genotype 1b: amino acid substitutions in the core region and low-density lipoprotein cholesterol levels. *J Hepatol* 2007; 46: 403–10.
31 Adinolfi LE, Gambardella M, Andreana A, Tripodi MF, Utili R, Ruggiero G. Steatosis accelerates the progression of liver damage of chronic hepatitis C patients and correlates with specific HCV genotype and visceral obesity. *Hepatology* 2001; 33: 1358–64.
32 Ortiz V, Berenguer M, Rayon JM, Carrasco D, Berenguer J. Contribution of obesity to hepatitis C-related fibrosis progression. *Am J Gastroenterol* 2002; 97: 2408–14.
33 Muzzi A, Leandro G, Rubbia-Brandt L *et al.* Insulin resistance is associated with liver fibrosis in non-diabetic chronic hepatitis C patients. *J Hepatol* 2005; 42: 41–6.
34 Charlton MR, Pockros PJ, Harrison SA. Impact of obesity on treatment of chronic hepatitis C. *Hepatology* 2006; 43: 1177–86.
35 Di Bona D, Cippitelli M, Fionda C *et al.* Oxidative stress inhibits IFN-alpha-induced antiviral gene expression by blocking the JAK-STAT pathway. *J Hepatol* 2006; 45: 271–9.
36 Minuk GY, Weinstein S, Kaita KD. Serum cholesterol and low-density lipoprotein cholesterol levels as predictors of response to interferon therapy for chronic hepatitis C. *Ann Intern Med* 2000; 132: 761–2.
37 Gopal K, Johnson TC, Gopal S *et al.* Correlation between beta-lipoprotein levels and outcome of hepatitis C treatment. *Hepatology* 2006; 44: 335–40.
38 Agnello V, Abel G, Elfahal M, Knight GB, Zhang QX. Hepatitis C virus and other flaviviridae viruses enter cells via low density lipoprotein receptor. *Proc Natl Acad Sci USA* 1999; 96: 12766–71.
39 Leiter U, Buettner PG, Eigentler TK, Garbe C. Prognostic factors of thin cutaneous melanoma: an analysis of the central malignant melanoma registry of the german dermatological society. *J Clin Oncol* 2004; 22: 3660–7.

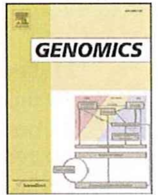
© 2009 The Japan Society of Hepatology

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.

© 2010 Published by Elsevier Inc.

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

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

ARTICLE IN PRESS

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 (ALB)* 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 (HP)* (NL: 631, NT: 329, T: 57) and *metallothionein 1G (MTIG)* (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 (MT2A)* (NL: 1027, NT: 872, T: 19), *metallothionein 1X (MT1X)* (NL: 547, NT: 644, T: 11), and *metallothionein 1E (MT1E)* (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 (RPS29)* (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 1A (XAGE 1A)*, *retinol dehydrogenase 11 (RDH11)*, *hydroxysteroid (17-beta) dehydrogenase 14 (HSD17B14) 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 (F13B)* 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 (ADCY1)* 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

ARTICLE IN PRESS

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 (ALB)
2484	2146	2042	0.864	0.822	Apolipoprotein C-I (APOC1)
1955	1603	1079	0.820	0.552	Apolipoprotein A-II (APOA2)
1653	1050	828	0.635	0.501	Apolipoprotein A-I (APOA1)
1252	1908	1203	1.524	0.961	Transthyretin (prealbumin, amyloidosis type I) (TTR)
1233	724	220	0.587	0.178	Serpin peptidase inhibitor, clade A, member 1 (SERPINA1)
1027	872	19	0.849	0.019	Metallothionein 2A (MT2A)
755	1144	762	1.515	1.009	Ferritin, light polypeptide (FTL)
713	632	680	0.886	0.954	Alpha-1-microglobulin/bikunin precursor (AMBP)
635	524	1336	0.825	2.104	Apolipoprotein E (APOE)
631	329	57	0.521	0.090	Haptoglobin (HP)
600	228	212	0.380	0.353	Fibrinogen gamma chain (FGG)
549	395	302	0.719	0.550	Apolipoprotein C-III (APOC3)
547	644	11	1.177	0.020	Metallothionein 1X (MT1X)
479	257	290	0.537	0.605	Tumor protein, translationally-controlled 1 (TPT1)
463	217	53	0.469	0.114	Serpin peptidase inhibitor, clade A, member 3 (SERPINA3)
393	204	206	0.519	0.524	Ribosomal protein L26 (RPL26)
392	169	2	0.431	0.005	Metallothionein 1G (MT1G)
372	1011	1768	2.718	4.753	Ribosomal protein S29 (RPS29)
306	163	223	0.533	0.729	Ribosomal protein S27 (RPS27)
279	135	159	0.484	0.570	Ribosomal protein S16 (RPS16)
275	340	2	1.236	0.007	Metallothionein 1E (MT1E)
269	170	246	0.632	0.914	Ribosomal protein S23 (RPS23)
260	142	92	0.546	0.354	Fibrinogen beta chain (FGB)
260	200	195	0.769	0.750	Aldolase B, fructose-bisphosphate (ALDOB)
255	228	286	0.894	1.122	Ribosomal protein S12 (RPS12)
248	162	198	0.653	0.798	Ribosomal protein S14 (RPS14)
246	175	70	0.711	0.285	Interferon induced transmembrane protein 3 (IFITM3)
239	198	273	0.828	1.142	Ribosomal protein L31(RPL31)
229	264	0	1.153	0.004	Hepcidin antimicrobial peptide (HAMP)
228	149	156	0.654	0.684	Ribosomal protein S20 (RPS20)
222	191	117	0.860	0.527	Ubiquitin B (UBB)
216	218	352	1.009	1.630	Ribosomal protein L41 (RPL41)
210	150	155	0.714	0.738	Ribosomal protein, large, P1(RPLP1)
201	110	90	0.547	0.448	Ribosomal protein, large, P2 (RPLP2)
198	102	64	0.515	0.323	Fibrinogen alpha chain (FGA)
196	143	408	0.730	2.082	Ribosomal protein L37 (RPL37)
192	123	56	0.641	0.292	Ribosomal protein L37a (RPL37A)
191	208	346	1.089	1.812	Ribosomal protein L30 (RPL30)
174	109	76	0.626	0.437	Ribosomal protein L35 (RPL35)
169	208	3	1.231	0.018	Cytochrome P450, family 2, subfamily E, polypeptide 1 (CYP2E1)
167	105	300	0.629	1.796	Apolipoprotein H (beta-2-glycoprotein I) (APOH)
162	106	33	0.654	0.204	Serum amyloid A4, constitutive (SAA4)
159	85	157	0.535	0.987	Ribosomal protein L34 (RPL34)
159	113	229	0.711	1.440	Transferrin (TF)
155	84	135	0.542	0.871	Ribosomal protein S11 (RPS11)
152	125	101	0.822	0.664	Ribosomal protein S13 (RPS13)
147	84	1	0.571	0.007	Nicotinamide *N*-methyltransferase (NNMT)
147	180	35	1.224	0.238	Hemopexin (HPX)
146	89	121	0.610	0.829	Alpha-2-HS-glycoprotein (AHSG)

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*,

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) (PAGE2)
18	10	1.8	Lectin, galactoside-binding, soluble, 4 (LGALS4)
16	3	5.33	Choline phosphotransferase 1 (CHPT1)
14	2	7	X antien family, member 1A (XAGE1A)
14	2	7	Dehydrogenase/reductase (SDR family) member 4 (DHRS4)
14	2	7	Sterol-C5-desaturase-like (SC5DL)
13	0.5	26	Coagulation factor XIII, B polypeptide (F13B)
13	2.33	5.58	Retinol dehydrogenase 11 (all-trans and 9-cis) (RDH11)
13	0.5	26	Transmembrane protein 14A (TMEM14A)
12	1.33	9.02	Dual specificity phosphatase 23 (DUSP23)
Down-regulated gene			
0.00436	0.0137	0.318	Hepcidin antimicrobial peptide (HAMP)
0.0051	ND		Metallothionein 1G (MT1G)
0.0068	0.04	0.17	Nicotinamide *N*-methyltransferase (NNMT)
0.00727	ND		Metallothionein 1E (functional) (MT1E)
0.0098	0.0526	0.186	C-reactive protein, pentraxin-related (CRP)
0.0145	ND		Metallothionein 1 M (MT1M)
0.0152	ND		Phospholipase A2, group IIA (platelets, synovial fluid) (PLA2G2A)
0.0178	0.111	0.16	Cytochrome P450, family 2, subfamily E, polypeptide 1 (CYP2E1)
0.0185	0.192	0.096	Metallothionein 2A (MT2A)
0.0201	ND		Metallothionein 1X (MT1X)

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 (LILRB1)
ND	12		Fibroblast growth factor 5 (FGF5)
1	11	0.909	Adenosine deaminase, tRNA-specific 1 (ADAT1)
5	11	0.454	px19-like protein (PRELID1)
4.4	11	0.4	Anaphase promoting complex subunit 11 (ANAPC11)
ND	10.3		Chromosome 21 open reading frame 77 (C21orf77)
ND	10		von Willebrand factor (VWF)
2.333	10	0.233	ATX1 antioxidant protein 1 homolog (yeast) (ATOX1)
18	10	1.8	Lectin, galactoside-binding, soluble, 4 (LGALS4)
ND	9.5		Solute carrier family 26 (sulfate transporter), member 2 (SLC26A2)
Down-regulated gene			
0.5	0.012	41.7	ELL associated factor 1 (EAF1)
0.5	0.0137	36.5	TGF beta-inducible nuclear protein 1 (NSA2)
0.000436	0.0137	0.032	Hepcidin antimicrobial peptide (HAMP)
1	0.0179	55.9	Basic, immunoglobulin-like variable motif containing (BIVM)
ND	0.0182		DNA fragmentation factor, 45 kDa, alpha polypeptide (DFFA)
1	0.0185	54.1	GRIP1 associated protein 1 (GRIPAP1)
ND	0.0189		Nuclear factor of activated T-cells 5, tonicity-responsive (NFAT5)
1	0.0204	49	Adenylate cyclase 1 (ADCY1)
0.333	0.0312	10.7	Dihydroorotate dehydrogenase (DHODH)
0.738	0.0312	23.7	Ribosomal protein, large, P1 (RPLP1)

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

ARTICLE IN PRESS

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, 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 *TMEM176B* 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.

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	Transmembrane protein 176B (TMEM176B)

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 (ALB)* 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 (ALDOB), antitrypsin (SERPINA1)*, and *haptoglobin (HP)* 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 (LGALS4)* was up-regulated and *HAMP, 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 *XAGE1A* 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

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

ARTICLE IN PRESS

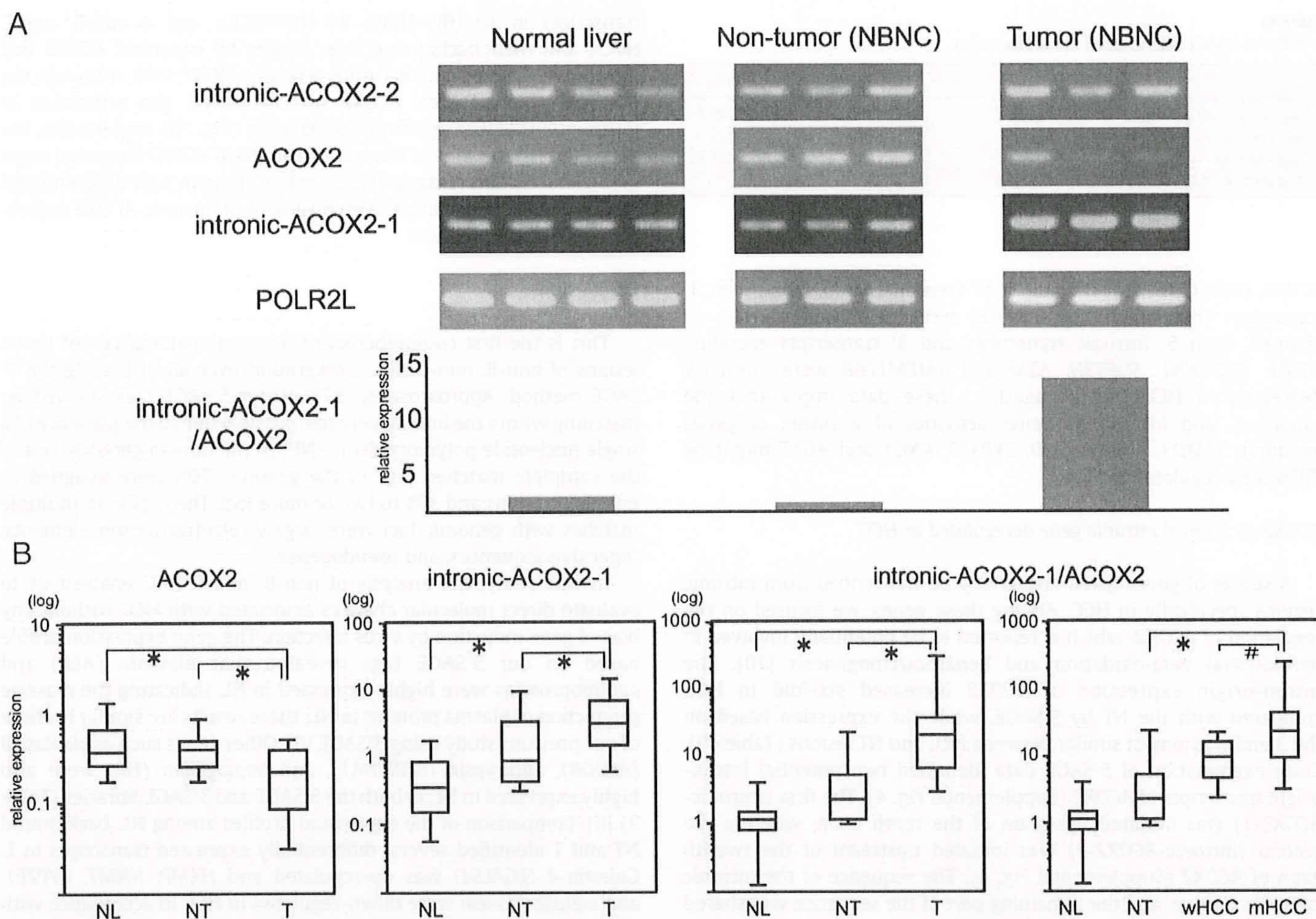


Fig. 1. (A) RT-PCR results of *ACOX2* and *ACOX2* intronic RNAs in independent NL, NT (non-B, non-C), and T (non-B, non-C) samples. RT-PCR was performed in triplicate for each sample-primer set from cDNA. The PCR products were semi-quantitatively analyzed with ImageJ software and calculated as levels 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. *$P<0.01$, #$P<0.05$.

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 alpha (PPARA). 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 (AMACR), 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.

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

ARTICLE IN PRESS

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′-TTCATAAAGTTGTGAGCA-GAGGAAA-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

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

ARTICLE IN PRESS

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.

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

Differential MicroRNA Expression Between Hepatitis B and Hepatitis C Leading Disease Progression to Hepatocellular Carcinoma

Shunsuke Ura,[1] Masao Honda,[1,2] Taro Yamashita,[1] Teruyuki Ueda,[1] Hajime Takatori,[1] Ryuhei Nishino,[1] Hajime Sunakozaka,[1] Yoshio Sakai,[1] Katsuhisa Horimoto,[3] and Shuichi Kaneko[1]

MicroRNA (miRNA) plays an important role in the pathology of various diseases, including infection and cancer. Using real-time polymerase chain reaction, we measured the expression of 188 miRNAs in liver tissues obtained from 12 patients with hepatitis B virus (HBV)-related hepatocellular carcinoma (HCC) and 14 patients with hepatitis C virus (HCV)-related HCC, including background liver tissues and normal liver tissues obtained from nine patients. Global gene expression in the same tissues was analyzed via complementary DNA microarray to examine whether the differentially expressed miRNAs could regulate their target genes. Detailed analysis of the differentially expressed miRNA revealed two types of miRNA, one associated with HBV and HCV infections (n = 19), the other with the stage of liver disease (n = 31). Pathway analysis of targeted genes using infection-associated miRNAs revealed that the pathways related to cell death, DNA damage, recombination, and signal transduction were activated in HBV-infected liver, and those related to immune response, antigen presentation, cell cycle, proteasome, and lipid metabolism were activated in HCV-infected liver. The differences in the expression of infection-associated miRNAs in the liver correlated significantly with those observed in Huh7.5 cells in which infectious HBV or HCV clones replicated. Out of the 31 miRNAs associated with disease state, 17 were down-regulated in HCC, which up-regulated cancer-associated pathways such as cell cycle, adhesion, proteolysis, transcription, and translation; 6 miRNAs were up-regulated in HCC, which down-regulated anti-tumor immune response. *Conclusion:* miRNAs are important mediators of HBV and HCV infection as well as liver disease progression, and therefore could be potential therapeutic target molecules. (HEPATOLOGY 2009;49:1098-1112.)

MicroRNA (miRNA) is an endogenous, small, single-strand, noncording RNA consisting of 20 to 25 bases and regulates gene expression of various cell types. It plays an important role in various biological processes, including organ development and differentiation as well as cellular death and proliferation, and is also involved in various diseases such as infection and cancer.[1-3]

miRNAs are produced as follows. A primary miRNA with a hairpin loop structure is cleaved into a precursor miRNA and transported out of the nuclei with a carrier protein (Exportin-5). The precursor miRNA is then processed by Dicer and converted into an active single-strand RNA in the cytoplasm. The miRNA binds to a target messenger RNA in a sequence-dependent manner and induces degradation of the target messenger RNA and translational inhibition. One miRNA regulates the expression of multiple target genes; bioinformatics analyses have suggested that the expression of more than 30% of human genes is regulated by miRNAs.[4-7]

Abbreviations: cDNA, complementary DNA; CH, chronic hepatitis; CH-B, chronic hepatitis B; CH-C, chronic hepatitis C; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HCC-B, hepatitis B-related hepatocellular carcinoma; HCC-C, hepatitis C-related hepatocellular carcinoma; HCV, hepatitis C virus; miRNA, microRNA; RTD-PCR, real-time detection polymerase chain reaction; SVM, support vector machine.

From the Departments of [1]Gastroenterology and [2]Advanced Medical Technology, Kanazawa University Graduate School of Medicine, Kanazawa, Japan; and the [3]Biological Network Team, Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, 2-42 Aomi, koto-ku, Tokyo 135-0064, Japan.

Received July 3, 2008; accepted November 15, 2008.

Address reprint requests to: Masao Honda, M.D., Ph.D., Department of Gastroenterology, Graduate School of Medicine, Kanazawa University, Takara-Machi 13-1, Kanazawa 920-8641, Japan. E-mail: mhonda@m-kanazawa.jp; fax: (81)-76-234-4250.

Copyright © 2009 by the American Association for the Study of Liver Diseases.

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI 10.1002/hep.22749
Potential conflict of interest: Nothing to report.

Additional Supporting Information may be found in the online version of this article.

Table 1. Characteristics of Patients Used for Analysis of miRNA and Microarray Samples

Patient No.	Virus	Age	Sex	ALT	Histology of Activity	Background Liver Fibrosis	Histological Grade of HCC	Tumor Size (mm)	TNM Staging	HCV-RNA (KIU/mL)	HBV-DNA (LEG/mL)
1	HBV	57	M	16	2	4	Moderate	20	II	–	3.4
2	HBV	51	M	57	1	2	Moderate	48	II	–	< 2.6
3	HBV	61	M	17	1	4	Well	16	II	–	< 3.7
4	HBV	47	M	19	1	4	Moderate	15	I	–	< 3.7
5	HBV	72	M	19	1	1	Well	25	II	–	NA
6	HBV	73	M	62	1	3	Moderate	45	III	–	5.7
7	HBV	42	M	36	1	4	Moderate	18	I	–	< 3.7
8	HBV	63	M	13	1	2	Moderate	15	I	–	2.8
9	HBV	68	F	54	1	2	Well	56	II	–	4.1
10	HBV	70	M	13	0	2	Well	40	II	–	< 3.7
11	HBV	58	M	29	1	4	Moderate	35	IVA*	–	3.3
12	HBV	72	M	22	1	4	Moderate	18	I	–	6
13	HCV	66	F	33	2	4	Well	25	II	423	–
14	HCV	67	M	89	1	4	Well	30	II	> 850	–
15	HCV	64	M	31	1	4	Moderate	75	III	< 5 (+)	–
16	HCV	68	M	30	0	4	Well	23	II	> 850	–
17	HCV	46	M	98	2	3	Moderate	20	I	> 850	–
18	HCV	68	F	32	2	4	Moderate	25	III	< 5 (+)	–
19	HCV	66	F	46	2	4	Well	25	II	> 850	–
20	HCV	47	M	246	1	3	Moderate	20	I	262	–
21	HCV	75	M	27	1	3	Moderate	19	II	85.1	–
22	HCV	77	M	21	0	1	Moderate	20	II	< 5 (–)	–
23	HCV	66	M	46	2	2	Well	60	II	50.3	–
24	HCV	65	M	89	1	1	Poorly	25	III	850	–
25	HCV	53	M	54	0	1	Moderate	28	II	< 5 (–)	–
26	HCV	75	F	212	1	4	Well	19	I	580	–
27	–	51	F	18	0	0	–	–	–	–	–
28	–	78	F	13	0	0	–	–	–	–	–
29	–	75	M	20	0	0	–	–	–	–	–
30	–	34	M	12	0	0	–	–	–	–	–
31	–	64	M	30	0	0	–	–	–	–	–
32	–	78	M	9	0	0	–	–	–	–	–
33	–	53	M	19	0	0	–	–	–	–	–
34	–	64	F	12	0	0	–	–	–	–	–
35	–	60	F	20	0	0	–	–	–	–	–

HCV RNA was assayed via Amplicor Monitor Test (KIU/mL); HBV DNA was assayed via transcription-mediated amplification (LEG/mL).

Abbreviations: ALT, alanine aminotransferase; F, female; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HCV, hepatitis C virus; M, male; TNM, tumor-node–metastasis.

*Vascular invasion (+).

Infection of the human liver with hepatitis B virus (HBV) and hepatitis C virus (HCV) induces the development of chronic hepatitis (CH), cirrhosis, and in some instances hepatocellular carcinoma (HCC).[8] The virological features of these two distinct viruses are completely different; however, the viruses infect the liver and cause CH, which is not distinguished by histological examination or clinical manifestations. We previously reported that gene expression profiles in chronic hepatitis B (CH-B) and chronic hepatitis C (CH-C) are different. Proapoptotic and DNA repair responses were predominant in CH-B, and inflammatory and antiapoptotic phenotypes were predominant in CH-C. However, factors inducing these differences in gene expression remain to be elucidated.[9,10]

We examined miRNA expression in liver tissue with HBV-related liver disease (CH-B and HCC-B) and HCV-related liver disease (CH-C and HCC-C) and in normal liver tissue via real-time detection polymerase chain reaction (RTD-PCR). We also performed global analysis of messenger RNA expression in these tissues using complementary DNA (cDNA) microarray. These analyses allowed us to find characteristic miRNAs associated with HBV or HCV infection as well as the progression of liver disease.

Patients and Methods

Patients. The study subjects included 12 patients with CH-B complicated by HCC and 14 patients with

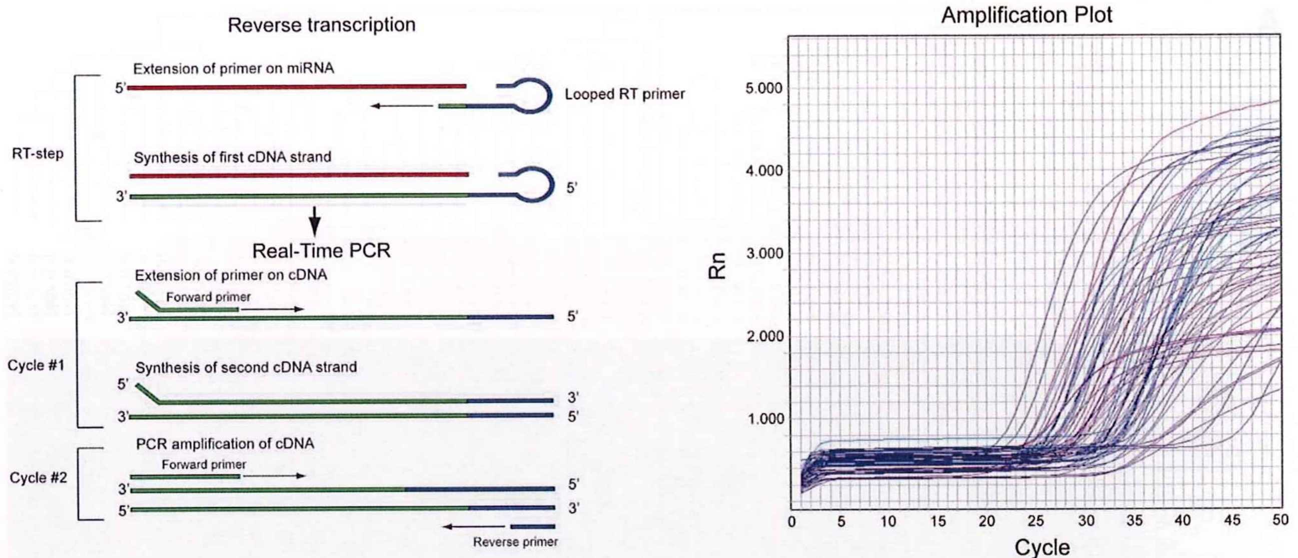


Fig. 1. (A) miRNA-specific RTD-PCR using sheet hairpin primers. (B) miRNA amplification curves by RTD-PCR.

CH-C complicated by HCC. Gene expression analysis was approved by the ethics committee of the Graduate School of Medicine, Kanazawa University Hospital, Japan, between 1999 and 2004. In addition, nine normal liver tissue samples obtained during surgery for metastatic liver cancer were used as control samples. Surgically removed liver tissues were stored in liquid nitrogen until analysis. Histological classification of HCC and histological evaluation of hepatitis in noncancerous regions for each patient are shown in Table 1. HCV viremia in two patients with CH-C was persistently cleared by interferon therapy before HCC development. There were no significant differences in the histological findings of HCC and noncancerous regions, as well as in sex, age, and hepatic function between the HBV and HCV infection groups.

Quantitative RTD-PCR. Approximately 1 mg of each liver tissue sample stored in liquid nitrogen was ground with a homogenizer while still frozen, and total RNA containing miRNA was isolated according to the protocol of the mirVana miRNA Isolation kit (Ambion, Austin, TX) and stored at −80°C until analysis. miRNA expression levels were quantitated using the TaqMan MicroRNA Assays Human Panel Early Access kit (Applied Biosystems, Foster City, CA). cDNA was prepared via reverse transcription using 10 ng each of the isolated total RNA and 3 μL each of the reverse transcription primers with specific loop structures. Reverse transcription was performed using the TaqMan MicroRNA Reverse Transcription kit (Applied Biosystems) according to the manufacturer's protocol. Then, a mixture of 6.67 μL of nuclease-free water, 10 μL of TaqMan 2 × Universal PCR Master Mix (No AmpErase UNG; Applied Biosystems), and 2 μL of TaqMan MicroRNA Assay Mix, which was included in the kit, was prepared for each sample on a 384-well plate; 1.33 μL of the reverse transcription product was added to the mixture, and amplification reaction was performed on an ABI PRISM 7900HT (Applied Biosystems). Expression levels of 188 miRNAs in each sample were quantitated.

Analysis of RTD-PCR Data. The measured 188 miRNAs included RNU6B, which is commonly used as a control for miRNA. β-Actin and glyceraldehyde 3-phosphate dehydrogenase were also measured simultaneously for correcting RNA amount. The mean Ct values and standard deviations of each miRNA were calculated from expression data of all patients obtained by RTD-PCR. miRNA with the lowest expression variation was used as the internal control. Ct values of each miRNA were then corrected by the Ct value of the internal control to yield −ΔCt values defined as relative miRNA expression levels and used for analyses. Statistical analyses and hierarchical cluster analyses of expression data were performed using BRB ArrayTools (http://linus.nci.nih.gov/BRB-ArrayTools.html). Relative miRNA expression levels were further normalized using the median over the all patients so that the normalized expression levels of each patient have a median log ratio of 0. A class prediction method was used for classifying two patient groups based on the supervised learning method, and a binary tree classification method was used for classifying three or more patient groups with a statistical algorithm of the support vector machine (SVM). Class prediction was performed using SVM incorporating genes differentially expressed at a univariate parametric significance level of $P = 0.01$. The prediction rate was estimated via cross-validation and the bootstrap method for small sample data.[11] (It is worth

A

B

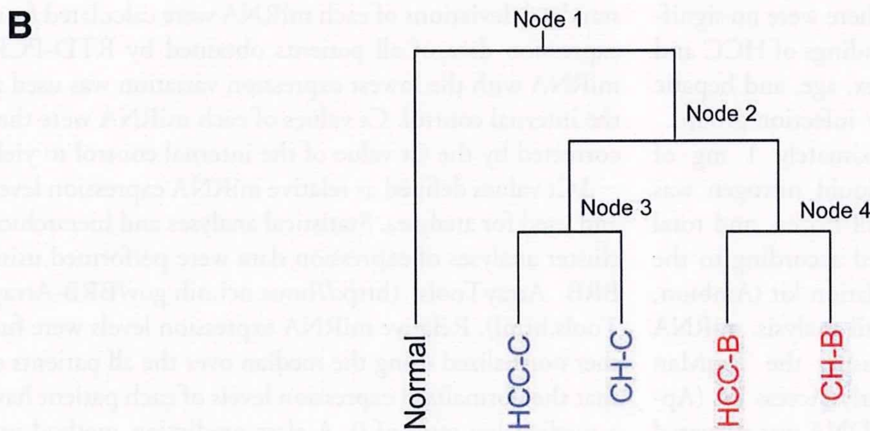
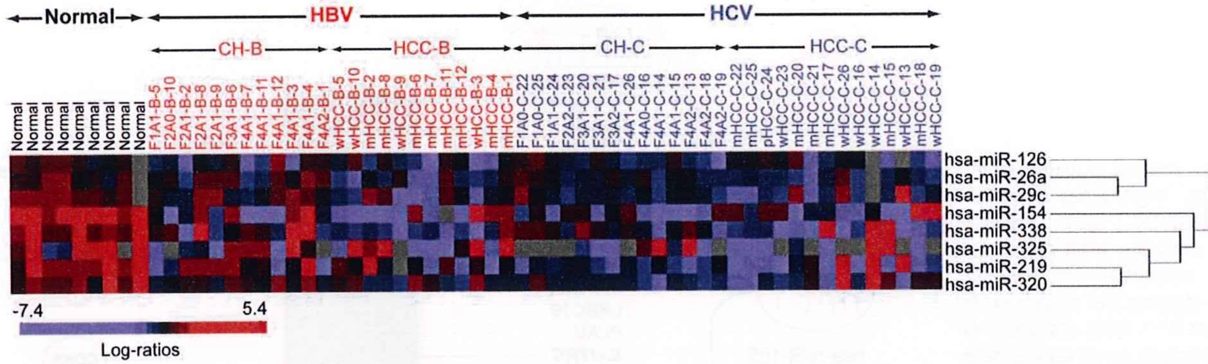
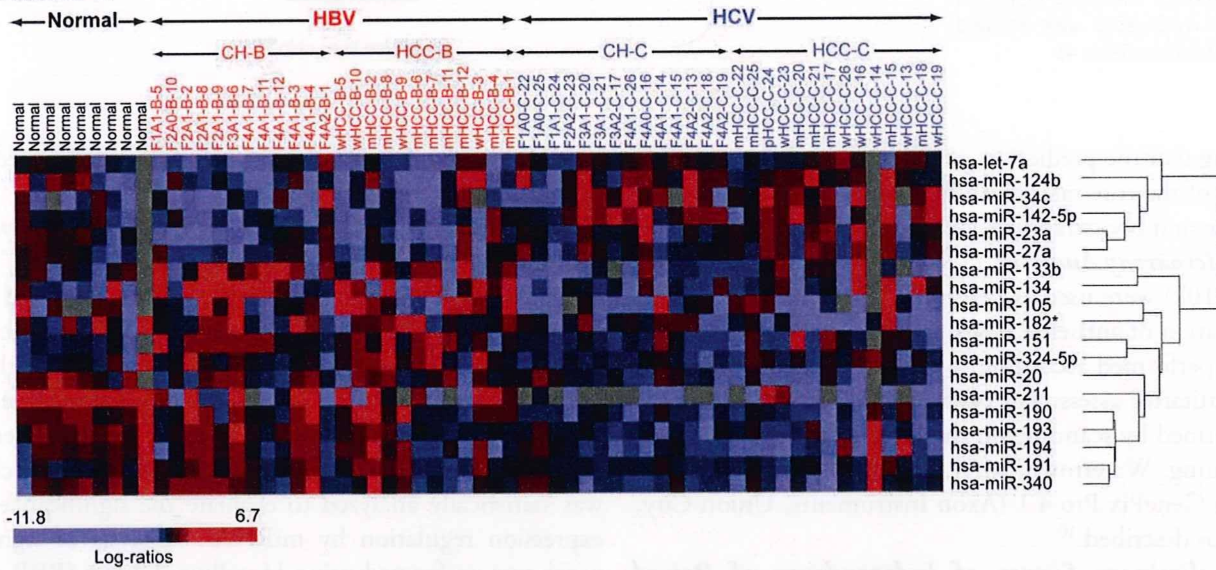


Fig. 2. (A) Hierarchical cluster analysis using total miRNA. Chronic hepatitis is indicated by histological stage and grade (F, fibrosis; A, activity) and type of infecting virus (B or C). HCC is indicated by histological grade (w, well differentiated; m, moderately differentiated; p, poorly differentiated) and type of infecting virus (B or C), with the patient number added at the end. (B) Relationship between five classes divided by binary tree classification. Expression profiles were first classified into normal liver and non-normal liver groups (node 1), then into HBV and HCV groups (node 2). The HBV group was further divided into HCC-B and CH-B (node 3), and the HCV group into HCC-C and CH-C (node 4).





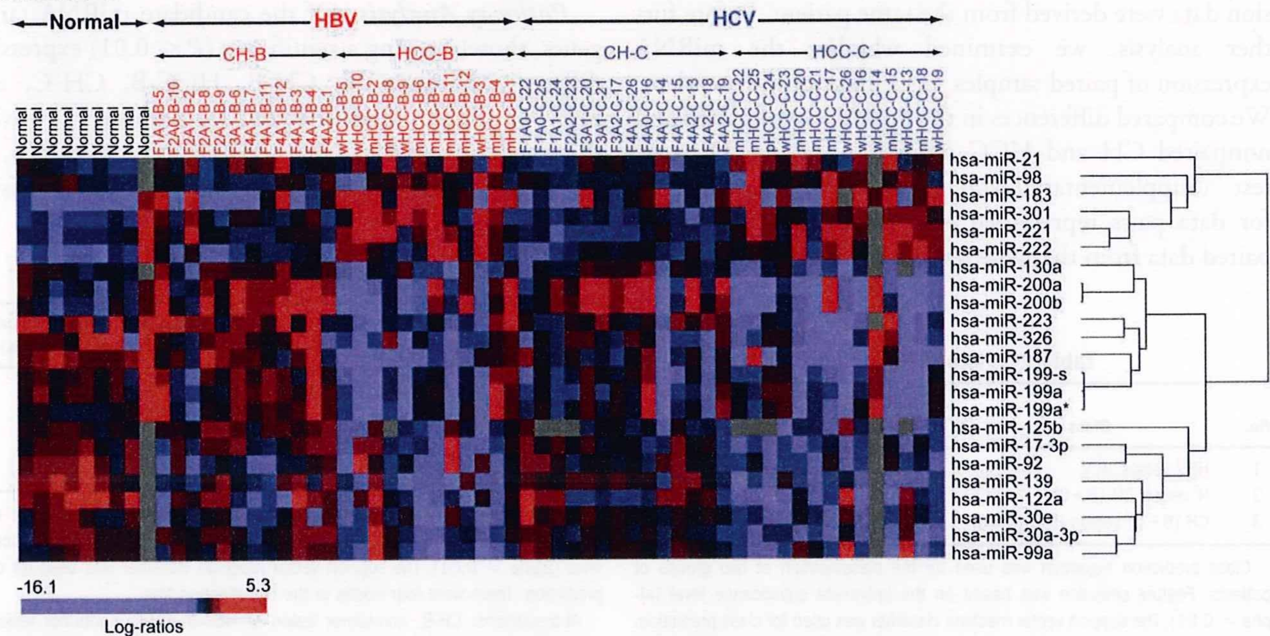


Fig. 3. Cluster 1: Eight miRNAs specifically differentiated node 1 classification. Cluster 2: Nineteen miRNAs specifically differentiated node 2 classification. Cluster 3: Twenty-three miRNAs differentiated CH-B and HCC-B as well as CH-C and HCC-C.