[11] Satoh, H., Ono, I. and Kobayashi, S.: A new generation alternation model of genetic algorithm and its assessment, *J. of Japanese Society for Artificial Intelligence*, 15(2) (1997) 743-744.

[12] Ritt, J.F.: Differential Algebra, *Dover Publications Inc., NY*, (1950).

[13] Kolchin, E.R.: Differential Algebra and Algebraic Groups, *Academic Press, NY*, (1973).

Joint 48th IEEE Conference on Decision and Control and
28th Chinese Control Conference
Shanghai, P.R. China, December 16-18, 2009

Detection of network structure changes by graphical chain modeling: a case study of hepatitis C virus-related hepatocellular carcinoma

Shigeru Saito, Masao Honda, Shu-ichi Kaneko and Katsuhisa Horimoto

Abstract—**One of the most characteristic features of biological molecular networks is that the network structure itself changes, depending on the cellular environment. Indeed, activated molecules show a variety of responses to distinctive cell conditions, and subsequently the network structures of active molecules also change. Here we present an approach to trace the network structure changes by using the graphical chain model developed from the gene expression data. The previous procedure for applying the graphical chain model to the expression profiles of a limited number of genes has been improved to analyze the entire set of genes. Furthermore, the chain model has been rearranged according to the association strength, and was scrutinized to identify the candidates of essential gene-gene relationships for the network changes, by using the path consistency algorithm. The improved procedure was applied to the expression profiles of 8,427 genes, which were measured in two distinctive stages of liver cancer progression. As a result, the chain model of the 18 gene cluster relationships with strong associations was inferred, in which the coordination of clusters was described in the cell stage progression, and the gene-gene relationships between known cancer-related genes causing the progression were further refined. Thus, the present procedure is a useful method to model the network structure changes in the cell stage progression, and to clarify the gene candidates for the progression.**

I. Introduction

One of the remarkable relationships between molecules in living organisms is the drastic changes of network structures in response to the environment. For example, it is well known that a specific set of molecules, among all of the molecules in a cell, is activated, in response to environmental stress [1, 2]. Unfortunately, the experimental techniques for monitoring the activated molecules in a living cell still require further development. Thus, it is desirable to be able to infer the network structure of activated molecules from data measured under distinctive conditions.

Manuscript received January 28, 2009. This work was supported by a Grant-in-Aid for Scientific Research on Priority Areas "Systems Genomics" (grant 20016028) and for Scientific Research (A) (grant 19201039) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

S. S. is with National Institute of Advanced Industrial Science and Technology, Tokyo 135-0064, Japan, and also with INFOCOM Corporation, Tokyo 150-0001, Japan (e-mail: sh.saito@ infocom.co.jp).

M. H. is with Kanazawa University Graduate School of Medical Science, Kanazawa 920-8641, Japan (e-mail: mhonda@m-kanazawa.jp).

S. K. is with the Kanazawa University Graduate School of Medical Science, Kanazawa 920-8641, Japan (e-mail: skaneko@m-kanazawa.jp).

K. H. is with National Institute of Advanced Industrial Science and Technology, Tokyo 135-0064, Japan, and also with Shanghai University, Shanghai 200444,, China (phone: +81-3-3599-8711; fax: +81-3-3599-8081; e-mail: k.horimoto@ aist.go.jp).

In human liver infected by hepatitis C virus (HCV), the infection induces the development of chronic hepatitis (CH), cirrhosis, and in some instances, hepatocellular carcinoma (HCC) [3]. The virological features of the infection were described [4, 5], and we previously reported that the gene expression profiles in chronic hepatitis C (CH-C) predominantly induced inflammatory and anti-apoptotic phenotypes. However, the network structure changes inducing these modifications in gene expression still remain to be elucidated.

Recently, we have developed a procedure [6] for tracing the network structure changes from the gene expression data by using the graphical chain model (GCM) [7-10]. In the application of GCM to the expression data measured in progressive cell stages, the block and the variables in blocks correspond the cell stage and the genes characteristically expressed in each stage, respectively. Since GCM exhibits the overlaps of the variables between the blocks, the genes responsible for distinctive stages should be selected among the set of entire genes. Indeed, in our previous application of GCM to the yeast cell-cycle [6], we adopted the gene sets of about 700 genes that characterized each cell stage, from a previous study [11]. In general, the genes that are characteristically expressed in distinctive stages are identified most effectively by discriminating between the stages. In this case, however, the genes that are continuously up-regulated or down-regulated over the stages are not selected. In the case of progressive processes, the continuously up (or down)-regulated genes may be important for identifying the molecular mechanisms underlying the stage progression.

Here, we have improved the previous procedure [6] to detect the changes of network structures more efficiently, by using the entire set of genes. The procedure was applied to the expression profiles measured in two stages of hepatocellular carcinoma, from CH to HCC [4, 5]. For each stage, all of the genes in the analyzed data were systematically classified into three groups, up-, down-, and unchanged regulated gene groups, and based on the classification, the three blocks in GCM-CH, HCC, and background-were defined. In particular, the background includes the genes that are continuously up- and down-regulated, and the influence of these genes on the progression from CH to HCC was estimated. Thus, the improved procedure allowed us to describe the network structure changes of entire sets of genes. Furthermore, the transformation of the inferred network structure helped to reveal the candidates of the gene-gene relationships causing

978-1-4244-3872-3/09/$25.00 ©2009 IEEE

the cancer progression, including known cancer-related genes.

II. MATERIALS AND METHODS

A. Expression Profile Data

The expression profiles of 8,427 genes were monitored in 6 normal, 32 CH and 17 HCC samples [4]. Relative expression ratios of 8,427 genes were obtained by comparing hybridization of Cy5-labeled cDNAs from chronic hepatitis lesions and Cy3-labeled cDNA from normal liver tissue.

B. Graphical Chain Modeling (GCM)

The graphical chain model (GCM) is a probability model for multivariate random observations, in which the independence of the structure can be represented by a graph [7-10]. Here, we will briefly describe GCM.

The graph $G = (V, E)$ consists of a set of vertices V, representing the variables, and a set of edges E, representing the associations between pairs of variables. E is a set of ordered pairs (A, B), $A, B \in V$. The chain graph is based on the partitioning of V into disjointed subsets: $V = V_1 \cup V_2 \cup \ldots \cup V_T$. The subsets are called blocks or chain components. The edges within blocks are undirected, reflecting the systematic associations, and the edges between blocks are arrows pointing from blocks with lower index numbers to those with higher indices. A graphical chain model displays the independence between variables conditioned on all of the other variables in the current and previous blocks. In a graphical chain model, any direct association between two variables in the same block is assumed to be non-causal, and is represented by an undirected edge (line) in a graph. Any direct association between two variables from different blocks is assumed to be potentially causal, and is represented by a directed edge (arrow). The absence of a line or arrow between two variables in the graph indicates that there is no direct association between the variables, i.e. the variables are independent, after controlling for all of the other variables in the same and previous blocks.

The graphical chain model is fitted in a number of stages. When fitting a graphical chain model, the first step is to partition the variables into a number of ordered blocks. Then, the significant direct associations between the variables in the first block are determined. For each pair of variables, the null hypothesis when tested shows that the variables are independent, given all of the other variables in the first block, and the deviance statistics in graphical Gaussian modeling (GGM) is used [12].

Next, the significant direct associations between the variables in the second block and between the first and second blocks are determined. For each pair of variables, the null hypothesis when tested shows that the variables are independent, given all of the other variables in the first and second blocks, and again the deviance statistics is used.

The fitting continues, block by block, by determining all of the significant direct associations between the variables in the current block and between all of the variables in the current and previous blocks. The null hypothesis is now independence, given the other variables in the current and previous blocks, and again the deviance statistics in GGM is used. In other words, the procedure of GCM is the iteration of GGM. All of these tests were carried out at the 5% level, using the χ^2 distribution in deviance statistics.

In the present study, the block in the graphical chain model simply corresponds to the cell stages that are defined by biological information. By the intact correspondence to graphical chain modeling, the variable is the gene that has an expression profile with numerical values. However, since the expression profiles often show similar patterns, the genes are highly related to one another. Thus, hierarchical clustering is performed for the genes within each block, as a preprocessing step for the graphical chain modeling [7-10], and then, each gene cluster corresponds with the variable in the present procedure.

C. Improved Procedure for Applying GCM to Gene Expression Profiles

In the present study, we improved the procedure for applying GCM to gene expression measured in two cell stages. The overview of the present procedure is described in Fig. 1.

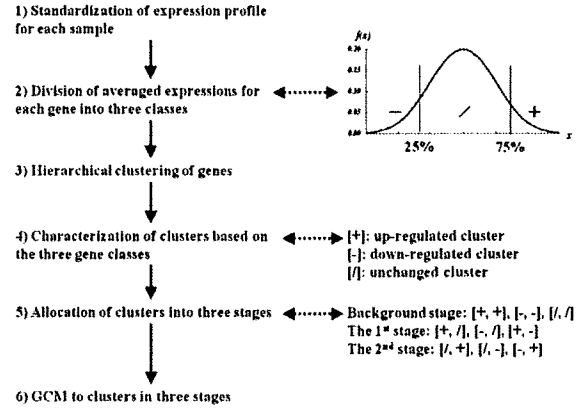

Fig. 1. Improved procedure for applying GCM to gene expression profiles in two stages.

1) Standardization

The expression profiles of all genes are standardized by the average and the standard deviation for each sample, i.e.,

$$z_{ij} = \frac{x_{ij} - AV_j}{SD_j},$$

where z_{ij} and x_{ij} are the standardized and intact expression values of the i-th gene and the j-th sample, respectively, and AV_j and SD_j are the average and the standard deviation of all genes over the j-th sample, respectively. This allows the noise of the expression profiles of each gene, due to the differences between samples, to be excluded by the transformation of the

intact expression degree into a standardized value, the z-value.

2) Division of Average Expression Values into Three Classes

The average expression values for each gene are calculated over the samples in one stage. Then, the genes are divided into three classes in terms of the average value: up-regulated gene class ('+', abbreviation of class), down-regulated gene class ('-'), and the other class ('/'). If the average value of a gene ranges more than the 25 th percentile, less than the 75 the percentile, and between the 25 th and 75 th percentiles, then the corresponding gene is regarded as up-regulated, down-regulated, and the other class, respectively.

3) Hierarchical Clustering

All genes in the analyzed data were subjected to hierarchical clustering. In the present clustering, the metric is Pearson's correlation coefficient of genes between the expressions of samples, and the technique is the Un-weighted Pair Group Method using the Arithmetic average (UPGMA). The number of clusters was estimated by using the variance inflation factor, defined in the previous study [13].

4) Cluster Characterization in Terms of Expression Class

Based on the gene classification into the three classes mentioned in 2), the clusters are characterized: each stage is characterized by the maximum number of gene classes. For example, if 50, 10, and 5 genes belonging one cluster in one stage are '+', '/', and '-', respectively, then the cluster in the stage is characterized by [+]. According to the above rule, the class pairs in the two stages were defined: the nine class pairs in the two stages are [+, +], [+, /], [+, -], [/, +], [/, /], [/, -], [-, +], [-, /], and [-, -].

5) Allocation of Three Expression Classes into Clusters

The clusters with the pairs of three expression classes for the two stages described in 4) were allocated to three groups. The rules for allocation are simple. First, if a cluster shows an up-regulated class at only one stage, then the cluster is allocated into a group that represents the corresponding stage. Second, among the remaining clusters, if a cluster also shows a down-regulated class at only one stage, then the cluster is allocated into a group that represents the corresponding stage. Finally, the remaining clusters are allocated into a hypothetical group, named the background stage. This is because the up-regulation of the gene indicates that the corresponding gene product increases, and plays a important role in the stage. Thus, for example, the clusters allocated into the first cell stage are composed as follows: according to the first rule, the up-regulated cluster class in the first cell stage and the other cluster class in the second cell stage [+, /] and the up-regulated cluster class in the first cell stage and the down-regulated cluster class in the second cell stage [+, -] are allocated. The down-regulated cluster class in the first cell stage and the other cluster class in the second cell stage [-, /] are then allocated, according to the second rule. The clusters are allocated into the second cell stage according to the same rule: [/, +], [-, +], and [/, -]. Finally, the clusters allocated into the background stage are the remaining clusters, i.e., [+, +], [-, -], and [/, /]. Here, the background stage is a hypothetical stage for considering the influence of the expression of uncharacterized genes on that of well-characterized genes, depending on the two cell stages. In other words, the variables in the hypothetical stage are viewed as purely explanatory variables, whereas the variables in the second and subsequent blocks are viewed as responses to the variables in the preceding blocks. Note that the usual approach, based on the gene selection by discrimination between the two stages, does not consider the effect of the genes showing a constant degree of gene expression.

6) GCM

Finally, we perform GCM for the three groups of clusters that were allocated in 5). In this case, the order of the above groups is set in the order of the background stage, the first stage, and the second stage. It seems natural that the background stage influences both the first and second stages. Thus, we can estimate the causal relationship between the clusters in the first stage and the background and between the clusters in the second stage and the background, as well as between the clusters in the first and second stages.

D. Securitization of Chain Model

The chain model obtained by the standard algorithm of GCM frequently has many edges, and thereby adopts a messy form with many nodes and edges. Even in a sparse form, each node obtained by cluster analysis for the entire set of genes also frequently contains many genes. To scrutinize the genes responsible for the network structure change in the chain model, therefore, we further devised two techniques: one for the former issue is the evaluation of the association strength of each edge in the model, and the other for the latter issue is the inference of the gene-gene association in the selected clusters. The details of the two techniques are described below.

1) Evaluation of Association Strength in Chain Model

When there are many edges, drawing them all on one graph produces a mess or 'spaghetti' pattern, which would be difficult to read. Indeed, in some examples of the application of GGM to actual profiles, the intact networks derived by GGM still showed complicated forms with many edges [14]. The similar situation may be expected in GCM, which is the iteration of GGM. Thus, the strength of the association between clusters is evaluated in a statistical way: the intact network can be rearranged according to the partial correlation coefficient value, to interpret the associations between clusters. The strength of the association can be assigned by a standard test for the partial correlation coefficient, $r_{ij,rest}$, between the variables i and j, given the resting variables [15]. By Fisher's Z transformation of partial correlation coefficients, i.e.,

$$Z = \frac{1}{2}\log\left(\frac{1+r_{ij\cdot rest}}{1-r_{ij\cdot rest}}\right),$$

Z distributes according to the following normal distribution:

$$N\left(\frac{1}{2}\log\left(\frac{1+r_{ij\cdot rest}}{1-r_{ij\cdot rest}}\right),\frac{1}{\{N_c-(M-2)\}-3}\right),$$

where N_c and M are the number of conditions and the number of clusters, respectively. Thus, we can statistically test the observed correlation coefficients under the null hypothesis with a significance probability.

2) Application of Path Consistency Algorithm

The application of GCM to the entire data set with the combination of cluster analysis produces a macroscopic view of the causal relationships between them. This is partly because the entire data set contains a similar pattern of gene profiles, and partly because the noise in the data due to various effects prevents us from inferring the entire relationship at a fine level, such as the causal relationship between the genes. However, if partial sets of genes are selected by GCM, then the selected sets of genes may be appropriate for applying a fine analysis to infer a causal relationship between the genes. Thus, the relationships between the constituent genes in the clusters whose relationships are inferred by GCM are also inferred by another graphical modeling method, the path consistency (PC) algorithm [16]. The PC algorithm is composed of two parts: the undirected graph inference by the partial correlation coefficient and the following directed graph, obtained by using the orientation rule based on the inferred undirected graph. A brief overview of the PC algorithm and the modifications for the present analysis is described below.

The first part of the algorithm is simple. The relationship between two variables is tested from the lower partial correlation coefficient to the higher one. For example, the relationship between the two variables is first tested by the zero-th partial correlation coefficient. If the null hypothesis is accepted, i.e., no association between the two variables, then, the relationship between the two variables is tested by the first partial correlation coefficient. If it is rejected, then the test is not performed. In general, the (m-2)-th order of the correlation coefficient is calculated between two variables, given (m-2) variables, i.e., $r_{ij,rest}$, between X_i and X_j, given the '*rest*' variables, $\{X_k\}$ for k=1, 2, ..., m, and $k\neq i$, j, and after calculating the (m-2)-th order of correlation coefficient, the algorithm naturally stops. However, the algorithm does not usually request the (m-2)-th order of correlation coefficient for the natural stop. This is because no adjacent variables will be found after excluding the variables, even in the calculation of the lower order of the correlation coefficient. In the practical analysis of sample data, the zeroth-order of the correlation coefficient is calculated by Pearson's correlation coefficient, r_{ij}, expressed by

$$r_{ij}=\frac{\mathrm{cov}(X_i,X_j)}{\sqrt{\mathrm{var}(X_i)\,\mathrm{var}(X_j)}},$$

where cov (X_i, X_j) and var(X_i) are the covariance between X_i and X_j, and the variance of X_i, respectively. The higher order of the correlation coefficients is the partial correlation coefficient, $r_{ij,rest}$, expressed by

$$r_{ij,rest}=\frac{-r^{ij}}{\sqrt{r^{ii}\cdot r^{jj}}},$$

where (ij, *rest*) means $\{1,2,\ldots,p\}\backslash\{i,j\}$, and r^{ij} is the i-j element of the inverse correlation coefficient matrix. Note that the dimension of the correlation coefficient matrix corresponds to the orders of the correlation coefficients. The n th-order correlation coefficient is calculated from the (n+2) dimension of the correlation coefficient matrix. The correlation coefficient is statistically tested by using the Z-statistic [15], as in the case of the association strength evaluation in i).

Based on the inferred undirected graph, C, the direction of each graph is decided in the second part, according to the orientation rule [17]. The rules for the direction decision in C are as follows:

i) If there is an undirected relationship, X−Y−Z, and X and Z are not adjacent, then the direction is decided as being X→Y←Z.

ii) If there is a relationship, X→Y−Z, and X and Z are not adjacent, then the direction is decided as being X→Y→Z.

iii) If there is a directed path between X and Y, and there is a relationship, X−Y, then the direction is decided as being X→Y.

The key point in the present network inference is a modification of the original PC algorithm, for application to the expression profiles. The modification corrects the algorithm in the calculation of the partial correlation coefficient. Since many genes frequently show very similar patterns of expressions, the difficulty arises in the numerical calculation of correlation coefficients, due to the multi-colinearity between the variables. The original PC algorithm accidentally stops, if only one correlation between a pair of variables shows a violation of the numerical calculation, against the high similarity of the expressions. To escape the accidental stop by the highly associated gene pairs, the original PC algorithm was modified as follows: if the calculation of any order of correlation coefficient between the variables is violated, then the corresponding pair of variables is regarded as being dependent.

E. Software

All calculations of the present clustering and GGM were performed by the ASIAN web site [18](http://eureka.cbrc.jp /asian) and "Auto Net Finder", the commercialized PC version of ASIAN, from INFOCOM CORPORATION (http://www.infocom.co.jp/bio/download/).

III. RESULTS AND DISCUSSION

A. Clustering and Its Allocation into the Three Stages

According to the procedure described in the preceding section, first, all genes characterized by the three degrees of expression were subjected to the cluster analysis, and then the

TABLE I
ALLOCATION OF CLUSTERS INTO THE THREE STAGES

Cluster Number	Number of members	Cluster class pair	Number of genes in corresponding class pair	Allocated stage
1	146	[+, /]	66	CH
2	476	[+, /]	271	CH
3	654	[/, /]	394	B
4	198	[/, +]	54	HCC
5	440	[-, -]	173	B
6	229	[/, -]	127	HCC
7	51	[+, +]	47	B
8	9	[+, +]	9	B
9	288	[-, /]	85	CH
10	400	[/, +]	137	HCC
11	318	[-, /]	77	CH
12	272	[/, +]	115	HCC
13	188	[-, /]	85	CH
14	277	[-, -]	191	B
15	179	[/, /]	70	B
16	63	[+, /]	14	CH
17	310	[/, /]	86	B
18	184	[+, /]	38	CH
19	47	[-, -]	44	B
20	110	[-, -]	102	B
21	160	[-, /]	28	CH
22	950	[/, /]	565	B
23	1011	[+, /]	287	CH
24	23	[-, -]	23	B
25	392	[+, +]	281	B
26	127	[-, -]	92	B
27	274	[/, +]	126	HCC
28	40	[-, +]	6	HCC
29	316	[-, -]	293	B
30	115	[/, -]	22	HCC

clusters were allocated into the three stages. The clusters and their allocations are listed in TABLE I

As seen in TABLE I, all of the genes were grouped into 30 clusters, with various numbers of members. Indeed, the minimum and maximum numbers of genes among the 30 clusters were 9 in cluster 8 and 1,011 in cluster 23, respectively, and the average number of genes was 274.9, with a standard deviation of 243.3 genes. Based on the rule in II-C-5), the clusters were allocated to the three stages, Background, CH, and HCC. The classes of clusters in the CH and HCC stages are listed in the third column in the table, and the numbers of genes belonging to the class pairs of the two stages are listed in the fourth column. The total numbers of genes belonging to the class pairs were 3,908 among the 8,247 genes analyzed in the present study: the average fraction of the numbers of genes in the fourth column to those in the second column was 0.513, with a standard deviation of 0.275. Thus, more than half of the genes were responsible for the allocation of the clusters into the three stages. Finally, the numbers of clusters allocated to Background, CH, and HCC were 14 (2,370 genes), 9 (951 genes), and 7 (587 genes), respectively. The present allocation of clusters, with its small bias, may reveal the relationships between the clusters in the progression from CH to HCC.

B. Chain Model

The partial correlation coefficient matrix generated by

TABLE II
PARTIAL CORRELATION COEFFICIENT MATRIX FOR THE THREE STAGES BY GCM

	Background														CH									HCC						
Cluster	3	5	7	8	14	15	17	19	20	22	24	25	26	29	1	2	9	11	13	16	18	21	23	4	6	10	12	27	28	30
3	-1																													
5	0.275	-1																												
7	0	0.255	-1																											
8	0	-0.16	0	-1																										
14	0.248	0.181	0	0	-1																									
15	0	-0.28	0	0	0	-1																								
17	0	0	0	0	0.178	0.298	-1																							
19	0	0	-0.223	0	0.617	0	-0.220	-1																						
20	0.147	-0.26	0.290	0	0	0	0	-0.249	-1																					
22	-0.248	0.222	-0.350	0	-0.26	0	0	-0.255	0	-1																				
24	-0.253	0	0	-0.174	0	0	0	0.574	0.681	0.332	-1																			
25	0.421	0.165	0	0.158	-0.65	0	0	0.299	0	0	0.132	-1																		
26	-0.184	0.188	-0.226	-0.376	-0.24	0	0.448	0.215	0.315	0	-0.37	0	-1																	
29	0	0.53	-0.512	0.232	0	0	-0.132	-0.198	0.644	-0.31	0	-0.25	0	-1																
1	0	0	0.286	-0.224	0	-0.201	0	-0.440	-0.257	-0.23	0	0	0	0.394	-1															
2	0	0	0	0.186	-0.33	0	0.233	0	0	0.292	0	0	0	-0.11	0	-1														
9	0	-0.44	-0.475	0	0	-0.260	0	0	-0.322	0	0.158	0	-0.41	0	0	0	-1													
11	-0.146	0	-0.321	0.270	0	-0.289	-0.395	0	0	0	0	0	0	-0.31	0	-0.398	-0.381	-1												
13	0.443	0	0	0.277	0	-0.196	0.207	0.275	-0.224	0	0	0	-0.33	0	-0.21	-0.356	-0.209	-0.19	-1											
16	0.281	-0.21	0	0	-0.24	0.226	0.201	0.284	0	0.241	0	0	0	0	0.174	-0.377	0	0	-0.27	-1										
18	-0.203	0	0	0	-0.3	0.305	0.468	0	0	0	0.244	0	0.417	-0.22	0	-0.402	0	0	0	-0.22	-1									
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1								
23	0.329	-0.21	-0.290	0	0	0	0	0	-0.289	0.264	0.206	0.261	-0.42	0	0	0	-0.625	-0.16	-0.34	-0.17	0	0	-1							
4	0	0.263	0	0.458	0	0	0	0.212	-0.530	-0.42	0.349	-0.41	0	-0.37	0	0	-0.291	0.242	0	0	-0.25	0	-0.258	-1						
6	0.221	-0.22	0.474	-0.451	0	0.144	0.341	0.222	0.416	0.336	-0.28	-0.17	-0.38	0.746	0.743	0	0	0	-0.32	0	0.190	0.228	0.335	0.447	-1					
10	-0.238	-0.2	0.164	0	0	-0.243	-0.522	-0.442	-0.153	-0.14	0	0	0	0	-0.41	0	0.858	0.157	0.247	-0.24	0	-0.58	0	0.169	0	-1				
12	-0.237	-0.48	0.337	0	-0.35	-0.415	-0.357	-0.183	-0.484	-0.52	0.238	-0.56	-0.31	0	-0.27	-0.191	0.157	0.676	0	-0.45	-0.32	-0.5	-0.342	-0.180	0	-0.47	-1			
27	-0.263	-0.41	0.295	-0.250	-0.23	-0.475	0	-0.125	-0.366	-0.48	0	-0.32	-0.46	-0.18	-0.23	-0.512	-0.215	0	0	-0.46	-0.49	-0.4	-0.477	0	0	-0.17	-0.42	-1		
28	-0.149	-0.16	-0.232	0	0	-0.231	-0.234	-0.113	0	-0.19	0	0	0	-0.11	0	-0.123	0	-0.25	0	0	0	-0.24	0	0	0	0	0	0	-1	
30	-0.280	-0.43	0.272	-0.357	-0.21	-0.437	-0.218	-0.180	-0.166	-0.49	0	-0.46	-0.32	0	-0.13	-0.230	0	0.242	0	-0.38	-0.3	-0.42	-0.240	0	-0.19	-0.31	-0.66	-0.52	-0.18	-1

GCM is shown in TABLE II. By GCM, 195 (44.8%) of the possible 435 edges in the full model of 30 clusters were deleted; thus, there were still 242 edges in the inferred chain model, and its form was too complicated to interpret the entire association between the clusters. Within and between the three stages, the numbers of deleted edges were relatively uniform, except for those between Background and HCC: only 26 (26.5%) of 98 edges (=14x7) were deleted. In other words, the influence of the clusters in the Background to those in HCC was inferred in the model.

C. *Arrangement of Chain Model*

According to the procedure for evaluating the association strength, the chain model corresponding to the partial correlation coefficient matrix was arranged. To interpret the model, the strong associations with the significant (p<0.01, r>0.510) edges were selected: in the rearranged model, 22 clusters showed the strong association, and 8 clusters were isolated (not depicted in the figure).

By the above arrangement, a simple chain model emerged in Fig. 2. As seen in the figure, the number of edges within the three stages was less than those between them. In particular, there was only one edge in CH, and HCC had two edges, whereas there were 6 edges in Background. In contrast, 10 edges emerged between the three stages: there were 5 edges from Background to HCC, and in contrast, there were no edges from Background to CH. Thus, the bias of the numbers of edges between the stages to those within each stage suggests that the genes were coordinately expressed in accordance with the cancer progression.

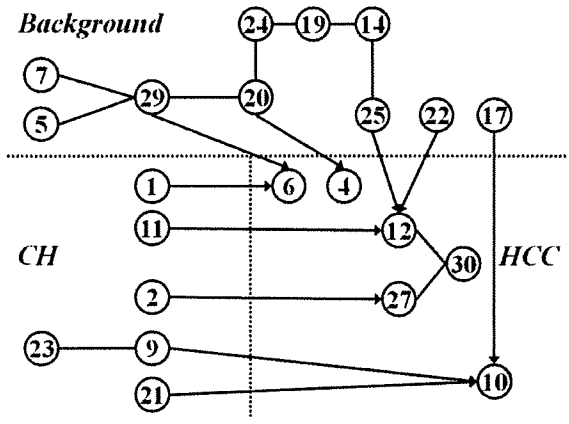

Fig. 2. Chain model for hepatitis C virus-related hepatocellular carcinoma. In the chain model, the edges showing strong associations with less than 1% significance probability are depicted, but the isolated edges are not. The numbers in the circles correspond to the cluster numbers in TABLE I. The clusters allocated to the background, CH, and HCC are located in the upper, left, and right parts of the figure, respectively, and are separated by dotted lines.

The following relationships are particularly remarkable: cluster 29 in Background, cluster 1 in CH, and cluster 6 in HCC ({29→6←1}), {(22, 25)→12←11}, and {17→10←(9, 21)}. In the above cluster sets, the clusters in Background and in CH were coordinately related to the cluster in HCC. Since conventional approaches consider only the differences in gene expression between the two stages, the above relationships between the three stages, Background, CH, and HCC, would frequently be neglected. Thus, these clusters have the possibility of containing the essential genes responsible for cancer progression, which were not detected by the conventional discrimination approaches.

D. *Gene-Gene Relationships between Clusters*

Here, we focus on the genes responsible for the relationships between the three cluster sets in the preceding section. However, since the clusters still contain many genes, further analysis is needed to refine the candidates of the genes responsible for the cluster-cluster relationships. For this purpose, the PC algorithm was performed for the genes belonging to the respective sets. Then, the gene-gene relationships consistent with the relationship between the clusters were selected. For example, in the first set, {29→6←1}, all of the genes belonging to clusters 29, 1 and 6 were attributed by the PC algorithm. If the inferred relationship of the genes belonging to the respective clusters is consistent with that between clusters 29, 1, and 6, then the gene sets were extracted from all of the gene-gene relationships inferred by the PC algorithm.

TABLE III
GENE-GENE RELATIONSHIPS IN COORDINATED CLUSTER-CLUSTER RELATIONSHIPS

Cluster Relationships	Gene Relationships
{29→6←1}	ARL6IP→CDH1←FGFR2, Hs.209450→BHMT2← SFRS11, SCP2→BHMT2←SFRS11, CYP4X1→NDRG1← ELYS, GPAM→SYP←KIAA0182, RRAS2→CLDN12← PXMP3, ARCN1→VDAC1←Hs.433078, SHANK2→ PDSS2←COX7B, MYO1B→KTN1←FKBP3, ARCN1→ KIAA0073←RNF146
{22→12←11}	NEO1→KRT14←FLJ12270, NEO1→KRT14←IGHMBP2, RANBP3→IL1RL1←SSRP1, Hs.79241→CD2←NFKB2, MGC4606→SLA/LP←EFNB2, PKD2L1→TK2←GNAI2, ADA→FLJ46603←HOXA9, GGA3→RAP1A←LSM14B, LMLN→RAP1A←LSM14B, ZNF638→GTPBP6← Hs.298289, DCT→GTPBP6←Hs.298289, Hs.191356→ PTMS←ATP2A3, Hs.378847→UBE2G2←TEAD3, LIM→ UBE2G2←TEAD3, TCF7L2,FLJ23556→ CTAG1B,CTAG1A←Hs.107410, SAPS3→MYO1E← ZNF175, SYNE1→SLC12A2←Hs.127657, C4.4A→PTEN ←NAPB, MGC23985→DDB1←HOXA9
{25→12←11}	T→MYCN←SLC30A4, CENTB1→CD2←NFKB2, SPINT2 →IRF3←ACVRL1, Hs.504960→TRAF2←BYSL, Hs.504960→TRAF2←SP192, MGC11266→CSNK1E← MCRS1, Hs.23871→ELAVL2←Hs.439153, GFPT1→BCL3 ←Hs.2173, KIAA1030→SEC5L1,HUS1B←TEB4
{17→10←9}	NAB2→STARD3←OPHN1, CINP→KRT8P12←MYF5

The gene-gene relationships narrowed down by the above procedure are listed in Table III. As seen in the table, 40 gene-gene relationships were narrowed down from a large number of possible gene relationships in the cluster

relationships, {29→6←1}, {22→12←11}, {25→12←11}, and {17→10←9}: in {17→10←21}, no gene relationships were detected. Interestingly, known cancer-related genes were included in the constituent genes of the relationships. Indeed, 17 genes in 17 relationships are described as the cancer-related genes in OMIM (Online Mendelian Inheritance in Man) [19]: CDH1, FGFR2, RRAS2, NEO1, NFKB2, TK2, HOXA9, ATP2A3, TCF7L2, CTAG1B, C4.4A, PTEN, MYCN, SPINT2, TRAF2, BCL3, and STARD3. Note that both known cancer-related genes and genes with other functions were included in most relationships. In these cases, genes with other functions may be involved in the cancer progression. In addition, some genes with unidentified functions were also included in the relationships above. This may suggest that one of the functions of the genes may be related to the cancer progression. At any rate, the gene-gene relationships, which were narrowed down with reference to network structure changes, reflect well the knowledge about the genes responsible for the cancer, and show the possibility of unknown relationships related to the cancer progression.

E. Merits and Pitfalls

We proposed a procedure for inferring a model for progressive stages from the entire data set, by using the graphical chain model. Furthermore, the following analyses of the evaluation of association strength by a statistical test and the selection of gene-gene relationships by the PC algorithm narrowed down the candidates of the gene sets causing the inferred cluster-cluster relationships. By using the above procedure, we analyzed 8,427 gene expression profiles in the two stages of hepatocellular carcinoma from CH to HCC. By the analyses, the chain model including the background stage was constructed, and several gene cluster connections were found to cause the progression from CH to HCC. Furthermore, 40 candidates of gene-gene relationships responsible for the progression emerged, with reference to the cluster-cluster relationships. Thus, we successfully described a framework of network structure changes for cancer progression, and based on the inferred changes, further refined the causal gene-gene relationships for the cell stage progression in a rational and systematic manner.

It is interesting in considering the present procedure in the case of more than two cell stages. For example, we can allocate 27 cluster sets of three cell stages into four groups, according to the rule adopted in the two cell stage: the first group, [+, /, /], [+, -, /], [+, -, -], [+, /, -], and [-, /, /]; the second group, [/, +, /], [-, +, /], [-, +, -], [/, +, -], and [/, -, /]; the third group, [/, /, +], [-, /, +], [-, -, +], [/, -, +], and [/, /, -]; the background group, [+, +, +], [-, -, -], [/, /, /], [+, +, -], [+, +, /], [+, -, +], [+, /, +], [-, +, +], [/, +, +], [-, -, /], [-, /, -], and [/, -, -]. In the allocation of the above clusters into the background group, some ambiguity emerged; the clusters showing up- or down-regulation at two cell stages are included. However, the present allocation rule may be reasonable, on the assumption that the network structure change is responsible for the up- and down-regulated classes characterizing each cell stage. Although the ambiguity of the allocation into the background group emerges as the number of stages increases, the present procedure may help to reveal the gene-gene relationships as well as to capture the network structure change through the distinctive cell stages in a systematic manner.

REFERENCES

[1] T. Finkel and N.J. Holbrook, "Oxidants, oxidative stress and the biology of ageing," Nature, 408, 239-247, 2000.

[2] H.C. Causton, B. Ren, S.S. Koh, C.T. Harbison, E. Kanin, E.G. Jennings, T.I. Lee, H.L. True, E.S. Lander and R.A. Young, "Remodeling of Yeast Genome Expression in Response to Environmental Changes," Mol. Biol. Cell, 12, 323-337, 2001.

[3] K. Kiyosawa, T. Sodeyama, E. Tanaka, Y. Gibo, K. Yoshizawa, Y. Nakano, S. Furuta, et al., "Interrelationship of blood transfusion, non-A, non-B hepatitis and hepatocellular carcinoma: analysis by detection of antibody to hepatitis C virus," Hepatology, 12, 671-675, 1990.

[4] M. Honda, S. Kaneko, H. Kawai, Y. Shirota, K. Kobayashi, "Differential gene expression between chronic hepatitis B and C hepatic lesion," Gastroenterology, 120, 955-966, 2001.

[5] M. Honda, T. Yamashita, T. Ueda, H. Takatori, R. Nishino, S. Kaneko, "Different signaling pathways in the livers of patients with chronic hepatitis B or chronic hepatitis C," Hepatology, 44, 1122-1138, 2006.

[6] S. Aburatani, S. Saito, H. Toh and K. Horimoto, "A Graphical Chain Model for Inferring Regulatory System Networks from Gene Expression Profiles," Statistical Methodology, 3, 17-28, 2006

[7] N. Wermuth and S. L. Lauritzen, "On substantive research hypotheses, conditional independence graphs and graphical chain models," J. R. Statist. Soc. B, 52, 21-50, 1990.

[8] J. Whittaker, *Graphical Models in Applied Multivariate Statistics*, Wiley, Chichester, 1990.

[9] D. Edwards, *Introduction to Graphical Modelling*, Springer, New York, 1995.

[10] S. L. Lauritzen, *Graphical Models*, Oxford University Press, Oxford, 1996.

[11] P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein and B. Futcher, "Comprehensive identification of cell cycle–regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization," Mol. Biol. Cell, 9, 3273–3297, 1998.

[12] P. Dempster, "Covariance selection," Biometrics, 28, 157-175, 1972.

[13] K. Horimoto and H. Toh, "Statistical Estimation of Cluster Boundaries in Gene Expression Profile Data," Bioinformatics, 17, 1143-1151, 2001.

[14] S. Aburatani, F. Sun, S. Saito, M. Honda, S. Kaneko and K. Horimoto, "Gene systems network inferred from expression profiles in hepatocellular carcinogenesis by graphical Gaussian model," EURASIP J. Bioinfo. Systems Biol., 47214, 2007.

[15] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 2nd Edition. New York, John Wiley & Sons, 1984.

[16] P. Spirtes, C. Glymour and R. Scheines *Causation, Prediction, and Search* (Springer Lecture Notes in Statistics, 2nd edition, revised). MIT Press, Cambridge, 2001.

[17] T. Verma and J. Peral, "An algorithm for deciding if a set of observed independence has a causal explanation," Proc. 8th Conf. on Uncertainty in AI, Stanford, Morgan Kaufmann, p. 323-330, 1992.

[18] S. Aburatani, K. Goto, S. Saito, H. Toh and K. Horimoto, "ASIAN: A Web Server for Inferring a Regulatory Network Framework from Gene Expression Profiles," Nucleic Acids Res, 33, W659-W664, 2005.

[19] http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim

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.

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

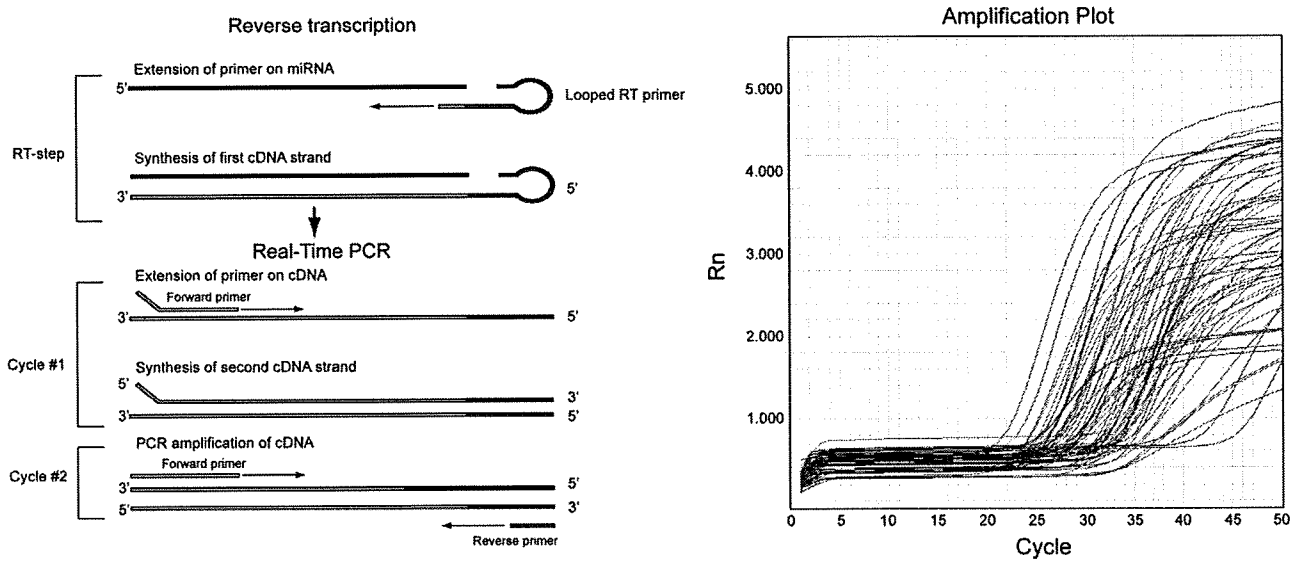

Fig. 1. (A) miRNA-specific RTD-PCR using sheet hairpin primers. (B) miRNA amplification curves by RTD-PCR.

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

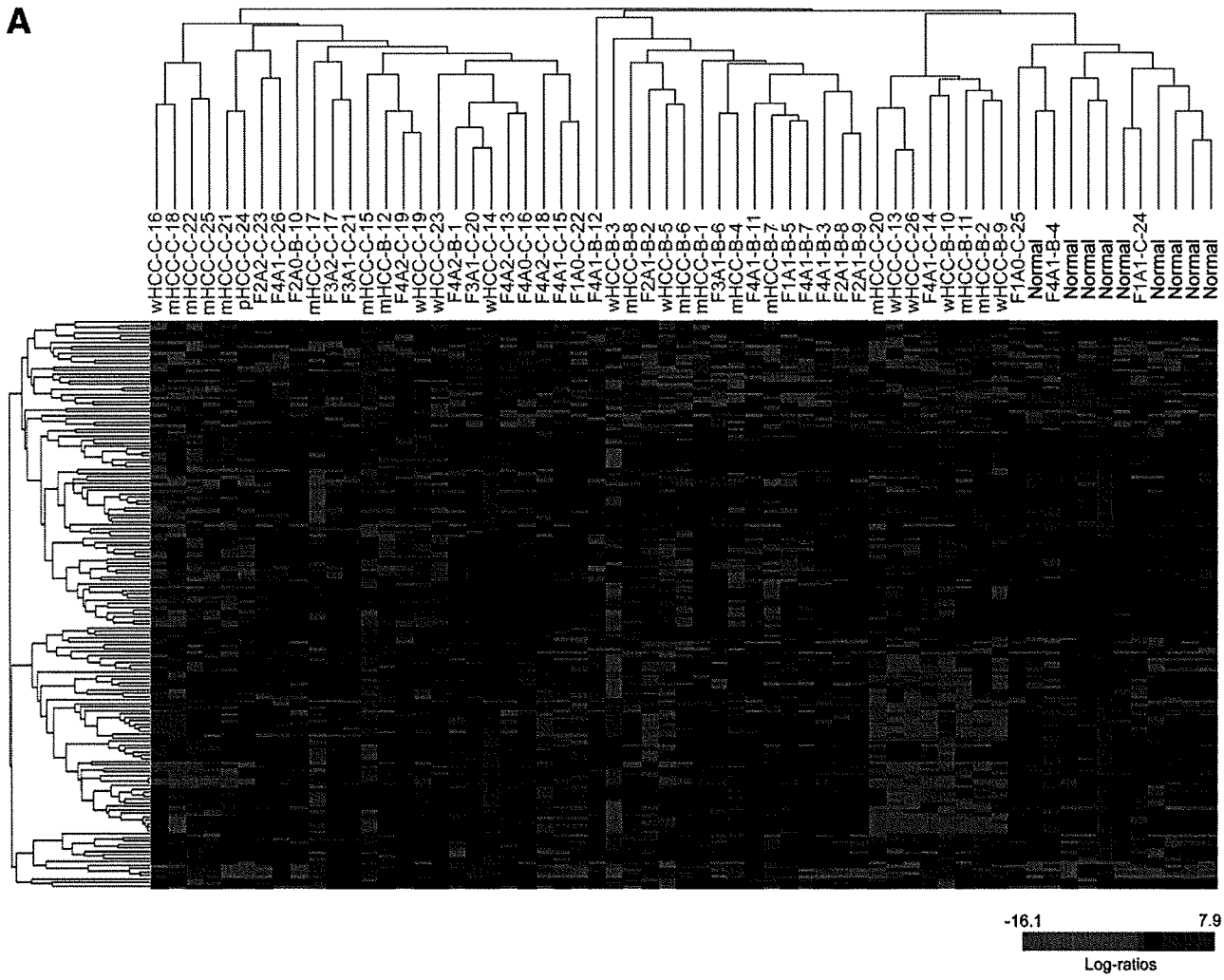

Fig. 2. (A) Hierarchical cluster analysis using total miRNA. Chronic hepatitis is indicated by histological stage and grade (F, fibrosis; A, activity) and type of infecting virus (B or C). HCC is indicated by histological grade (w, well differentiated; m, moderately differentiated; p, poorly differentiated) 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).

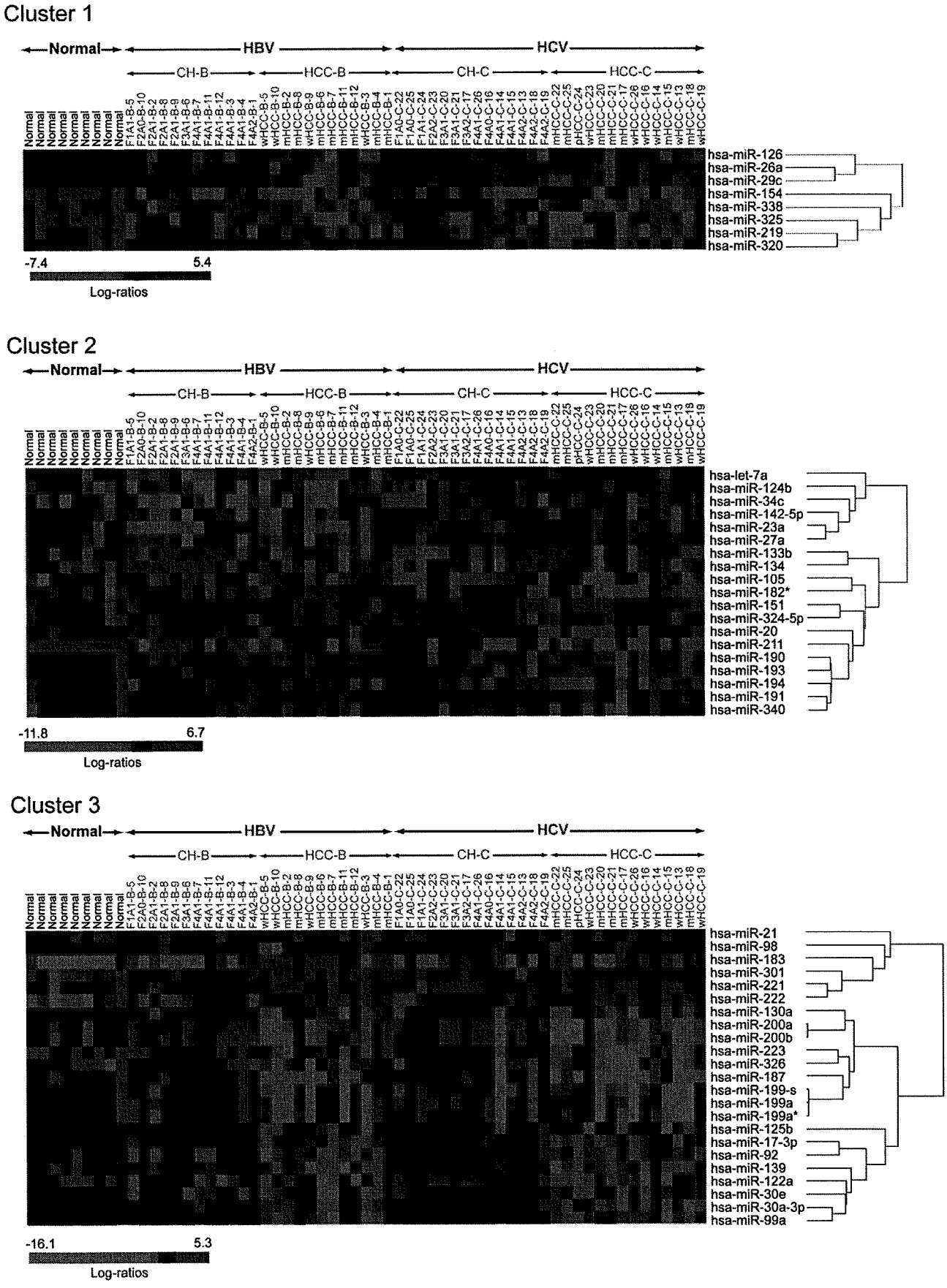

Fig. 3. Cluster 1: Eight miRNAs specifically differentiated node 1 classification. Cluster 2: Nineteen miRNAs specifically differentiated node 2 classification. Cluster 3: Twenty-three miRNAs differentiated CH-B and HCC-B as well as CH-C and HCC-C.

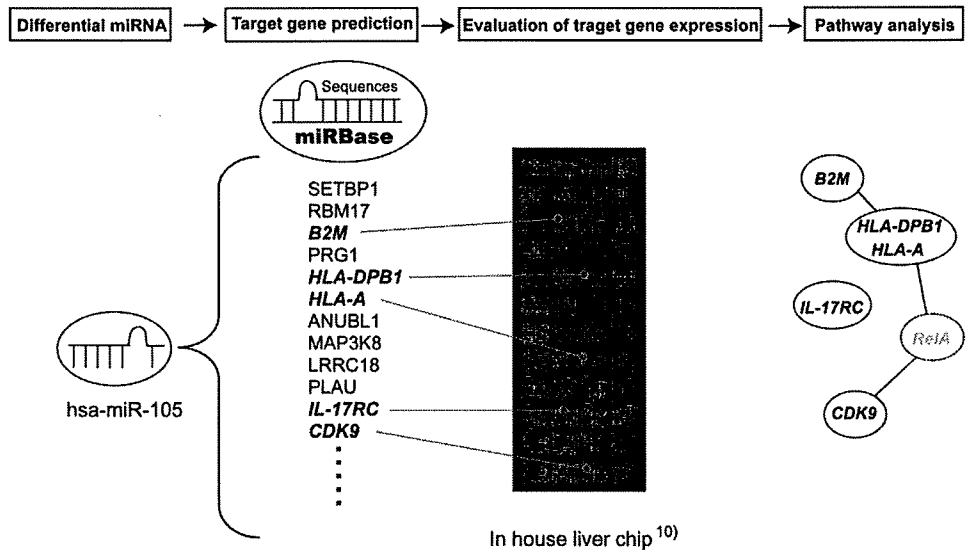

Fig. 4. Analysis of miRNA expression data. Target genes of miRNAs were predicted using MIRANDA Pro3.0; candidate target genes spotted on microarray were identified; number of genes that actually exhibit significant ($P < 0.05$) changes in expression among the genes was determined; and signal pathways involving genes regulated by the miRNAs that had exhibited differential expression between each group were analyzed using MetaCore (Table 4).

noting that the prediction rate may be likely an overestimate of the true rate, given the weaknesses of cross-validation and bootstrapping methods in a strict sense.)

Microarray Analysis. cDNA microarray slides (Liver chip 10k) were used as described.[10] RNA isolation, amplification of antisense RNA, labeling, and hybridization were performed according to the protocols described.[9,10] Quantitative assessment of the signals on the slides was performed by scanning on the ScanArray 5000 (General Scanning, Watertown, MA) followed by image analysis using GenePix Pro 4.1 (Axon Instruments, Union City, CA) as described.[10]

Preliminary Survey of Independency of Paired Samples from the Same Patient. CH and HCC expression data were derived from the same patient. Before further analysis, we examined whether the miRNA expression of paired samples was similar or independent. We compared differences in the expressions of paired and nonpaired CH and HCC samples using the Dunnett test[12] (Supplementary Data). All possible tests performed for data pairs represented no dependency due to the paired data from the same patients. For data analysis, we used the standard pairwise class comparison and prediction tool in BRB ArrayTools.

Identification of Candidate miRNA Target Genes. Candidate target genes predicted to be regulated by miRNAs based on sequence comparison were selected using MIRANDA Pro3.0 (Sanger Institute). Of the selected genes, those represented on a microarray chip were then examined for expression (Fig. 4). The number of genes showing a significant ($P < 0.05$) expression difference among the candidate target genes represented on the chip was statistically analyzed to evaluate the significance of expression regulation by miRNAs. Analysis of significance was performed using Hotelling T2 test (BRB ArrayTools).

Pathway Analysis. Of the candidate miRNA target genes, those showing a significant ($P < 0.01$) expression difference between N, CH-B, HCC-B, CH-C, and HCC-C samples were analyzed for pathways involving these genes using MetaCore software suite (GeneGo, St. Joseph, MI). Significance probability was calculated using

Table 2-1. Class Prediction

No.	Class	Prediction (%)	No. of Predictors	*P* Value
1	HBV versus HCV	87	32	<0.001
2	N versus CH (B+C)	91	26	0.007
3	CH (B+C) versus HCC (B+C)	92	34	0.003

Class prediction algorithm was used for the classification of two groups of patients. Feature selection was based on the univariate significance level (alpha = 0.01). The support vector machine classifier was used for class prediction.

Abbreviations: CH, nontumor lesion of HCC; HCC, hepatocellular carcinoma; N, normal.

Table 2-2 Binary Tree Classification

Node	Group 1 Class	Group 2 Class	No. of Predictors	Misclassification Rate (%)
1	HCC-B, HCC-C, CH-B, CH-C	N	20	4.9
2	HCC-B, CH-B	HCC-C, CH-C	19	13.5
3	HCC-B	CH-B	15	29.2
4	HCC-C	CH-C	14	17.9

Binary tree classification algorithm was used for the classification of each category of patients. Feature selection was based on the univariate significance level (alpha = 0.01). The support vector machine classifier was used for class prediction. There were four nodes in the classfication tree.

Abbreviations: CH-B, non-tumor lesion of HCC-B; CH-C, nontumor lesion of HCC-C; HCC-B, hepatitis B virus-related hepatocellular carcinoma; HCC-C, hepatitis C virus-related hepatocellular carcinoma; N, normal

Table 3-1. Representative miRNAs That Were Commonly Repressed in CH-B, CH-C, HCC-B, and HCC-C Compared with Normal Liver (Cluster)

miRNA	Parametric *P* Value	Ratio*	No. of Significant Genes/Predicted Target Genes†	Hotelling Test *P* Value‡	Differentially Expressed Target Genes§	Pathway of Regulated Genes¶
hsa-miR-219	7.3E-05	0.28	25/109	2.59E-04	Glypican-3, ERP5, PLK2, HIRA, HMG2	Regulatory T cell differentiation
					ACOX1	Fatty acid beta-oxidation
					NF-X1	MHC class II biosynthetic process
hsa-miR-320	9.8E-05	0.50	26/88	3.50E-06	Vimentin, ALP (*N*-acetyltransferase-like), SEC61 beta, G-protein alpha-i2, Filamin A	Protein kinase cascade
					Rac1, RhoG	Organelle organization and biogenesis
					Vinexin beta, Profilin I, Ca-ATPase3	Actin cytoskeleton organization and biogenesis
hsa-miR-154	2.7E-04	0.15	22/70	5.40E-06	OTR, NET1(TSPAN1), NAP1, Vimentin, PDIA3, cytochrome P-450 reductase	Regulation of apoptosis
					DLX2	Morphogenesis
					GUAC, ACAT1	Branched chain family amino acid catabolic process
hsa-miR-29c	1.8E-03	0.55	53/133	1.00E-06	FBXO7, ASPP1, HSPA4, Cathepsin O, PDF, COL4A1, HSPA4, TIP30, CXADR	Cell-substrate adhesion
					NS1-BP, ALP (*N*-acetyltransferase-like), ACTR10, Beclin 1	Transcription, DNA-dependent
					SMAD6, LTBR(TNFRSF3), ENPP7	Apoptosis
hsa-miR-338	5.2E-03	0.46	30/101	3.60E-06	ID3, GATA-4, NFIA, FR-beta, CREST, HYOU1	Developmental process
					G3ST1, CAD,	Immune effector process
					FKBP12, LZIP, PDIA3, Schwannomin (NF2), CREST	Immune system process
hsa-miR-26a	6.3E-03	0.70	37/119	2.64E-05	*LIG4*, c-FLIP, GADD45 beta, DAPK1, PRDX4, LRP130	Response to stimulus
					Cyclin E, ZDHHC6, Tx1, ATG8 (GATE-16), WASP, C1s	DNA replication initiation
					COPG1	Ion transport
hsa-miR-126	8.1E-03	0.65	27/101	4.04E-03	ANP32B (april), HSPA4, RLI, LIV-1 (SLC39A6), PTP-MEG2, CD97, DHPR	Regulation of cellular protein metabolic process
					NFKBIA, NMI, MDH1, PDCD2	Response to stress
					SMAD6, ATP6AP2, ANP32B (april), NMI, HSPA4	Apoptosis
hsa-miR-325	8.7E-03	0.20	18/63	2.03E-04	TRADD, CREST, NEDD8, annexin IV, GPX2, PDF, TNFAIP1	Developmental process
					Glypican-3, ID1, PC-TP,	Multicellular organismal development
					SNRPB (Sm-B)	RNA splicing

*Ratio of HCC-B, HCC-C, CH-B, and CH-C to normal.
†The number of significant genes ($P < 0.05$) out of predicted target genes in which expression was evaluated in microarray.
‡Statistical assessment of presence of differentially expressed genes out of predicted target genes of miRNAs.
§Representative differentially expressed genes out of predicted target genes of miRNAs.
¶Representative pathway of differentially expressed genes out of predicted target genes of miRNAs.

the hypergeometrical distribution based on gene ontology terms. Because one gene is frequently involved in multiple pathways, all pathways corresponding to the genes with significance probability were listed.

Verification of Regulation of Candidate Target Genes by miRNAs. Anti-miRNA (Ambion) specific to 13 miRNAs (has-miR-17*, has-miR-20a, has-miR-23a, has-miR-26a, has-miR-27a, has-miR-29c, has-miR-30a, has-miR-92, has-miR-126, has-miR-139, has-miR-187, has-miR-200a, and has-miR-223) showing significant differences in expression were transfected into Huh7 cells using TransMessenger transfection reagent (QIAGEN, Valencia, CA), and loss of function of each miRNA was evaluated. Similarly, precursor miRNAs of five miRNAs (has-miR-23a, has-miR-26a, has-miR-27a, has-miR-92, and has-miR-200a) were also transfected into Huh7 cells, and gain of function of each miRNA was evaluated. The loss- and gain-of-function of miRNAs were evaluated via RTD-PCR. In addition, different gene expressions regulated by miRNAs were also evaluated via RTD-PCR.

HBV/HCV Infection Model Using Cultured Cells. The plasmid pHBV 1.2 coding the 1.2-fold length of the HBV genome was transfected into Huh7.5 cells using Fugene6 transfection reagent (Roche Applied Science, Indianapolis, IN). HBeAg production in culture medium was measured using Immunis HBeAg/Ab EIA (Institute of Immunology Co., Ltd., Tokyo, Japan).[13] The amount of HBV-DNA was measured via RTD-PCR (Supplementary Fig. 1A,B). JFH1-RNA was transfected into Huh7.5 cells using TransMessenger transfection reagent (QIAGEN) and the expression of the core protein was examined via immunofluorescence staining using anti-HCV core antibody (Affinity BioReagent, CO).[14,15] HCV-RNA amount was also measured via RTD-PCR (Supplementary Fig. 1A,B). JFH1/GND was used as a negative control. miRNA expression was quantitated by RTD-PCR 48 hours after transfection.

Results

Expression of miRNA in Liver Tissue. A panel of miRNA was successfully amplified from liver tissues via RTD-PCR. The representative amplification profile of miRNA as determined with RTD-PCR is shown in Fig. 1. To assess the reliability and reproducibility of this assay system, we first measured RNU6B in duplicate from all samples in different plates. The mean difference in Ct values of RNU6B expression within the same samples was 0.08 ± 0.05 (mean ± standard deviation), indicating the high reproducibility of this assay. All Ct values from each reaction were collected, and Ct variation obtained by each probe from all patients was calculated. Although RNU6B was frequently used as the internal control, the standard Ct variation was relatively high (Ct, 27 ± 1.94), suggesting that the variances in its value depend on the state of liver disease (N, CH and HCC). Therefore, we selected has-miR-328 as the internal control with the smallest standard deviation (Ct, 30 ± 0.60). The relative expression ratio of individual miRNA to has-miR-328 was calculated and applied to the following analysis using a BRB-array tool.

Hierarchical cluster analysis revealed that the expression profiles of the 188 miRNAs from each patient were roughly classified into normal liver, HBV-infected liver (CH-B+HCC-B; HBV group), and HCV-infected liver (CH-C+HCC-C; HCV group) (Fig. 2A). HCV viremia in two patients with CH-C was persistently cleared by interferon therapy before HCC development. The background liver of one of these patients was clustered in the normal group and those of others in the HCV group. Although these two patients were not clearly differentiated from others, some miRNAs such as miR-194, miR-211, and miR-340 that were down-regulated in the HCV group were significantly up-regulated in two patients (Fig. 3, cluster 2).

The present CH and HCC expression data were obtained from the same patient; however, each sample clustered irrespective of pairs in all but two patients. miRNA expression profiling was therefore more dependent on the disease condition than on the paired condition, as also confirmed by the Dunnett test.[12] We then attempted to classify the expression profiles into HBV and HCV groups using supervised learning methods (Table 2-1). HBV and HCV groups were significantly differentiated at an 87% accuracy ($P < 0.001$). The normal liver and CH (CH-B + CH-C) and CH and HCC (HCC-B + HCC-C) were also significantly differentiated at a 90% rate of accuracy. These results suggest that different stages of liver disease (normal, CH, and HCC) can be differentiated from each other based on the miRNA expression profile, as well as HBV and HCV infection.

To examine the relationship among five categories of groups, namely, N, CH-B, CH-C, HCC-B and HCC-C, we attempted to differentiate the five groups using a supervised learning algorithm (binary tree classification) used for classifying three or more groups. SVM was used as a prediction method. Expression profiles were first classified into groups N (normal) and non-N (non-normal) (CH-C, CH-B, HCC-C, and HCC-B) (node 1) ($P < 0.01$). The non-N group was then classified into HBV and HCV (node 2) ($P < 0.01$). The HBV group was further classified into CH-B and HCC-B (node 3) ($P < 0.01$), and the HCV group was further classified into CH-C and HCC-C (node 4) ($P < 0.01$) (Fig. 2B, Table 2-2). Thus, the findings support the notion that differences in miRNA expression between HBV and HCV are as distinct as those between CH and HCC.

Out of 20 miRNAs that differentiated node 1 classification (Table 2-2), 12 also differentiated node 3 or node 4 classification. The remaining eight miRNAs specifically differentiated node 1 classification. They were down-regulated in the HBV and HCV groups compared with the normal group (Fig. 3, cluster 1). Nineteen miRNAs differentiated node 2 classification (Table 2-2) and the hierarchical clustering using these miRNAs clearly differentiated the HBV and HCV groups (Fig. 3, cluster 2). There were 15 and 14 miRNAs that differentiated node 3 and 4 classifications, respectively (Table 2-2). Hierarchical clustering using these miRNAs revealed that these miRNAs differentiated CH-B and HCC-B as well as CH-C and HCC-C, respectively; 17 miRNAs were down-regulated in HCC, and six were up-regulated in HCC (Fig. 3, cluster 3).

Table 3-2. Differentially Expessed miRNA Between HCC-B, CH-B, and HCC-C, CH-C, and Their Representative Target Genes (Cluster 2)

miRNA	Parametric *P* Value	Ratio*	No. of Significant Genes/Predicted Target Genes†	Hotelling Test *P* Value‡	Differentially Expressed Target Genes§	Pathway of Regulated Genes¶
hsa-miR-190	1.2E-05	2.06	21/68	4.47E-02	Chk1, C2orf25, VRK2, USP16, STAF65(gamma)	Regulation of cell cycle
					AP1S2, RNASE4	Mitotic cell cycle
					PPP2R1B, ARHGAP15, UBPY	Negative regulation of apoptosis
hsa-miR-134	2.3E-04	5.74	11/58	3.40E-06	VKDGC, SH2B, MALS-1, DDB2	Multicellular organismal process
					BCRP1	Regulation of viral reproduction
					DDB2	Lipid biosynthetic process
hsa-miR-151	2.8E-04	1.82	12/62	6.41E-01	RGS2, UFO, AK2, USP7	G-protein signaling
					eIF4G2, USP7	Regulation of translation
					SLC22A7	Organic anion transport
hsa-miR-193	5.0E-04	1.67	23/95	9.30E-01	G-protein alpha-11, p130CAS, VAV-1, PDCD11	Cell motility
					Colipase, ACSA	Energy coupled proton transport
					DCOR	Intracellular signaling cascade
hsa-miR-133b	1.7E-03	2.42	20/97	3.69E-02	DDB2, Bcl-3, Cystatin B	Proteasomal protein catabolic process
					Rab-3, RAG1AP1, KCNH2, DCOR	Regulation of biological quality
					AL1B1	Carbohydrate metabolic process
hsa-miR-324-5p	2.9E-03	1.51	27/121	1.90E-06	SKAP55, VAV-1, DDB2, E2A, NIP1	Cellular developmental process
					MEMO (CGI-27), Rab-3	Cellular structure morphogenesis
					COPG1, GPX3, OAZ2	Glutathione metabolic process
hsa-miR-182*	3.1E-03	2.23	28/123	< 1e-07	Alpha-endosulfine, HCCR-2, Thioredoxin-like 2, TPT1, USP7	Translation initiation in response to stress
					DDB2, TPT1	Cellular developmental process
					JIP-1	JNK cascade
hsa-miR-105	4.6E-03	4.38	18/68	4.74E-05	Beta-2-microglobulin, HLA-B27	Antigen processing and presentation
					PIMT, IL-17RC	Immune response
					MHC class I, CDK9, ERG1, Desmocollin 3	
hsa-miR-211	5.3E-03	25.61	10/56	2.00E-04	PSMD5, SLC26A6	Proteasomal protein catabolic process
hsa-miR-20	5.7E-03	1.52	27/113	5.28E-03	Noelin, SC4MOL, Thioredoxin-like 2, CCL5, NALP3	Regulation of apoptosis
					Hic-5/ARA55, USP16, MAP4, Ferroportin 1	Positive regulation of cellular process
					TOP3A, PLRP1	Oxygen transport
hsa-miR-191	6.7E-03	1.39	25/79	7.55E-04	CDK9, GPS2, CLTA, LXR-alpha	Nucleic acid metabolic process
					ACSA	Acetyl-CoA biosynthetic process
					UGCGL1, SGPP1	Metal ion transport
hsa-miR-340	8.5E-03	1.48	17/81	3.73E-03	FKBP12, DCOR,	Calcium ion transport
					Gelsolin, VAV-1, ARF6	Actin cytoskeleton organization and biogenesis
					HXK3	Glucose catabolic process
hsa-miR-194	8.7E-03	1.67	13/74	5.90E-01	Cyclin B1, Serglycin	M phase of mitotic cell cycle
					PTE2	Acyl-CoA metabolic process
					SLC7A6	Carbohydrate utilization
hsa-miR-23a	1.9E-04	0.46	14/97	< 1e-07	*RGL2*, MANR, MEK1 (MAP2K1), Caspase-3, AZGP1	Protein kinase cascade
					FRK, Pyk2(FAK2), CSE1L	Cellular developmental process
					AZGP1	Defense response
hsa-miR-142-5p	4.9E-04	0.40	25/89	9.10E-06	Sirtuin4, PAI2, PSAT, RIL, CDC34, SPRY1	Metabotropic glutamate receptor
					E4BP4, DNAJC12, WWP1, PAIP1, PASK, rBAT	Regulation of gene expression
					VCAM1, CaMK I, WWP1, FHL3	Cell-matrix adhesion
hsa-miR-34c	5.1E-04	0.20	31/129	7.30E-06	Diacylglycerol kinase, zeta, PLC-delta 1, ATP2C1, PAI2	Manganese ion transport
					MLK3(MAP3K11), MEK1(MAP2K1), CDC25C, MRF-1, XPC	Protein kinase cascade
					GNT-IV	Inflammatory cell apoptosis

Table 3-2. Continued

miRNA	Parametric *P* Value	Ratio*	No. of Significant Genes/Predicted Target Genes†	Hotelling Test *P* Value‡	Differentially Expressed Target Genes§	Pathway of Regulated Genes¶
hsa-miR-124b	8.6E-04	0.32	25/120	7.10E-05	E2F5, Rad51, Jagged1	Muscle development
					MLK3(MAP3K11), RGS1	Intracellular signaling cascade
					COL16A1	MAPKKK cascade
hsa-let-7a	1.0E-03	0.45	28/136	9.35E-04	RAD51C, CoAA, hASH1, Cockayne syndrome B, Caspase-1, PP5	Response to DNA damage stimulus
					PLC-delta 1, MANR, ACADVL	Fibroblast proliferation
					HGF, NGF	Cellular developmental process
hsa-miR-27a	3.9E-03	0.59	18/108	1.19E-02	COL16A1, RIL, RhoGDI gamma, ANP32B (april)	Cytoskeleton organization and biogenesis
					VE-cadherin, NTH1, GATA-2, E4BP4	Response to external stimulus
					RAD51C	DNA recombination

*Ratio of HCC-B, CH-B, to HCC-C,CH-C.
†The number of significant genes ($p < 0.05$) out of predicted target genes in which expression was evaluated in microarray.
‡Statistical assesment of presence of differentially expressed genes out of predicted target genes of miRNAs.
§Representative differentially expressed genes out of predicted target genes of miRNAs.
¶Representative pathway of differentially expressed genes out of predicted target genes of miRNAs.

These results indicate that there were two types of miRNAs—one associated with HBV and HCV infection (cluster 2), the other associated with the stages of liver disease (clusters 1 and 2) that were irrelevant to the differences in HBV and HCV infection.

Differential miRNAs and Their Candidate Target Genes and Signaling Pathways. Differentially expressed miRNAs are shown in Table 3. In addition to the expression ratios of miRNAs in each group, the number of genes analyzed on the microarray predicted to be the target genes of miRNAs and that which actually showed significant ($P < 0.05$) differences in expression are also shown. Based on the frequencies and levels of expression of differential genes, the significance of regulation of these gene groups by miRNAs was evaluated using Hotelling T2 test (BRB ArrayTools) (Table 3). The representative candidate target genes and their signaling pathways by each miRNA were shown one by one (Table 3). The signaling pathways regulated by all differential miRNAs in each category of groups are shown in Table 4.

Eight miRNAs were down-regulated in the HBV and HCV groups compared with the normal group (Table 3-1; Fig. 3, cluster 1). These miRNAs were associated with an increased expression of genes related to cell adhesion, cell cycle, protein folding, and apoptosis (Tables 3-1, 4-1), and possibly with the common feature of CH irrespective of the differences in HBV and HCV infection.

Nineteen miRNAs clearly differentiated the HBV and HCV groups (Fig. 3, cluster 2, Table 3-2). Thirteen miRNAs exhibited a decreased expression in the HCV group, and six showed a decreased expression in the HBV group. miRNAs exhibiting a decreased expression in the HCV group regulate genes related to immune response, antigen presentation, cell cycle, proteasome, and lipid metabolism. On the other hand, those exhibiting a decreased expression in the HBV group regulate genes related to cell death, DNA damage and recombination, and transcription signals. These findings reflected the differences in the gene expression profile between CH-B and CH-C described (Tables 3-2, 4-2).[10] Interestingly, although these miRNAs were HBV and HCV infection–specific, some of them were reported to be tumor-associated miRNAs, suggesting the possible involvement of infection-associated miRNAs in HCC development.

Twenty-three miRNAs clearly differentiated CH and HCC that were irrelevant to the differences in HBV and HCV infection. Seventeen miRNAs were down-regulated in HCC that up-regulated cancer-associated pathways such as cell cycle, adhesion, proteolysis, transcription, translation, and the Wnt signaling pathway (Tables 3-3, 4-3). Six miRNAs were up-regulated in HCC that down-regulated all inflammation-mediated signaling pathways, potentially reflecting impaired antitumor immune response.

Relationship Between Expressions of Infection-Associated miRNA in Liver and Cultured Cells Infected with HBV and HCV. To clarify whether the expression of infection-associated miRNA is regulated by HBV and HCV infection, we investigated the relationship between changes in miRNA in liver tissues and those in miRNA in Huh7.5 cells in which infectious HBV or HCV clones replicated. To evaluate the replication of each clones in Huh7.5 cells, we measured time-course changes in the amounts of HBV-DNA and HCV-RNA in Huh7.5 cells transfected with pHBV1.2 and JFH1-RNA, respectively, by RTD-PCR (Supplementary Fig. 1A). The expression of HBV proteins was examined by measuring the amount

Table 3-3. Differentially Expessed miRNA Between CH and HCC and Their Representative Target Genes (Cluster 3)

miRNA	Parametric p-value	Ratio*	No. of Significant Genes/Predicted Target Genes†	Hotelling Test *P* Value‡	Differentially Expressed Target Genes§	Pathway of Regulated Genes¶
hsa-miR-139	4.50E-06	0.42	19/106	2.70E-03	Cyclin B1, DHX15, MCM5, Histone H2A	Mitotic cell cycle
					RBCK1, SYHH	Protein catabolic process
					ILK, IGFBP7, SAFB, CTR9	Response to external stimulus
hsa-miR-30a-3p	2.50E-05	0.49	26/144	1.73E-02	GGH, Pirin, ZNF207, Annexin VII	Regulation of oxidoreductase activity
					ILK, LTA4H, ABC50, GNPAT	Cell-matrix adhesion
					DLC1	Morphogenesis
hsa-miR-130a	7.00E-05	0.50	22/108	1.07E-02	SPHM, PPP2R5D, RHEB2, SPHM	Mitotic cell cycle
					MLK3(MAP3K11), Otubain1, TIMP4	Protein modification process
					NRBP	Cell differentiation
hsa-miR-223	3.40E-04	0.39	14/90	6.52E-03	Ephrin-A1, Midkine, FDPS	Cell morphogenesis
					K(+) channel, subfamily J	Notch signaling pathway
hsa-miR-187	3.55E-04	0.12	16/66	6.76E-04	HFE2, Otubain1	Negative regulation of programmed cell death
					PRSS11, SUPT5H, RAG1AP1	Developmental process
					PLOD3	Mitochondrial ornithine transport
hsa-miR-200a	6.86E-04	0.18	20/141	2.15E-02	*CDC25B, KAP3, CDK2AP2,* CHKA	Cell communication
					POLD	DNA replication
					CPSF4	RNA splicing
hsa-miR-17-3p	8.42E-04	0.58	28/108	8.98E-04	MLK3(MAP3K11), Tip60, ACBD6, DOC-1R, DAX1, RBCK1	Protein kinase cascade
					WNT5A, 14-3-3 gamma, DHX15	BMP signaling pathway
					HFE2, MCM5	DNA recombination
hsa-miR-99a	1.17E-03	0.53	33/163	9.52E-03	Calpain small subunit, Thoredoxin-like 2, Survivin	Cytokinesis
					IBP2, DNA-PK, KAP3,	Intracellular signaling cascade
					NFE2L1, PARP-1, HDAC11	Regulatory T cell differentiation
hsa-miR-200b	1.57E-03	0.18	24/147	2.72E-02	HSP47, HMG2, NRBP	Regulation of cell cycle
					SNX17	Cell motility
					Ephrin-A1	Receptor protein signaling pathway
hsa-miR-125b	1.82E-03	0.55	26/114	1.03E-01	COL4A2, TIP30, HSP47, MSP58	Cell adhesion
					MLK3(MAP3K11), ERK2 (MAPK1), ERK1 (MAPK3), PLOD3	Nuclear translocation of MAPK
					Otubain1, SCN4A(SkM1)	Ubiquitin-dependent protein catabolic process
hsa-miR-30e	2.10E-03	0.65	24/151	4.30E-02	Cyclin B1, XTP3B, GAK, Annexin VII, MIC2, NRBP	Mitotic cell cycle
					MSS4	Protein localization
					S100A10	Calcium ion transport
hsa-miR-199a*	4.26E-03	0.35	11/71	7.16E-02	BUB3, Cyclin B1, LMNBR	Mitotic cell cycle
					PRAME	Cardiac muscle cell differentiation
hsa-miR-122a	6.31E-03	0.51	11/80	1.01E-03	JAB1, APEX, Clathrin heavy chain	Base-excision repair
					PARN	Translational initiation
					DDAH2	Regulation of cellular respiration
hsa-miR-199a	8.77E-03	0.35	18/94	3.56E-02	IL-13, MLK3(MAP3K11), CLK2, ACP33	Protein amino acid phosphorylation
					PAFAH beta, SPA1, CLCN4	Small GTPase mediated signal transduction
hsa-miR-326	9.00E-03	0.57	29/147	2.25E-01	Midkine, ENT1, IP3KA, PSMC5, ANCO-1	Regulation of programmed cell death
					Thy-1, MCM6, Tip60, VILIP3	Cell-matrix adhesion
					COMP, Cathepsin A	Blood vessel development
hsa-miR-92	9.60E-03	0.81	28/140	2.47E-02	*TUBGCP2,* Fibrillin 1, PIPKI gamma, KAP3	Rho protein signal transduction
					SNX15, BCAT2	LDL receptor and BCAA metabolism
					IGFBP7, FZD6, COPS6	Adenosine receptor signaling pathway
hsa-miR-221	3.40E-06	3.34	16/67	3.59E-01	Lck, Kallistatin, Neuromodulin, LFA-3, PA24A, AZGP1, MSH2	Immune response-activating signal transduction
					KYNU, PMCA3	DNA repair

Table 3-3. *Continued*

miRNA	Parametric p-value	Ratio*	No. of Significant Genes/Predicted Target Genes†	Hotelling Test *P* Value‡	Differentially Expressed Target Genes§	Pathway of Regulated Genes¶
hsa-miR-222	6.50E-06	2.23	18/85	1.59E-02	Thrombospondin 1, Lck, MSH2, ATF-2, CITED2, Kallistatin	Cell motility
					PGAR	Triacylglycerol metabolic process
					KYNU	DNA replication
hsa-miR-301	5.22E-05	1.96	14/71	1.16E-01	Beta-2-microglobulin, PPCKM, PRC, Fra-1, PPCKM, ACAT2	Antigen processing and presentation
					BMPR1B, ARMER, EHM2, RBBP8	Meiotic recombination
					Neuromodulin, LDLR	Cell motility
hsa-miR-21	7.67E-03	1.57	19/81	1.86E-04	Btk, Fra-1, MSH2, Collectrin, Adipophilin	Regulation of T cell proliferation
					RNASE4, AGXT2L1	Peptidyl-tyrosine phosphorylation
					SARDH	Natural killer cell activation during immune response
hsa-miR-183	2.46E-02	3.51	13/86	3.36E-01	Hdj-2, PEMT, Lck, MKP-5, Chondromodulin-I, ABCA8	Cell differentiation
					IL-16, MTRR, SerRS	Methionine biosynthetic process
hsa-miR-98	5.22E-02	1.32	24/130	2.95E-04	ACAA2, LTB4DH, ACADVL, DECR, S14 protein,	Fatty acid metabolic process
					Rapsyn, Kallistatin, ENPEP, Beta crystallin B1	Multicellular organismal process
					CYP4F8	Prostaglandin metabolic process

*Ratio of HCC to CH.
†The number of significant genes ($P<0.05$) out of predicted target genes in which expression was evaluated in microarray.
‡Statistical assesment of presence of differentially expressed genes out of predicted target genes of miRNAs.
§Representative differentially expressed genes out of predicted target genes of miRNAs.
¶Representative pathway of differentially expressed genes out of predicted target genes of miRNAs.

of HBeAg released in culture medium (Supplementary Fig. 1B). HCV protein expression was examined by evaluating the core protein expression after 48 hours by fluorescence immunostaining (Supplementary Fig. 1C). RNA was extracted from the Huh7.5 cells 48 hours after gene transfection, and miRNA expression pattern in the cells was compared with those in liver tissues. We found a strong correlation between differences in miRNA expression between liver tissues of the HBV and HCV groups, and those in miRNA expression between Huh7.5 cells transfected with HBV and HCV clones ($r = 0.73$, $P = 0.0006$) (Fig. 5). These results revealed that differences in the expression of infection-associated miRNA in the liver between the HBV and HCV groups are explained by changes in miRNA expression caused by HBV and HCV infections.

Verification of Regulation of Candidate Target Genes by miRNA. Anti-miRNAs (Ambion) specific to 13 miRNAs (has-miR-17*, has-miR-20a, has-miR-23a, has-miR-26a, has-miR-27a, has-miR-29c, has-miR-30a, has-miR-92, has-miR-126, has-miR-139, has-miR-187, has-miR-200a, and has-miR-223) showing significant differences in expression were transfected into Huh7 cells to examine loss of function of the miRNAs. Five miRNAs (has-miR-23a, has-miR-26a, has-miR-27a, has-miR-92, and has-miR-200a) showed a decreased expression by more than 50%. Precursor miRNAs of these miRNAs were also transfected into the cells to examine the gain of function of the miRNAs (Supplementary Fig. 2). It was confirmed that the expressions of target genes of the five miRNAs (LIG4 [by has-miR-26a]; RGL2 [by has-miR-23a]; Rad51C [by has-miR-27a]; KAP3, CDC25B, KAP3, CDK2AP2, POLD, and CPSF4 [by has-miR-200a]; and TUBGCP2, SNX15 and BCAT2 [by has-miR-92]) were increased by the suppression of the miRNAs induced by anti-miRNAs and were decreased by the overexpression of precursor miRNAs (Supplementary Fig. 3).

Discussion

miRNA plays an important role in various diseases such as infection and cancer.[1-3] In this study, we examined miRNA expression profiles in normal liver and HCC, including nontumor lesions infected with HBV or HCV. Although the expression profiles of miRNAs in HCC have been reported,[16-18] most of the studies were performed using a microarray system. Because we thought that miRNAs could not produce enough detection signals owing to their short length, we applied a highly sensitive and quantitative RTD-PCR method for miRNAs. Moreover, global gene expression in the same tissues was ana-