of the correlation coefficient between the variables is needed, and therefore, GGM can be easily applied to a wide variety of data. However, straightforward applications of statistical theory to practical data fail in some cases, and GGM also fails frequently when applied to gene expression profiles; here the expression profile indicates a set of the expression degrees of one gene, measured under various conditions. This is because the profiles often share similar expression patterns, which indicate that the correlation coefficient matrix between the genes is not regular. Thus, we have devised a procedure, named ASIAN (automatic system for inferring a network), to apply GGM to gene expression profiles, by a combination of hierarchical clustering [14]. First, the large number of profiles is grouped into clusters, according to the standard approach of profile analysis [15]. To avoid the generation of a nonregular correlation coefficient matrix from the expression profiles, we adopted a stopping rule for hierarchical clustering [10]. Then, the relationship between the clusters is inferred by GGM. Thus, our method generates a framework of gene regulatory relationships by inferring the relationships between the clusters [11, 16], and provides clues toward estimating the global relationships between genes on a large scale.

Methods for extracting biological knowledge from large amounts of literature and arranging it in terms of gene function have been developed. Indeed, ontologies have been made available by the gene ontology (GO) consortium [17] to construct a functional categorization of genes and gene products, and by using the GO terms, the software determines whether any GO terms annotate a specified list of genes at a frequency greater than that expected by chance [18]. Furthermore, various software applications, most of which are commercial software, such as MetaCore from GeneGo http://www.genego.com/, have been developed for the navigation and analysis of biological pathways, gene regulation networks, and protein interaction maps [19]. Thus, advances in the processing of biological knowledge have enabled us to correspond to the results of gene expression analyses for a large amount of data with the biological functions.

In this study, we analyzed the gene expression profiles from the CHC and HCC cell stages, by ASIAN based on the graphical Gaussian Model, to reveal the framework of gene group associations in hepatocellular carcinogenesis. For this purpose, first, the genes characteristically expressed in hepatocellular carcinogenesis were selected, and then, the profiles of the genes thus selected were subjected to the association inference method. In addition to the association inference, which was presented by the network between the clusters, the network was further interpreted systematically by the biological knowledge of the gene interactions and by the functional categories with GO terms. The combination of the statistical network inference from the profiles with the systematic network interpretation by the biological knowledge in the literature provides a snapshot of the orchestration of gene systems in hepatocellular carcinogenesis, especially for bridging the gap between the information on the disease mechanisms at the molecular level and at more macroscopic levels.

## 2. MATERIALS AND METHODS

### 2.1. Gene selection

We selected the up- and downregulated genes characteristically expressed in the CHC and HCC stages, as a prerequisite for defining the variables in the network inference by the graphical Gaussian modeling. This involved the following steps. (1) The averages and the standard deviations in the respective conditions, $AV_j$ and $SD_j$, for $j = 1, \ldots, N_c$, are calculated. (2) The expression degree of the $i$th gene in the $j$th condition, $e_{ij}$, is compared with $|AV_j \pm SD_j|$. (3) The gene is regarded as a characteristically expressed gene, if the number of conditions that $e_{ij} \geq |AV_j \pm SD_j|$ is more than $N_c/2$. Although the criterion for a characteristically expressed gene is usually $|AV_j \pm 2SD_j|$, the present selection procedure described above is simply designed to gather as many characteristically expressed genes as possible, and is suitable to capture a macroscopic relationship between the gene systems estimated by the following cluster analysis.

### 2.2. Gene systems network inference

The present analysis is composed of three parts: first, the profiles selected in the preceding section are subjected to the clustering analysis with the automatic determination of cluster number, and then the profiles of clusters are subjected to the graphical Gaussian modeling. Finally, the network inferred by GGM is rearranged according to the magnitude of partial correlation coefficients, which can be regarded as the association strength, between the clusters. The details of the analysis are as follows.

#### 2.2.1. Clustering with automatic determination of cluster number

In clustering the gene profiles, here, the Euclidian distance between Pearson's correlation coefficients of profiles and the unweighted pair group method using arithmetic average (UPGMA or group average method) were adopted as the metric and the technique, respectively, with reference to the previous analyses by GGM [11, 16]. In particular, the present metric between the two genes is designed to reflect the similarity in the expression profile patterns between other genes as well as between the measured conditions, that is,

$$d_{ij} = \sqrt{\sum_{l=1}^{n} (r_{il} - r_{jl})^2}, \tag{1}$$

where $n$ is the total number of the genes, and $r_{ij}$ is the Pearson correlation coefficient between the $i$ and $j$ genes of the expression profiles that are measured at $N_c$ conditions, $p_{ik}$, $(k = 1, 2, \ldots, N_c)$:

$$r_{ij} = \frac{\sum_{k=1}^{l} (p_{ik} - \overline{p_i}) \cdot (p_{jk} - \overline{p_j})}{\sqrt{\sum_{k=1}^{l} (p_{ik} - \overline{p_i})^2 \cdot \sum_{k=1}^{l} (p_{jk} - \overline{p_j})^2}}, \tag{2}$$

where $\overline{p_i}$ is the arithmetic average of $p_{ik}$ over $N_c$ conditions.

In the cluster number estimation, various stopping rules for the hierarchical clustering have been developed [20]. Recently, we have developed a method for estimating the cluster number in the hierarchical clustering, by considering the following application of the graphical model to the clusters [10]. In our approach, the variance inflation factor (VIF) is adopted as a stopping rule, and is defined by

$$\mathrm{VIF}_i = r_{ii}^{-1}, \tag{3}$$

where $r_{ii}^{-1}$ is the $i$th diagonal element of the inverse of the correlation coefficient matrix between explanatory variables [21]. In the cluster number determination, the popular cutoff value of 10.0 [21] was adopted as a threshold in the present analysis, also with reference to the previous analyses.

After the cluster number determination, the average expression profiles are calculated for the members of each cluster, and then the average correlation coefficient matrix between the clusters is calculated from them. Finally, the average correlation coefficient matrix between the clusters is subjected to the graphical Gaussian modeling. Note that the average coefficient correlation matrix avoids the difficulty of the above numerical calculation, due to the distinctive patterns of the average expression profiles of clusters. This means that the GGM works well for the average coefficient correlation matrix.

### 2.2.2. *Graphical Gaussian modeling*

The concept of conditional independence is fundamental to graphical Gaussian modeling (GGM). The conditional independence structure of the data is characterized by a conditional independence graph. In this graph, each variable is represented by a vertex, and two vertices are connected by an edge if there is a direct association between them. In contrast, a pair of vertices that are not connected in the graph is conditionally independent.

In the procedure for applying the GGM to the profile data [11], a graph, $G = (V, E)$, is used to represent the relationship among the $M$ clusters, where $V$ is a finite set of nodes, each corresponding to one of the $M$ clusters, and $E$ is a finite set of edges between the nodes. $E$ consists of the edges between cluster pairs that are conditionally dependent. The conditional independence is estimated by the partial correlation coefficient, expressed by

$$r_{i,j|\mathrm{rest}} = -\frac{r^{ij}}{\sqrt{r^{ii}}\sqrt{r^{jj}}}, \tag{4}$$

where $r_{ij|\mathrm{rest}}$ is the partial correlation coefficient between variables $i$ and $j$, given the rest variables, and $r_{ij}$ is the $(i, j)$ element in the reverse of the correlation coefficient matrix.

In order to evaluate which pair of clusters is conditionally independent, we applied the covariance selection [22], which was attained by the stepwise and iterative algorithm developed by Wermuth and Scheidt [23]. The algorithm is presented as Algorithm 1.

The graph obtained by the above procedure is an undirected graph, which is called an independence graph. The independence graph represents which pair of clusters is conditionally independent. That is, when the partial correlation coefficient for a cluster pair is equal to 0, the cluster pair is conditionally independent, and the relationship is expressed as no edge between the nodes corresponding to the clusters in the independence graph.

The genes grouped into each cluster are expected to share similar biological functions, in addition to the regulatory mechanism [24]. Thus, a network between the clusters can be approximately regarded as a network between gene systems, each with similar functions, from a macroscopic viewpoint. Note that the number of connections in one vertex is not limited, while it is only one in the cluster analysis. This

*Step 1.* Prepare a complete graph of $G(0) = (V, E)$. The nodes correspond to $M$ clusters. All of the nodes are connected. $G(0)$ is called a full model. Based on the expression profile data, construct an initial correlation coefficient matrix $C(0)$.

*Step 2.* Calculate the partial correlation coefficient matrix $P(\tau)$ from the correlation coefficient matrix $C(\tau)$. $\tau$ indicates the number of the iteration.

*Step 3.* Find an element that has the smallest absolute value among all of the nonzero elements of $P(\tau)$. Then, replace the element in $P(\tau)$ with zero.

*Step 4.* Reconstruct the correlation coefficient matrix, $C(\tau + 1)$, from $P(\tau)$. In $C(\tau + 1)$, the element corresponding to the element set to zero in $P(\tau)$ is revised, while all of the other elements are left to be the same as those in $C(\tau)$.

*Step 5.* In the Wermuth and Sheidt algorithm, the termination of the iteration is judged by the "deviance" values. Here, we used two types of deviance, dev1 and dev2, with the following:

$$\begin{aligned} \mathrm{dev1} &= N_c \log\left(\frac{|C(\tau+1)|}{|C(0)|}\right), \\ \mathrm{dev2} &= N_c \log\left(\frac{|C(\tau+1)|}{|C(\tau)|}\right). \end{aligned} \tag{5}$$

Calculate dev1 and dev2. The two deviances follow an asymptotic $\chi^2$ distribution with a degree of freedom $= n$, and that with a degree of freedom $= 1$, respectively. $n$ is the number of elements that are set to zero until the $(\tau+1)$th iteration. In our approach, $n$ is equal to $(\tau + 1)$. $|C(\tau)|$ indicates the determinant of $C(\tau)$. $N_c$ is the number of different conditions under which the expression levels of $M$ clusters are measured.

*Step 6.* If the probability value corresponding to dev1 $\leq 0.05$, or the probability value corresponding to dev2 $\leq 0.05$, then the model $C(\tau + 1)$ is rejected, and the iteration is stopped. Otherwise, the edge between a pair of clusters with a partial correlation coefficient set to zero in $P(\tau)$ is omitted from $G(\tau)$ to generate $G(\tau + 1)$, and $\tau$ is increased by 1. Then, go to Step 1.

ALGORITHM 1

feature of the network reflects the multiple relationships of a gene or a gene group in terms of the biological function.

#### 2.2.3. Rearrangement of the inferred network

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 by GGM still showed complicated forms with many edges [11, 16]. Since the magnitude of the partial correlation coefficient indicates the strength of the association between clusters, the intact network can be rearranged according to the partial correlation coefficient value, to interpret the association between clusters. The strength of the association can be assigned by a standard test for the partial correlation coefficient [25]. By Fisher's $Z$ transformation of partial correlation coefficients, that is,

$$Z = \frac{1}{2}\log\left(\frac{1 + r_{ij\cdot\text{rest}}}{1 - r_{ij\cdot\text{rest}}}\right), \tag{6}$$

$Z$ is approximately distributed according to the following normal distribution:

$$N\left(\frac{1}{2}\log\left(\frac{1 + r_{ij\cdot\text{rest}}}{1 - r_{ij\cdot\text{rest}}}\right), \frac{1}{\{N_c - (M-2)\} - 3}\right), \tag{7}$$

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.

### 2.3. Statistical significance of the inferred network with the biological knowledge

The inferred network can be statistically evaluated in terms of the gene-gene interactions. The chance probability was estimated by the correspondence between the inferred cluster network and the information about gene interactions. The following steps were used. (1) The known gene pairs with interactions in the database were overlaid onto the inferred network. (2) The number of cluster pairs, upon which the gene interactions were overlaid, was counted. (3) The chance probability, in which the cluster pairs connected by the established edges in the network were found in all possible pairs, was calculated by using the following equation:

$$P = 1 - \sum_{i=0}^{f-1} \frac{\binom{g}{i}\binom{N-g}{n-i}}{\binom{N}{n}}, \tag{8}$$

where $N$ is the number of possible cluster pairs in the network, $n$ is the number of cluster pairs with edges in the inferred network, $f$ is the number of cluster pairs with edges in the inferred network, including the known gene pairs with interactions, and $g$ is the number of cluster pairs, including the known gene pairs with interactions.

### 2.4. Evaluation of the inferred network in terms of the biological knowledge

The inferred network can be evaluated in terms of the biological knowledge. For this purpose, we characterize the clusters by GO terms, and overlay the knowledge about the gene interactions onto the network. For this purpose, we first use GO::TermFinder [18] to characterize the clusters by GO terms with the user-defined significance probability (http://search.cpan.org/dist/GO-TermFinder). Then, Pathway Studio [19] is used to survey the biological information about the gene interactions between the selected genes.

### 2.5. Software

All calculations of the present clustering and GGM were performed by the ASIAN web site [26, 27] (http://www.eureka.cbrc.jp/asian) and "Auto Net Finder," the commercialized PC version of ASIAN, from INFOCOM CORPORATION, Tokyo, Japan (http://www.infocom.co.jp/bio/download).

### 2.6. Expression profile data

The expression profiles of 8516 genes were monitored in 27 CHC samples and 17 HCC samples [28].

## 3. RESULTS AND DISCUSSION

### 3.1. Clustering

Among the 8516 genes with expression profiles that were measured in the previous studies [28], 661 genes were selected as those characteristically expressed in the CHC and HCC stages. As a preprocessing step for the association inference, the genes thus selected were automatically divided into 18 groups by ASIAN [26, 27]. Furthermore, each cluster was characterized in terms of the GO terms, which define the macroscopic features of the cluster in terms of the biological function.

Figure 1 shows the dendrogram of clusters, together with their expression patterns. As seen in Figure 1, the genes were grouped into 18 clusters, in terms of the number of members and the expression patterns in the clusters. The average number of cluster members was 36.7 genes (SD, 14.2), and the maximum and minimum numbers of members were 69 in cluster 14 and 18 in cluster 9, respectively. As for the expression pattern, five clusters (10, 12, 14, 15, and 18) and ten clusters (1–7, 9, 16, and 17) were composed of up- and downregulated genes, respectively, and three clusters (8, 11, and 13) showed similar mixtures of up- and downregulated genes.

Table 1 shows the GO terms for the clusters (clusterGOB), which characterized them well (see details at http://www.cbrc.jp/~horimoto/HCGO.pdf). Among the 661 genes analyzed in this study, 525 genes were characterized by the GO terms, and among the 18 clusters, 11 clusters were characterized by GO terms with $P < .05$. In addition, 188 genes (28.3% of all characterized genes) corresponded to the GO terms listed in Table 1. As seen in the table, although

most clusters are characterized by several GO terms, reflecting the fact that the genes function generally in multiple pathways, the clusters are not composed of a mixture of genes with distinctive functions. For example, cluster 2 is characterized by 10 terms, and most of the terms are related to the energy metabolism. Thus, the GO terms in the respective clusters share similar features of biological functions, which cause the hierarchical structure of the GO term definitions.

In Table 1, most of the clusters characterized by GO terms with $P < .05$ are related to response function and to metabolism. Clusters 1, 6, 8, 12, and 13 are characterized by GO terms related to different responses, and clusters 2, 3, 4, and 7 are characterized by GO terms related to different aspects of metabolism. Although the genes in two clusters, 14 and 16, did not adhere to this dichotomy, the genes characteristically expressed in HCC in the above nine clusters were related to the responses and the metabolic pathways. As for the remaining clusters with lower significance, three clusters (9, 10, and 11) were also characterized by response functions, and four clusters (5, 15, 17, and 18) were related to morphological events at the cellular level. Note that none of the clusters characterized by cellular level events attained the significance level. This may be because the genes related to cellular level events represent only a small fraction of genes relative to all genes with known functions, in comparison with the genes related to molecular level events in the definition of GO terms.

It is interesting to determine the correspondence between the up- and downregulated genes and the GO terms in the clusters. In the five clusters of upregulated genes, clusters 10 and 12 were characterized by different responses, and two clusters were characterized by morphological events, which were the categories of "cell proliferation" in cluster 15 and of "development" in cluster 18. The remaining cluster, 14, was characterized by regulation, development, and metabolism. As for the clusters of downregulated genes, four of the ten clusters were characterized by GO terms related to various aspects of metabolism. In the remaining six clusters, three clusters were characterized by GO terms related to responses, two clusters were characterized by morphological events, and one cluster was characterized by mixed categories.

In summary, the present gene selection and the following automatic clustering produced a macroscopic view of gene expression in hepatocellular carcinogenesis. Although the clusters contain many genes that do not always share the same functions, the clusters were characterized by their responses, morphological events, and metabolic aspects from a macroscopic viewpoint. The clusters of upregulated genes were characterized by the former two categories, and those of the downregulated genes represented all three categories. Thus, the present clustering serves to interpret the network between the clusters in terms of the biological function and the gene expression pattern.

### 3.2. *Known gene interactions in the inferred network*

The association between the 18 clusters inferred by GGM is shown in Figure 2. In the intact network by ASIAN, 96 of 153 possible edges between 18 clusters (about 63%) were established by GGM. Since the intact network is still messy, the network was rearranged to interpret its biological meaning by extracting the relatively strong associations between the clusters, according to the procedure in Section 2.2.3. After the rearrangement, 34 edges remained by the statistical test of the partial correlation coefficients with 5% significance. In the rearranged network, all of the clusters were nested, but each cluster was connected to a few other clusters. Indeed, the average number of edges per cluster was 2.3, and the maximum and minimum numbers of edges were seven in cluster 15 and one in cluster 9, respectively. In particular, the numbers of edges are not proportional to the numbers of constituent genes in each cluster. For example, while the numbers of genes in clusters 15 and 17 are equal to each other (24 genes), the number of edges from cluster 15 (2 edges) differs from that from cluster 17 (5 edges). Thus, the number of edges does not depend on the number of genes belonging to the cluster, but rather on the gene associations between the cluster pairs.

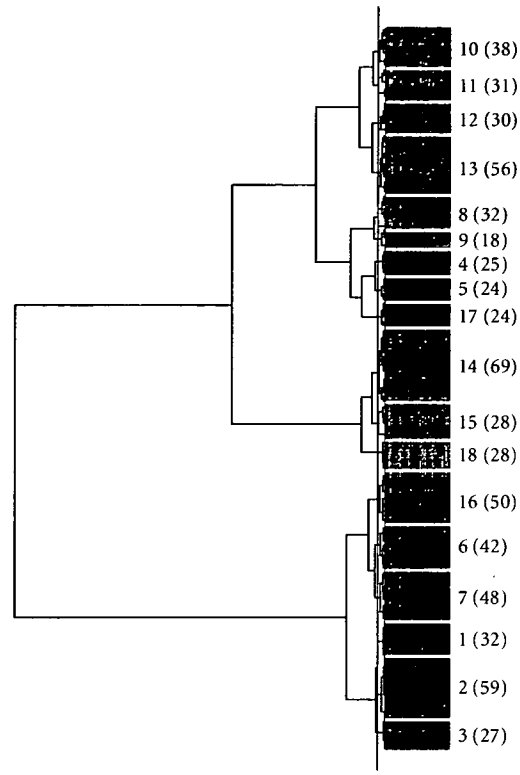


FIGURE 1: *Dendrogram of genes and profiles.* The dendrogram was constructed by hierarchical clustering with the metric of the Euclidian distances between the correlation coefficients and the UPGMA. The blue line on the dendrogram indicates the cluster boundary estimated automatically by ASIAN. The gene expression patterns of the respective clusters in the CHC and HCC stages are shown by the degree of intensity: the red and green colors indicate relatively higher and lower intensities. The cluster number and the number of member genes in each cluster (in parentheses) are denoted on the right side of the figure.

To test the validity of the inferred network in terms of biological function, the biological knowledge about the gene interactions is overlaid onto the inferred network. For this purpose, all of the gene pairs belonging to cluster pairs are surveyed by Pathway Assist, which is a database for biological knowledge about molecular interactions, compiled based on the gene ontology [17]. Among the 661 genes analyzed in this study, the interactions between 90 gene pairs were detected by Pathway Assist, and 50 of these pairs were found in Figure 2. Notice that the number of gene pairs reported in the literature does not directly reflect the importance of the gene interactions, and instead is highly dependent on the number of scientists who are studying at the corresponding genes. Thus, we counted the numbers of cluster pairs in which at least one gene pair was known, by projecting the gene pairs with known interactions onto the network. By this projection, the interactions were found in 35 ($g$ in the equation of Section 2.3) cluster pairs among 153 ($N$) possible pairs (see details of the gene pair projection at http://www.cbrc.jp/~horimoto/GPPN.pdf). Then, 19 ($f$) of the 35 cluster pairs were overlapped with 34 ($n$) cluster pairs in the rearranged network. The chance probability that a known interaction was found in the connected cluster pairs in the rearranged network was calculated as $P < 10^{-4.3}$. Thus, the rearranged network faithfully captures the known interactions between the constituent genes.

Furthermore, the genes with known interactions were corresponded to the genes responsible for the GO terms of each cluster, as shown in Table 1. The genes responsible for the GO terms were distributed over all cluster pairs, including gene pairs with known interactions, except for only two pairs, clusters 15 and 17, and 15 and 18. Thus, the network can be interpreted not only by the known gene interactions but also by the GO terms characterizing the clusters.

### 3.3. Gene systems network characterized by GO terms

#### 3.3.1. Coarse associations between the clusters

To elucidate the associations between the clusters, the cluster associations with 1% significance probability were further discriminated from those with 5% probability. This generated four groups of clusters, shown in Figure 3(a).

First, we will focus on the groups including the clusters that were characterized by GO terms with a significance probability, and that were definitely occupied by up- or downregulated genes (clusters depicted by triangles with bold lines in the figure). Groups I and III attained the above criteria. In group I, the clusters were a mixture of the clusters of the up- and downregulated genes. Note that three of the six clusters were composed of upregulated genes, which were characterized by responses (cluster 12), mixed categories (cluster 14), and morphological events (cluster 15). In group III, all three clusters were of downregulated genes. One cluster was characterized by responses, and two were characterized by amino-acid-related metabolism. In contrast, groups II and IV were composed of the clusters that were somewhat inadequately characterized by GO terms and expression patterns. Thus, groups I and III provide the characteristic features about the orchestration of gene expression in hepatocellular carcinogenesis.

Secondly, a coarse grinning for group associations provides another viewpoint, shown in Figure 3(b). When the groups with at least one edge between the clusters in the respective groups were presented, regardless of the number of edges, groups I, II, and IV were nested, and group III was connected with only group I. In the second view, group I, which includes three of the five clusters of upregulated genes in all clusters, was associated with all of the other groups. This suggests that group I represents a positive part of the gene expression in hepatocellular carcinogenesis, which is consistent with the interpretation by the first view, from the significant GO terms and the clear expression patterns. Interestingly, among the clusters characterized by morphological events (clusters 5, 15, 17, and 18), three of the four clusters were distributed over groups I, II, and IV, and the distribution was consistent with the nested groups. This suggests that the upregulated genes of the clusters in group I are responsible for the events at the cellular level.

Thirdly, the clusters not belonging to the four groups were clusters 1, 3, and 5. Clusters 1, 3, and 5 were directly connected with groups I, III, and IV, groups I and III, and group IV, respectively. Interestingly, cluster 1, characterized by only "*anti-inflammatory response*," was connected with five clusters belonging to three groups, in which four clusters were downregulated clusters. Although cluster 5 was not clearly characterized by the GO terms, cluster 3 was characterized by metabolic terms that were quite similar to those for cluster 2, a downregulated cluster. Thus, the three clusters may be concerned with downregulation in hepatocellular carcinogenesis.

#### 3.3.2. Interpretations of the inferred network in terms of pathogenesis

The coarse associations between the clusters in the preceding section can be interpreted on the macroscopic level, such as the pathological level. The interpretation of the network inferred based on the information at the molecular level will be useful to bridge the gap between the information about the disease mechanisms at the molecular and more macroscopic levels.

One of the most remarkable associations is found in group I. Cluster 12, with upregulation, was associated at a 1% significance level with cluster 2, with downregulation. The former cluster is characterized by the GO terms related to the immune response, and the latter is characterized by those involved with metabolism. In general, CHC and HCC result in serious damage to hepatocytes, which are important cells for nutrient metabolism, and the damage induces different responses. Indeed, HCC is a suitable target for testing active immunotherapy [29]. Furthermore, cluster 2 was also associated at a 1% significance level with cluster 14, characterized by prostaglandin-related terms. This may reflect the fact that one mediator of inflammation, prostaglandin, shows elevated expression in human and animal HCCs [30]. Thus, the associations in group I are involved in the molecular pathogenesis of the CHC and HCC stages.

TABLE 1: Cluster characterization by GO terms[#].

| Cluster no. | GO no. | Category | $P$-value | Fraction |
|---|---|---|---|---|
| 1 | GO:0030236 | Anti-inflammatory response | 0.18% | 2 of 22/6 of 26081 |
| 2 | GO:0006094 | Gluconeogenesis | 0.06% | 3 of 37/19 of 26081 |
| 2 | GO:0006066 | Alcohol metabolism | 0.12% | 6 of 37/312 of 26081 |
| 2 | GO:0006091 | Generation of precursor metabolites and energy | 0.14% | 9 of 37/961 of 26081 |
| 2 | GO:0019319 | Hexose biosynthesis | 0.34% | 3 of 37/33 of 26081 |
| 2 | GO:0046165 | Alcohol biosynthesis | 0.34% | 3 of 37/33 of 26081 |
| 2 | GO:0046364 | Monosaccharide biosynthesis | 0.34% | 3 of 37/33 of 26081 |
| 2 | GO:0006067 | Ethanol metabolism | 0.48% | 2 of 37/5 of 26081 |
| 2 | GO:0006069 | Ethanol oxidation | 0.48% | 2 of 37/5 of 26081 |
| 2 | GO:0006629 | Lipid metabolism | 1.47% | 7 of 37/722 of 26081 |
| 2 | GO:0009618 | Response to pathogenic bacteria | 4.96% | 2 of 37/15 of 26081 |
| 3 | GO:0006094 | Gluconeogenesis | 0.61% | 2 of 15/19 of 26081 |
| 3 | GO:0019319 | Hexose biosynthesis | 1.87% | 2 of 15/33 of 26081 |
| 3 | GO:0046165 | Alcohol biosynthesis | 1.87% | 2 of 15/33 of 26081 |
| 3 | GO:0046364 | Monosaccharide biosynthesis | 1.87% | 2 of 15/33 of 26081 |
| 3 | GO:0009069 | Serine family amino acid metabolism | 4.49% | 2 of 15/51 of 26081 |
| 4 | GO:0006725 | Aromatic compound metabolism | 0.07% | 4 of 20/140 of 26081 |
| 4 | GO:0009308 | Amine metabolism | 0.38% | 5 of 20/454 of 26081 |
| 4 | GO:0006570 | Tyrosine metabolism | 0.59% | 2 of 20/11 of 26081 |
| 4 | GO:0050878 | Regulation of body fluids | 1.65% | 3 of 20/113 of 26081 |
| 4 | GO:0006950 | Response to stress | 2.70% | 6 of 20/1116 of 26081 |
| 4 | GO:0006519 | Amino acid and derivative metabolism | 4.12% | 4 of 20/398 of 26081 |
| 4 | GO:0007582 | Physiological process | 4.63% | 20 of 20/17195 of 26081 |
| 5 | GO:0006917 | Induction of apoptosis* | 16.06% | 2 of 13/132 of 26081 |
| 5 | GO:0012502 | Induction of programmed cell death* | 16.06% | 2 of 13/132 of 26081 |
| 6 | GO:0009613 | Response to pest, pathogen, or parasite | 0.00% | 8 of 29/522 of 26081 |
| 6 | GO:0043207 | Response to external biotic stimulus | 0.00% | 8 of 29/557 of 26081 |
| 6 | GO:0006950 | Response to stress | 0.00% | 10 of 29/1116 of 26081 |
| 6 | GO:0009605 | Response to external stimulus | 0.05% | 10 of 29/1488 of 26081 |
| 6 | GO:0006953 | Acute-phase response | 0.05% | 3 of 29/25 of 26081 |
| 6 | GO:0006955 | Immune response | 0.34% | 8 of 29/1098 of 26081 |
| 6 | GO:0006956 | Complement activation | 0.48% | 3 of 29/52 of 26081 |
| 6 | GO:0006952 | Defense response | 0.68% | 8 of 29/1209 of 26081 |
| 6 | GO:0050896 | Response to stimulus | 1.15% | 11 of 29/2619 of 26081 |
| 6 | GO:0009607 | Response to biotic stimulus | 1.65% | 8 of 29/1372 of 26081 |
| 6 | GO:0006629 | Lipid metabolism | 2.20% | 6 of 29/722 of 26081 |
| 7 | GO:0006559 | L-phenylalanine catabolism | 0.83% | 2 of 31/9 of 26081 |
| 7 | GO:0019752 | Carboxylic acid metabolism | 1.00% | 6 of 31/590 of 26081 |
| 7 | GO:0006082 | Organic acid metabolism | 1.02% | 6 of 31/592 of 26081 |
| 7 | GO:0006558 | L-phenylalanine metabolism | 1.26% | 2 of 31/11 of 26081 |
| 7 | GO:0009074 | Aromatic amino acid family catabolism | 1.26% | 2 of 31/11 of 26081 |
| 7 | GO:0006519 | Amino acid and derivative metabolism | 1.67% | 5 of 31/398 of 26081 |
| 7 | GO:0019439 | Aromatic compound catabolism | 1.79% | 2 of 31/13 of 26081 |
| 7 | GO:0006629 | Lipid metabolism | 3.04% | 6 of 31/722 of 26081 |
| 7 | GO:0009308 | Amine metabolism | 3.09% | 5 of 31/454 of 26081 |
| 8 | GO:0001570 | Vasculogenesis | 0.09% | 2 of 21/4 of 26081 |
| 8 | GO:0006950 | Response to stress | 0.42% | 7 of 21/1116 of 26081 |
| 8 | GO:0050896 | Response to stimulus | 2.33% | 9 of 21/2619 of 26081 |

Table 1: Continued.

| | | | | |
|---|---|---|---|---|
| 9 | GO:0009611 | Response to wounding* | 11.19% | 3 of 13/394 of 26081 |
| 10 | GO:0009607 | Response to biotic stimulus* | 6.66% | 6 of 19/1372 of 26081 |
| 11 | GO:0050896 | Response to stimulus* | 72.68% | 6 of 17/2619 of 26081 |
| 12 | GO:0006955 | Immune response | 0.01% | 8 of 18/1098 of 26081 |
| 12 | GO:0006952 | Defense response | 0.01% | 8 of 18/1209 of 26081 |
| 12 | GO:0050874 | Organismal physiological process | 0.02% | 10 of 18/2432 of 26081 |
| 12 | GO:0009607 | Response to biotic stimulus | 0.03% | 8 of 18/1372 of 26081 |
| 12 | GO:0050896 | Response to stimulus | 0.39% | 9 of 18/2619 of 26081 |
| 12 | GO:0030333 | Antigen processing | 0.97% | 3 of 18/108 of 26081 |
| 12 | GO:0019882 | Antigen presentation | 2.62% | 3 of 18/151 of 26081 |
| 12 | GO:0019884 | Antigen presentation, exogenous antigen | 3.97% | 2 of 18/32 of 26081 |
| 12 | GO:0019886 | Antigen processing, exogenous antigen via MHC class II | 4.22% | 2 of 18/33 of 26081 |
| 13 | GO:0009611 | Response to wounding | 0.08% | 6 of 30/394 of 26081 |
| 13 | GO:0009613 | Response to pest, pathogen, or parasite | 0.38% | 6 of 30/522 of 26081 |
| 13 | GO:0043207 | Response to external biotic stimulus | 0.55% | 6 of 30/557 of 26081 |
| 13 | GO:0006955 | Immune response | 3.12% | 7 of 30/1098 of 26081 |
| 13 | GO:0006950 | Response to stress | 3.44% | 7 of 30/1116 of 26081 |
| 13 | GO:0050874 | Organismal physiological process | 3.98% | 10 of 30/2432 of 26081 |
| 14 | GO:0051244 | Regulation of cellular physiological process | 0.51% | 8 of 45/665 of 26081 |
| 14 | GO:0007275 | Development | 0.94% | 13 of 45/2060 of 26081 |
| 14 | GO:0001516 | Prostaglandin biosynthesis | 3.30% | 2 of 45/9 of 26081 |
| 14 | GO:0046457 | Prostanoid biosynthesis | 3.30% | 2 of 45/9 of 26081 |
| 14 | GO:0051242 | Positive regulation of cellular physiological process | 4.35% | 5 of 45/289 of 26081 |
| 15 | GO:0008283 | Cell proliferation* | 29.37% | 4 of 26/488 of 26081 |
| 16 | GO:0042221 | Response to chemical substance | 0.16% | 5 of 31/237 of 26081 |
| 16 | GO:0008152 | Metabolism | 1.29% | 25 of 31/11891 of 26081 |
| 16 | GO:0009628 | Response to abiotic stimulus | 1.89% | 5 of 31/400 of 26081 |
| 16 | GO:0006445 | Regulation of translation | 2.82% | 3 of 31/87 of 26081 |
| 17 | GO:0050817 | Coagulation* | 13.92% | 2 of 12/118 of 26081 |
| 18 | GO:0007275 | Development* | 11.67% | 6 of 16/2060 of 26081 |

* The gene ontology terms in each cluster, detected with 5% significance probability by using GO::TermFinder [18], are listed. When the terms with that significance probability were not found in the cluster, the terms with the smallest probability were listed as indicated by an asterisk. In the last column, "Fraction," the numbers of genes belonging to the corresponding category in the cluster, of genes belonging to the cluster, of genes belonging to the corresponding category in all genes of the GO term data set, and of all genes are listed.

The associated clusters 4 and 7 in group III, which were characterized by GO terms related to amino acid and lipid metabolism, also show downregulation. Indeed, the products of dysregulated (aberrant regulation) metabolism are widely used to examine liver function in common clinical tests [8]. In addition, the connection between the clusters in groups III and I implies that the downregulation of the clusters in group III may be related to abnormal hepatocyte function.

In addition, cluster 15 in group I, which is characterized by the GO term "proliferation," was associated with different clusters in groups I, II, and IV. It is known that abnormal proliferation is one of the obvious features of cancer [31]. This broad association may be responsible for the cellular level events in hepatocellular carcinogenesis.

In summary, the inferred network reveals a coarse snapshot of the gene systems related to the molecular pathogenesis and clinical characteristics of hepatocellular carcinogenesis. Although the resolution of the network is still low, due to the cluster network, the present network may provide some clues for further investigations of the pathogenic relationships involved in hepatocellular carcinoma.

#### 3.3.3. *Interpretations of the inferred network in terms of gene-gene interactions*

In addition to the macroscopic interpretations above, the gene functionality from the gene-gene interactions listed in Figure 2 is also discussed in the context of hepatocellular carcinoma. Although the consideration of gene-gene interactions is beyond the aim of the present study,

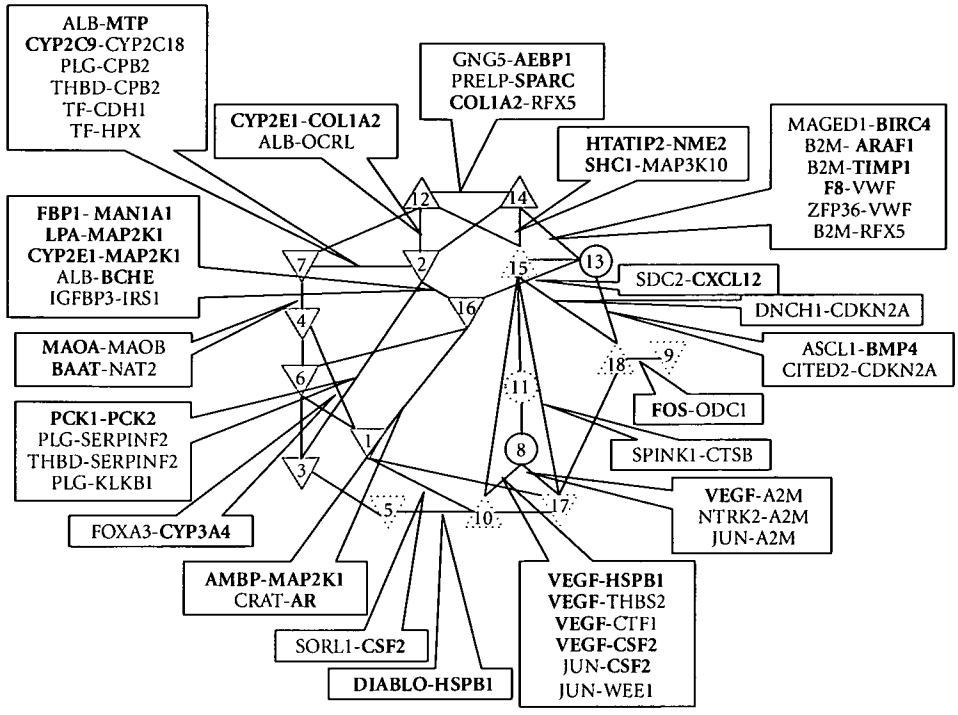


FIGURE 2: *Network between clusters, together with a projection of biological knowledge about the gene interactions.* The clusters are indicated by triangles and circles, in which the cluster numbers correspond to those in Figure 1, and the edges between the clusters are associations with 5% significance probability. The red triangles, the green upside-down triangles, and the circles indicate the clusters of up- and downregulated genes, and the mixture of them, respectively, and the dotted triangles indicate the clusters that were not characterized by GO terms with less than 5% significance probability. The known gene interactions in Pathway Assist are indicated between the clusters, in which the genes highlighted by bold letters are characterized by the GO terms in Table 1.

some examples may provide possible clues about the disease mechanisms.

First, we surveyed the frequencies of GO terms (geneGOB listed in the supplemental data at http://www.cbrc.jp/~horimoto/suppl/HCGO.pdf) in the selected genes in the present analysis, to investigate the features of gene-gene interactions in the inferred network. A few general terms appeared frequently, such as "response" (122 times in the geneGOB column of the supplemental data at http://www.cbrc.jp/~horimoto/suppl/HCGO.pdf) and "metabolism" (183), as expected from the coarse associations between the clusters in the preceding section. As for more specific terms about the gene function, "lipid" (46), "apoptosis" (31), and "cell growth" (27) are remarkably found in the list. The "lipid" is expected from the relationship between groups I and III, and the "apoptosis" and the "cell growth" are also expected from the frequent appearance of GO terms (clusterGOB listed in Table 1) related to the morphological events. Since the frequent appearance of "lipid" may be a sensitive reflection of the protein-protein interactions in lipid metabolic pathways to the expression profiles, here, we focus on the gene-gene interactions characterized by the "apoptosis" and the "cell growth."

Among the gene-gene interactions listed in Figure 2, the gene-gene interactions characterized by the cell growth or death are found in the coarse associations between the clusters. Group I contains the gene-gene interactions related to apoptosis. The expression of HTAIP2 (HIV-1 Tat interactive protein 2, 30 kd) in cluster 14 induces the expression of a number of genes, including NME2 (nonmetastatic cells 2, protein) in cluster 15 as well as the apoptosis-related genes Bad and Siva [32]. MAGED1 (melanoma antigen, family D, 1) in cluster 13, and its binding partner BIRC4 (baculoviral IAP repeat-containing 4) in cluster 14 are known to play some roles in apoptosis [33]. In addition, the expression of COL1A2 (collagen, type I, alpha 2) in cluster 12, which is related to cell adhesion and skeletal development, is regulated by RFX5 (regulatory factor X, 5) in cluster 14 [29, 34]. In group IV, the expression of CSF2 (colony-stimulating factor 2) in cluster 8 is dependent on the cooperation between NFAT (nuclear factor of activated T cells) and JUN (Jun oncogene) in cluster 10 [35]. Between groups I and II, ASCL1 (achaete-scute complex-like 1) in cluster 13 and BMP4 (bone morphogenetic protein 4) in cluster 18 share the function of cell differentiation [36].

As a result, the gene-gene interactions listed above are related to the mechanisms of cell growth or death at the molecular level. On the other hand, the cluster associations reveal the relationship between the cancer-induced events and various aspects of metabolisms at the pathogenesis and clinical characteristics. Thus, the metabolic pathways might directly

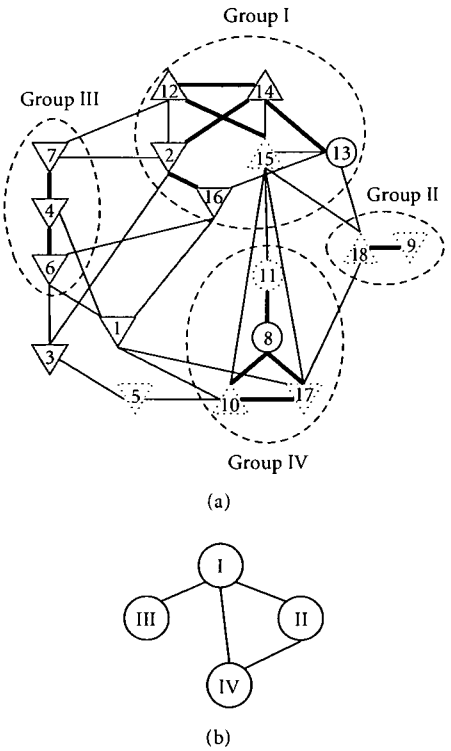


FIGURE 3: *Orchestration of gene systems.* (a) The association with 1% significance probability is indicated by a bold line, and the clusters with 1% significance association are naturally divided into four groups, which are enclosed by broken lines. (b) The connections between the groups are drawn schematically, as a coarse grinning of the cluster association.

influence the mechanisms of cancer-induced cell growth or death at the molecular level in unknown ways.

### 3.4. Merits and pitfalls of the present approach

The present analysis reveals a framework of gene system associations in hepatocellular carcinogenesis. The inferred network provides a bridge between the events at the molecular level and those at macroscopic levels: the associations between clusters characterized by cancer-related responses and those characterized by metabolic and morphological events can be interpreted from pathological and clinical views. In addition, the viewpoint of the gene-gene interactions in the inferred network indicates the relationship between cancer and cell growth/death. Thus, the gene systems network may also be useful as a bridge between the gene-gene interactions and the observations at macroscopic levels, such as clinical tests.

The present method assumes linearity in the cluster associations by using a partial correlation coefficient to identify the independence between clusters. It is well known that the interactions among genes and other molecular components are often nonlinear, and the assumption of linearity misses many important relationships among genes. In the present study, our aim was not the inference of detailed gene-gene interactions, but of coarse gene system interactions. Indeed, the use of a partial correlation coefficient is employed as a feasible approach for gene association inference as a first approximation in some studies [37, 38]. Thus, the assumption of the linearity is not suitable for a fine analysis of dynamic gene behaviors, but may be useful for the approximate analysis of static gene associations.

## ACKNOWLEDGMENTS

S. Aburatani was supported by a Grant-in-Aid for Scientific Research (Grant 18681031) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, and K. Horimoto was partly supported by a Grant-in-Aid for Scientific Research on Priority Areas "Systems Genomics" (Grant 18016008) and by a Grant-in-Aid for Scientific Research (Grant 19201039) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan. This study was supported in part by the New Energy and Industrial Technology Development Organization (NEDO) of Japan and by the Ministry of Health, Labour, and Welfare of Japan.

## REFERENCES

[1] M. J. Alter, H. S. Margolis, K. Krawczynski, et al., "The natural history of community-acquired hepatitis C in the United States. The sentinel counties chronic non-A, non-B hepatitis study team," *The New England Journal of Medicine*, vol. 327, no. 27, pp. 1899–1905, 1992.

[2] A. M. Di Bisceglie, "Hepatitis C," *The Lancet*, vol. 351, no. 9099, pp. 351–355, 1998.

[3] S. Zeuzem, S. V. Feinman, J. Rasenack, et al., "Peginterferon alfa-2a in patients with chronic hepatitis C," *The New England Journal of Medicine*, vol. 343, no. 23, pp. 1666–1672, 2000.

[4] S. S. Thorgeirsson, J.-S. Lee, and J. W. Grisham, "Molecular prognostication of liver cancer: end of the beginning," *Journal of Hepatology*, vol. 44, no. 4, pp. 798–805, 2006.

[5] N. Iizuka, M. Oka, H. Yamada-Okabe, et al., "Oligonucleotide microarray for prediction of early intrahepatic recurrence of hepatocellular carcinoma after curative resection," *The Lancet*, vol. 361, no. 9361, pp. 923–929, 2003.

[6] H. Okabe, S. Satoh, T. Kato, et al., "Genome-wide analysis of gene expression in human hepatocellular carcinomas using cDNA microarray: identification of genes involved in viral carcinogenesis and tumor progression," *Cancer Research*, vol. 61, no. 5, pp. 2129–2137, 2001.

[7] L.-H. Zhang and J.-F. Ji, "Molecular profiling of hepatocellular carcinomas by cDNA microarray," *World Journal of Gastroenterology*, vol. 11, no. 4, pp. 463–468, 2005.

[8] J. Jiang, P. Nilsson-Ehle, and N. Xu, "Influence of liver cancer on lipid and lipoprotein metabolism," *Lipids in Health and Disease*, vol. 5, p. 4, 2006.

[9] A. Zerbini, M. Pilli, C. Ferrari, and G. Missale, "Is there a role for immunotherapy in hepatocellular carcinoma?" *Digestive and Liver Disease*, vol. 38, no. 4, pp. 221–225, 2006.

[10] K. Horimoto and H. Toh, "Statistical estimation of cluster boundaries in gene expression profile data," *Bioinformatics*, vol. 17, no. 12, pp. 1143–1151, 2001.

[11] H. Toh and K. Horimoto, "Inference of a genetic network by a combined approach of cluster analysis and graphical Gaussian modeling," *Bioinformatics*, vol. 18, no. 2, pp. 287–297, 2002.

[12] S. Lauritzen, *Graphical Models*, Oxford University Press, Oxford, UK, 1996.
[13] J. Whittaker, *Graphical Models in Applied Multivariate Statistics*, John Wiley & Sons, New York, NY, USA, 1990.
[14] H. Toh and K. Horimoto, "System for automatically inferring a genetic network from expression profiles," *Journal of Biological Physics*, vol. 28, no. 3, pp. 449–464, 2002.
[15] D. K. Slonim, "From patterns to pathways: gene expression data analysis comes of age," *Nature Genetics*, vol. 32, no. 5, pp. 502–508, 2002.
[16] S. Aburatani, S. Kuhara, H. Toh, and K. Horimoto, "Deduction of a gene regulatory relationship framework from gene expression data by the application of graphical Gaussian modeling," *Signal Processing*, vol. 83, no. 4, pp. 777–788, 2003.
[17] M. Ashburner, C. A. Ball, J. A. Blake, et al., "Gene ontology: tool for the unification of biology," *Nature Genetics*, vol. 25, no. 1, pp. 25–29, 2000.
[18] E. I. Boyle, S. Weng, J. Gollub, et al., "GO::TermFinder—open source software for accessing gene ontology information and finding significantly enriched gene ontology terms associated with a list of genes," *Bioinformatics*, vol. 20, no. 18, pp. 3710–3715, 2004.
[19] A. Nikitin, S. Egorov, N. Daraselia, and I. Mazo, "Pathway studio—the analysis and navigation of molecular networks," *Bioinformatics*, vol. 19, no. 16, pp. 2155–2157, 2003.
[20] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley & Sons, New York, NY, USA, 1990.
[21] R. J. Freund and W. J. Wilson, *Regression Analysis: Statistical Modeling of a Response Variable*, Academic Press, San Diego, Calif, USA, 1998.
[22] A. P. Dempster, "Covariance selection," *Biometrics*, vol. 28, no. 1, pp. 157–175, 1972.
[23] N. Wermuth and E. Scheidt, "Algorithm AS 105: fitting a covariance selection model to a matrix," *Applied Statistics*, vol. 26, no. 1, pp. 88–92, 1977.
[24] L. F. Wu, T. R. Hughes, A. P. Davierwala, M. D. Robinson, R. Stoughton, and S. J. Altschuler, "Large-scale prediction of *Saccharomyces cerevisiae* gene function using overlapping transcriptional clusters," *Nature Genetics*, vol. 31, no. 3, pp. 255–265, 2002.
[25] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New York, NY, USA, 2nd edition, 1984.
[26] S. Aburatani, K. Goto, S. Saito, et al., "ASIAN: a website for network inference," *Bioinformatics*, vol. 20, no. 16, pp. 2853–2856, 2004.
[27] 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 Research*, vol. 33, pp. W659–W664, 2005.
[28] M. Honda, S. Kaneko, H. Kawai, Y. Shirota, and K. Kobayashi, "Differential gene expression between chronic hepatitis B and C hepatic lesion," *Gastroenterology*, vol. 120, no. 4, pp. 955–966, 2001.
[29] T. Wu, "Cyclooxygenase-2 in hepatocellular carcinoma," *Cancer Treatment Reviews*, vol. 32, no. 1, pp. 28–44, 2006.
[30] H. Xiao, V. Palhan, Y. Yang, and R. G. Roeder, "TIP30 has an intrinsic kinase activity required for up-regulation of a subset of apoptotic genes," *The EMBO Journal*, vol. 19, no. 5, pp. 956–963, 2000.
[31] W. B. Coleman, "Mechanisms of human hepatocarcinogenesis," *Current Molecular Medicine*, vol. 3, no. 6, pp. 573–588, 2003.
[32] Y. Xu, P. K. Sengupta, E. Seto, and B. D. Smith, "Regulatory factor for X-box family proteins differentially interact with histone deacetylases to repress collagen $\alpha$2(I) gene *(COL1A2)* expression," *Journal of Biological Chemistry*, vol. 281, no. 14, pp. 9260–9270, 2006.
[33] P. A. Barker and A. Salehi, "The MAGE proteins: emerging roles in cell cycle progression, apoptosis, and neurogenetic disease," *Journal of Neuroscience Research*, vol. 67, no. 6, pp. 705–712, 2002.
[34] Y. Xu, L. Wang, G. Buttice, P. K. Sengupta, and B. D. Smith, "Interferon $\gamma$ repression of collagen *(COL1A2)* transcription is mediated by the RFX5 complex," *The Journal of Biological Chemistry*, vol. 278, no. 49, pp. 49134–49144, 2003.
[35] F. Macian, C. Garcia-Rodriguez, and A. Rao, "Gene expression elicited by NFAT in the presence or absence of cooperative recruitment of Fos and Jun," *The EMBO Journal*, vol. 19, no. 17, pp. 4783–4795, 2000.
[36] J. Fu, S. S. W. Tay, E. A. Ling, and S. T. Dheen, "High glucose alters the expression of genes involved in proliferation and cell-fate specification of embryonic neural stem cells," *Diabetologia*, vol. 49, no. 5, pp. 1027–1038, 2006.
[37] J. Schäfer and K. Strimmer, "An empirical Bayes approach to inferring large-scale gene association networks," *Bioinformatics*, vol. 21, no. 6, pp. 754–764, 2005.
[38] A. de la Fuente, N. Bing, I. Hoeschele, and P. Mendes, "Discovery of meaningful associations in genomic data using partial correlation coefficients," *Bioinformatics*, vol. 20, no. 18, pp. 3565–3574, 2004.

# Lipid-Induced Oxidative Stress Causes Steatohepatitis in Mice Fed an Atherogenic Diet

Naoto Matsuzawa,[1,2] Toshinari Takamura,[1] Seiichiro Kurita,[1] Hirofumi Misu,[1] Tsuguhito Ota,[1] Hitoshi Ando,[1] Masayoshi Yokoyama,[1] Masao Honda,[1] Yoh Zen,[3] Yasuni Nakanuma,[3] Ken-ichi Miyamoto,[2] and Shuichi Kaneko[1]

**Recently, nonalcoholic steatohepatitis (NASH) was found to be correlated with cardiovascular disease events independently of the metabolic syndrome. The aim of this study was to investigate whether an atherogenic (Ath) diet induces the pathology of steatohepatitis necessary for the diagnosis of human NASH and how cholesterol and triglyceride alter the hepatic gene expression profiles responsible for oxidative stress. We investigated the liver pathology and plasma and hepatic lipids of mice fed the Ath diet. The hepatic gene expression profile was examined with microarrays and real-time polymerase chain reactions. The Ath diet induced dyslipidemia, lipid peroxidation, and stellate cell activation in the liver and finally caused precirrhotic steatohepatitis after 24 weeks. Cellular ballooning, a necessary histological feature defining human NASH, was observed in contrast to existing animal models. The addition of a high-fat component to the Ath diet caused hepatic insulin resistance and further accelerated the pathology of steatohepatitis. A global gene expression analysis revealed that the Ath diet up-regulated the hepatic expression levels of genes for fatty acid synthesis, oxidative stress, inflammation, and fibrogenesis, which were further accelerated by the addition of a high-fat component. Conversely, the high-fat component down-regulated the hepatic gene expression of antioxidant enzymes and might have increased oxidative stress. *Conclusion:* The Ath diet induces oxidative stress and steatohepatitis with cellular ballooning. The high-fat component induces insulin resistance, down-regulates genes for antioxidant enzymes, and further aggravates the steatohepatitis. This model suggests the critical role of lipids in causing oxidative stress and insulin resistance leading to steatohepatitis.** (HEPATOLOGY 2007;46:1392-1403.)

*Abbreviations: 4-HNE, 4-hydroxy-2-nonenal; α-SMA, α-smooth muscle actin; ALT, alanine aminotransferase; Ath, atherogenic; Ath+HF, atherogenic and high-fat; AUC, area under the curve; BW, body weight; Col1a1, procollagen type I alpha 1; Col1a2, procollagen type I alpha 2; Col4a1, procollagen type IV alpha 1; CPT-1a, carnitine palmitoyltransferase 1a; FAS, fatty acid synthase; FFA, free fatty acid; GPCR, G protein-coupled receptor; GPCRDB, G Protein-Coupled Receptor Database; GTT, glucose tolerance test; H&E, hematoxylin-eosin; HDL-C, high-density lipoprotein-cholesterol; HOMA-IR, homeostasis model assessment of insulin resistance; HPLC, high-performance liquid chromatography; HSC, hepatic stellate cell; IRS, insulin receptor substrate; ITT, insulin tolerance test; LDL, low-density lipoprotein; LDL-C; low-density lipoprotein-cholesterol; MAPK, mitogen-activated protein kinase; MCD, methionine- and choline-deficient; mRNA, messenger RNA; NADPH, reduced-form nicotinamide adenine dinucleotide phosphate; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; ND, not determined; PAI-1, plasminogen activator inhibitor 1; PCR, polymerase chain reaction; PPARα, peroxisome proliferator-activated receptor α; ROS, reactive oxygen species; SEM, standard error of the mean; SREBP-1c, sterol regulatory element binding protein 1c; TBS-T, trishydroxymethylaminomethane-buffered saline Tween 20; TCA, tricarboxylic acid cycle; TG, triglyceride; TGF-β, transforming growth factor β; TNF-α, tumor necrosis factor α; VLDL-C, very low density lipoprotein-cholesterol.*

*From the [1]Department of Disease Control and Homeostasis, Graduate School of Medical Science, [2]Department of Pharmacy and Health Science, Graduate School of Natural Science and Technology, and [3]Department of Human Pathology, Graduate School of Medical Science, Kanazawa University, Kanazawa, Japan.*

*Received January 9, 2007; accepted June 12, 2007.*

*Supported by grants-in-aid from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.*

*Address reprint requests to: Toshinari Takamura, M.D., Ph.D., Department of Disease Control and Homeostasis, Graduate School of Medical Science, Kanazawa University, 13-1 Takara-machi, Kanazawa, Ishikawa 920-8641, Japan. E-mail: ttakamura@m-kanazawa.jp; fax: (81) 76-234-4250.*



*Published online in Wiley InterScience (www.interscience.wiley.com).*

*DOI 10.1002/hep.21874*

*Potential conflict of interest: Nothing to report.*

*Supplementary material for this article can be found on the HEPATOLOGY Web site (http://interscience.wiley.com/jpages/0270-9139/suppmat/index.html).*

Nonalcoholic fatty liver disease (NAFLD) is currently the most common chronic liver condition in the Western world. Clinical, epidemiological, and biochemical data strongly support the concept that NAFLD is the hepatic manifestation of the metabolic syndrome, the constellation of metabolic abnormalities including obesity, dyslipidemia, and insulin resistance.[1] NAFLD includes not only steatosis (without other injury) but also various degrees of inflammation and fibrosis.[2] Simple steatosis is usually considered benign, but the development of inflammatory changes in the liver [called nonalcoholic steatohepatitis (NASH)] is recognized as a precursor to more severe liver disease and sometimes evolves into cryptogenic cirrhosis.[3] It has been recently proposed that NASH is strongly correlated with cardiovascular disease events independently of the metabolic syndrome.[4] Therefore, further investigations of NASH are required to elucidate the pathogenesis of this process and to develop treatments.

To date, however, studies of NASH have been hampered by the absence of a suitable experimental model. The use of genetic defects or targeted overexpression to produce obesity[5] or impaired hepatic lipid metabolism[6] in rodents has been used as an NAFLD model. Although these genetic manipulations can assess the biological importance of each gene *in vivo*, they might not reflect the natural etiology of NAFLD in patients and rarely lead to the pathology of NASH. The other models frequently used are based on nutritional manipulations. Natural nutritional models have been described, including the use of a sucrose-rich and fat-rich diet.[7] However, in these models, rodents accumulate minimal fat and develop subtle inflammation of the liver. The methionine- and choline-deficient (MCD) model, which is frequently used to produce more progressive liver pathology, leads to the development of steatosis with lobular inflammation and with perisinusoidal and pericentral fibrosis.[8,9] However, this model lacks lipotrophic factors, insulin resistance,[10] or the cellular ballooning that is observed only with the addition of a high-fat component to the MCD diet.[11]

In this study, we focused on an atherogenic (Ath) diet, which contains cholesterol and cholic acid. Because the diet produces not only an Ath lipoprotein profile but also vascular fatty streak lesions, it has been widely used to study atherosclerosis in animals, including mice.[12] Although the Ath diet has recently been reported to induce liver steatosis, inflammation, and fibrosis,[13] lipid metabolism, insulin resistance, and hepatic gene expression profiles responsible for liver pathology remain to be determined in this model. To address this issue, we investigated the time course of the pathological changes and gene expression profiles of the liver in mice fed the Ath diet. In addition, by adding a high-fat component to the Ath diet, we elucidated the impact of insulin resistance, which is commonly observed in NASH patients, on the development of oxidative stress in the liver and pathology of steatohepatitis.

**Table 1. The Composition of the 3 Diets**

| Composition | Control | Ath | Ath+HF |
|---|---|---|---|
| CRF-1 (%) | 100 | 90.75 | 38.25 |
| Cocoa butter (%) | – | 7.50 | 60.0 |
| Cholesterol (%) | – | 1.25 | 1.25 |
| Cholate (%) | – | 0.50 | 0.50 |
| | | | (wt/wt %) |
| **Energy composition** | **Control** | **Ath** | **Ath+HF** |
| Carbohydrate (g) | 60.9 | 55.2 | 23.3 |
| Protein (g) | 22.4 | 20.3 | 8.6 |
| Fat (g) | 6.0 | 14.0 | 60.0 |
| Total calorie (kcal) | 363 | 411 | 669 |
| | | | (/100 g) |

The contents of vitamins and minerals in each diet are presented in Supplementary Table 1.

## Materials and Methods

***Animals and Experimental Design.*** Male C57Bl/6J mice were purchased from Charles River Laboratories Japan (Yokohama, Japan) at 6 weeks of age. After 2 weeks of acclimation, the mice were divided into the following 3 groups: (1) control mice given a standard chow (CRF-1, Charles River Laboratories Japan), (2) mice given an Ath diet, and (3) mice fed an atherogenic and high-fat (Ath+HF) diet. The Ath and Ath+HF diets were prepared by the addition of cocoa butter, cholesterol, and cholate to CRF-1. These diets were prepared by Oriental Yeast (Tokyo, Japan). The compositions of each diet are shown in Table 1 and Supplementary Table 1. At 6 weeks of age, the mice were housed in colony cages with a 12-hour light/12-hour dark cycle, and they were given food and water *ad libitum*. All animal procedures were in accordance with the standards set forth in the Guidelines for the Care and Use of Laboratory Animals at the Takara-Machi campus of Kanazawa University (Japan).

***Blood Sampling and Analysis.*** At 6, 12, or 24 weeks, blood samples were obtained from the tail vein following a 12-hour fast. Enzymatic assays for the total cholesterol, free cholesterol, free fatty acids (FFAs), triglyceride (TG), and alanine aminotransferase (ALT) were performed with kits purchased from Wako Pure Chemical Industries (Osaka, Japan). The cholesterol and TG profiles in plasma lipoproteins were analyzed with a dual-detection high-performance liquid chromatography (HPLC) system with 2 tandem connected TSKgel LipopropakXL

columns (300 × 7.8 mm; Tosoh, Japan) by Skylight Biotech (Akita, Japan).[14]

***Glucose Tolerance Tests (GTTs) and Insulin Tolerance Tests (ITTs).*** At 12 weeks, GTTs and ITTs were conducted. For GTTs, glucose was administered (1.5 g/kg body weight) following a 12-hour fast. For ITTs, mice were injected intraperitoneally with insulin (0.5 U/kg of body weight; Humulin R, Eli Lilly, Indianapolis, IN) following a 4-hour fast. The glucose values were measured from whole venous blood with a blood glucose monitoring system (FreeStyle, Kissei, Matsumoto, Japan) 0, 15, 30, 60, and 120 minutes after the administration of glucose or insulin.

***Pyruvate Challenge Test.*** At 6 weeks, we conducted the pyruvate challenge test.[15,16] The mice, deprived of food for 16 hours, were injected intraperitoneally with pyruvate dissolved in saline (2 g/kg). The blood glucose values were measured 0, 15, 30, 60, and 90 minutes after the injection of pyruvate.

***Tissue Preparation and Histological Examination.*** At 6, 12, or 24 weeks, the mice were killed by cervical dislocation under diethyl ether anesthesia following a 12-hour fast. The livers were immediately removed and weighed. A large portion of each liver was snap-frozen in liquid nitrogen for later RNA studies. The remaining tissue was fixed in 10% buffered formalin, processed, and embedded in paraffin for hematoxylin-eosin (H&E), Azan, and Sirius red staining and was blindly scored by a single pathologist. Steatosis, fibrosis, and acinar inflammation were semiquantitatively evaluated according to the standard criteria of grading and staging for NASH, with minor modifications.[17] To evaluate steatosis, we used the absolute percentage of the macrovesicular fat droplet area in the section area (that is, $8 \times 10^5$ hepatocytes in 4 $mm^2$). For inflammation, 0 was defined as no hepatocyte injury or inflammation, 1 was defined as mild focal injury, 2 was defined as noticeable injury, and 3 was defined as severe zone 3 hepatocyte injury or inflammation. For fibrosis, 0 was defined as no fibrosis, 1 was defined as pericellular and perivenular fibrosis, 2 was defined as focal bridging fibrosis, 3 was defined as much bridging fibrosis with lobular distortion, and 4 was defined as cirrhosis.

Slides were immunostained with monoclonal mouse anti-human $\alpha$-smooth muscle actin ($\alpha$-SMA; Dako Japan, Kyoto, Japan). This was followed by the application of the immunoperoxidase technique with an Envision kit (Dako Japan). The peroxidase activity was identified by a reaction with 3′,3′-diaminobenzidine (Sigma, St Louis, MO). Areas staining for $\alpha$-SMA were quantified morphometrically with WinROOF version 5.7 (Mitani Shoji, Fukui, Japan) and expressed as percentages of the field area.

***Measuring the Hepatic Lipid Content.*** Hepatic lipids were extracted with chloroform/methanol (2:1) according to a published method.[18] With a kit (Wako), the extract was dissolved in water and subsequently analyzed for TG, total cholesterol, free cholesterol, and FFAs.

***Measuring the Hepatic Hydroxyproline Content.*** The hydroxyproline content in liver samples was quantified colorimetrically according to a published method.[19] Briefly, a 0.2-g liver sample was homogenized in 6 N HCl and hydrolyzed at 110°C for 16 hours. The hydrolysate was filtered, aliquots were evaporated under a vacuum, and the sediment was redissolved in 50% isopropanol. Then, the samples were incubated in a solution containing 0.84% chloramine-T, 42 mM sodium acetate, 2.6 mM citric acid, and 39.5% (vol/vol) isopropanol (pH 6.0) for 10 minutes at room temperature. Next, the samples were incubated in a solution containing 0.248 g of *p*-dimethylaminobenzaldehyde dissolved in 0.27 mL of 60% perchloric acid and 0.73 mL of isopropanol for 90 minutes at 50°C. The hydroxyproline content was quantified photometrically at 558 nm.

***Measuring Hepatic Protein Carbonyls.*** The concentration of hepatic proteins containing carbonyl groups (those that react with 2,4-dinitrophenylhydrazine to form the corresponding hydrazone) was determined spectrophotometrically according to the instructions with a protein carbonyl assay kit (Cayman Chemical, Ann Arbor, MI).

***RNA Preparation for the Microarray Analysis.*** Total RNA was isolated from the frozen liver with the TOTALLY RNA kit (Applied Biosystems, Foster City, CA). Each sample was prepared by equal amounts of total RNA being pooled from 3 mice in the same group. Three micrograms of total RNA was used to synthesize antisense RNA with the AminoAllyl MessageAmp II antisense RNA kit (Applied Biosystems) for oligo-microarrays (AceGene Mouse Oligo Chip 30K, DNA Chip Research, Yokohama, Japan). Each microarray hybridization sample and the reference amino allyl antisense RNA were labeled with Cy5 and Cy3, respectively. Hybridization and washing were performed according to the manufacturer's instructions; this was followed by scanning with a G2505B microarray scanner (Agilent Technologies, Palo Alto, CA) and then image analysis with GenePix Pro 4.1 software (Axon Instruments, Union City, CA). Microarray data were normalized (LOWESS [locally weighted polynomial regression] method) with GeneSpring version 7.2 software (Agilent Technologies). For the pathway analysis, we used the GenMAPP and MAPPFinder software package.[20,21] The GenMAPP program contains

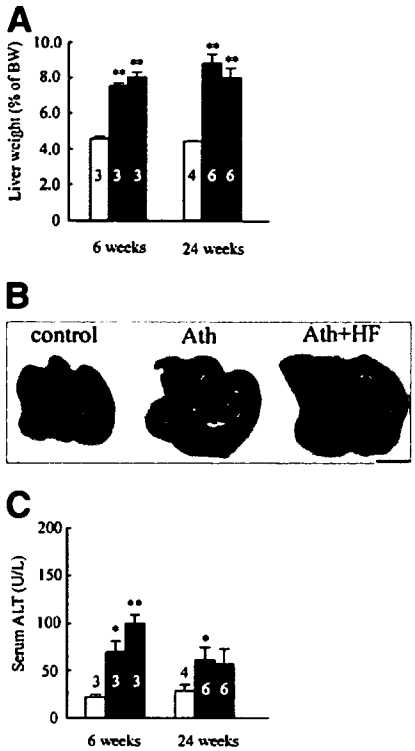


Fig. 1. Effects of 3 diets on the liver weight and morphology and serum ALT. (A) Liver weight with respect to the body weight (BW) of control mice fed standard chow (white bars), the Ath diet (gray bars), or the Ath+HF diet (black bars) after 6 or 24 weeks. (B) Photograph of livers after 12 weeks of feeding with the standard chow, Ath diet, or Ath+HF diet (scale bars: 10 mm). (C) Serum ALT levels after 6 or 24 weeks. The values represent the means ± the SEM. The number of animals per group is indicated in or just above the bars. $*P < 0.05$ and $**P < 0.01$ versus the control group.

many pathway maps that can be associated with imported microarray data. The MAPPFinder program, which links gene expression data to the pathway maps, can calculate the z score (standardized difference score) and the percentage of genes measured that meet user-defined criteria (±25% in the change fold in our analysis). With the z score and the percentage, the pathways were ranked according to the relative change in the gene expression. The microarray data sets have been submitted to the Genome Expression Omnibus Database (available at http://www.ncbi.nlm.nih.gov/geo/) under series GSE5852.

***Quantitative Real-Time Polymerase Chain Reaction (PCR).*** The reverse transcription of 100 ng of total RNA (the same sample used for the microarray analysis) was performed with Oligo(dT)$_{12\text{-}18}$ primer and SuperScript III reverse transcriptase (Invitrogen, Carlsbad, CA). PCR was performed on an ABI-Prism 7900HT (Applied Biosystems). The specific PCR primers and TaqMan probe used in this study were obtained from Applied Biosystems. The PCR conditions were 1 cycle at 50°C for 2 minutes and at 95°C for 10 minutes followed by 40 cycles at 95°C for 15 s and at 60°C for 1 minute.

***Western Blot Analysis.*** Livers were homogenized in a buffer containing 20 mM trishydroxymethylaminomethane-HCl (pH 7.5), 5 mM ethylene diamine tetraacetic acid, 1% NP-40, and a protease inhibitor cocktail (Pierce, Rockford, IL). Homogenated proteins (30 μg/lane) were separated by 4%-20% gradient sodium dodecyl sulfate–polyacrylamide gels (Daiichi Chemicals, Tokyo, Japan) and resolved with 130 V over 2 hours. Proteins were transferred to polyvinylidene difluoride membranes (Millipore, Billerica, MA) with a Transblot apparatus (Bio-Rad, Hercules, CA). The membranes were blocked in a buffer containing 5% nonfat milk, 50 mM trishydroxymethylaminomethane (pH 7.6), 150 mM NaCl, and 0.1% Tween 20 [trishydroxymethylaminomethane-buffered saline Tween 20 (TBS-T)] for 12 hours at 4°C. They were then probed with the monoclonal anti–4-hydroxy-2-nonenal (4-HNE) antibody (NOF, Tokyo, Japan) at a 1:200 dilution, with the polyclonal anti–insulin receptor substrate 2 (IRS-2) antibody (Millipore) at a 1:500 dilution, or with the polyclonal anti–glyceraldehyde 3-phosphate dehydrogenase antibody (Santa Cruz Biotechnology, Santa Cruz, CA) at a 1:3000 dilution in 5% bovine serum albumin TBS-T for 12 hours at 4°C. After the membranes had been washed in TBS-T, the blots were incubated with the horseradish peroxidase–linked secondary antibody (Cell Signaling Technology, Beverly, MA). Signals were detected with a chemiluminescence detection system (ECL Plus, GE Healthcare Bio-Sciences, Piscataway, NJ) and exposure to X-ray film. The hepatic 4-HNE contents were quantified with WinROOF version 5.7 (Mitani Shoji).

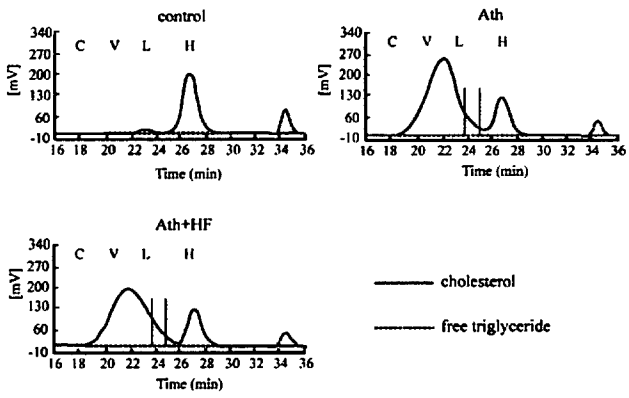


Fig. 2. HPLC analysis of plasma lipoproteins: fractionation by HPLC of cholesterol and free TG from mouse plasma after 24 weeks on the diet. The chromatograms for 1 representative sample are presented. The chylomicron, VLDL-C, LDL-C, and HDL-C fractions are labeled C, V, L, and H, respectively. The shaded fractions correspond to the level of small dense LDL-C.

**Table 2. Effects of the 3 Diets on Body Weight and Lipid Levels at 6 or 24 Weeks of Feeding**

| | 6 weeks | | | 24 weeks | | |
|---|---|---|---|---|---|---|
| Diet type | Control (n = 3) | Ath (n = 3) | Ath+HF (n =3) | Control (n = 4) | Ath (n = 6) | Ath+HF (n = 6) |
| Body weight (g) | 24.7 ± 0.5 | 24.9 ± 0.4 | 23.2 ± 0.4 | 29.0 ± 0.7 | 28.1 ± 1.8 | 26.4 ± 1.1** |
| Epididymal fat pad weight (g) | 0.14 ± 0.01 | 0.15 ± 0.01 | 0.15 ± 0.02 | 0.25 ± 0.01 | 0.09 ± 0.01** | 0.17 ± 0.01** |
| Plasma triglycerides (mg/dL) | 68.0 ± 5.2 | 54.6 ± 10.1 | 24.0 ± 3.6** | 41.5 ± 4.6 | 33.8 ± 3.5 | 20.8 ± 1.7* |
| Plasma total cholesterol (mg/dL) | 85.0 ± 8.5 | 173.6 ± 5.3** | 168.1 ± 5.3** | 84.6 ± 3.8 | 257.1 ± 10.8** | 204.4 ± 8.8** |
| Plasma free cholesterol (mg/dL) | 23.9 ± 2.2 | 45.9 ± 2.7** | 40.2 ± 0.7** | 17.3 ± 0.3 | 47.5 ± 5.6** | 31.3 ± 2.3* |
| Plasma FFA (mEq/L) | 0.58 ± 0.09 | 0.75 ± 0.10 | 0.46 ± 0.07 | 0.48 ± 0.04 | 0.48 ± 0.04 | 0.24 ± 0.06* |
| Plasma insulin (μU/mL) | N.D. | N.D. | N.D. | 6.2 ± 0.3 | 11.2 ± 1.7 | 13.8 ± 2.9 |
| Fasting blood glucose (mg/dL) | N.D. | N.D. | N.D. | 93 ± 4 | 85 ± 4 | 93 ± 8 |
| HOMA-IR | N.D. | N.D. | N.D. | 1.4 ± 0.1 | 2.3 ± 0.3 | 3.1 ± 0.4* |
| Hepatic triglycerides (μg/mg protein) | 67.2 ± 10.4 | 89.3 ± 19.7 | 150.8 ± 21.6* | 148.6 ± 20.9 | 52.8 ± 17.4* | 64.5 ± 9.2* |
| Hepatic total cholesterol (μg/mg protein) | 42.0 ± 2.8 | 206.8 ± 22.5** | 342.8 ± 40.8** | 34.3 ± 2.2 | 143.5 ± 24.1* | 192.8 ± 25.0** |
| Hepatic free cholesterol (μg/mg protein) | 22.1 ± 3.9 | 30.4 ± 4.2 | 52.6 ± 6.6* | 19.9 ± 2.3 | 33.0 ± 2.6* | 30.9 ± 5.9 |
| Hepatic FFA (μEq/mg protein) | 45.6 ± 4.0 | 52.6 ± 3.3 | 63.0 ± 2.8* | 53.1 ± 1.8 | 81.1 ± 6.3* | 83.6 ± 8.7* |

Data are means ± SEM. Significantly different from control value: $*P < 0.05$; $**P < 0.01$. Abbreviations: FFA, free fatty acid; HOMA-IR, homeostasis model assessment insulin resistance; N.D., Not determined.

***Statistical Analysis.*** The results are shown as the means ± the standard error of the mean (SEM). The data were analyzed with a 1-way analysis of variance to compare the means of all groups. The Bonferroni multiple-comparison procedure was used to determine which pairs of means were different. Differences in the histological scores between the Ath and Ath+HF groups were compared with the Mann-Whitney U test. All calculations were performed with SPSS version 12.0 software for Windows (SPSS, Chicago, IL).

## Results

***Ath Diet Causes Hepatomegaly and Liver Injury.*** Hepatomegaly was observed in the Ath and Ath+HF groups (Fig. 1A). As shown in Fig. 1B, the livers of mice fed the Ath diet were grossly enlarged and pale in color. The serum ALT level was also elevated in the Ath and Ath+HF groups (Fig. 1C). Splenomegaly, frequently associated with cirrhosis, was detected in the Ath and Ath+HF groups at 24 weeks.

***Effect of the Ath Diet on the Plasma Lipid Levels and Hepatic Lipid Content.*** As shown in Table 2, the plasma cholesterol levels were significantly elevated in the Ath diet group after both 6 and 24 weeks. An HPLC analysis revealed that the Ath and Ath+HF diets markedly increased the very low density lipoprotein-cholesterol (VLDL-C), low-density lipoprotein-cholesterol (LDL-C), and small dense LDL-C fractions, whereas they lowered high-density lipoprotein-cholesterol (HDL-C) in comparison with the controls (Fig. 2). As reported previously, we also confirmed atherosclerotic lesions in the mice fed the Ath and Ath+HF diets but not in the mice fed normal chow (data not shown).

The Ath and Ath+HF diets accumulated cholesterol in the liver after both 6 and 24 weeks. In addition to cholesterol, TG and FFA also accumulated with the Ath+HF diet. In comparison with hepatic lipid levels after 6 weeks, cholesterol and TG decreased in the livers of mice fed the Ath and Ath+HF diets after 24 weeks, and this indicated the progression of extensive hepatic fibrosis and impaired hepatocellular function. As is often found in patients with advanced liver disease, the serum ALT levels decreased with the progression of hepatic fibrosis, probably because of the impaired regeneration of hepatocytes and the production of ubiquitous liver enzymes.[22]

***Effects of the Ath Diet on Systemic or Hepatic Insulin Resistance.*** GTT and ITT after 12 weeks showed that the mice fed the Ath diet were remarkably sensitive to insulin (Fig. 3A,B). This ameliorating effect on the glucose tolerance and insulin sensitivity may be attributable to decreased adipose tissue in the mice fed the Ath or Ath+HF diet (Table 2). Therefore, we next evaluated the hepatic insulin sensitivity. For this purpose, we performed the pyruvate challenge test, an established method for evaluating hepatic insulin sensitivity,[15,16] by investigating the rise in blood glucose in response to the administration of pyruvate, a precursor for gluconeogenesis. The mice fed the Ath+HF diet showed an increased rise in the blood glucose concentration after pyruvate injection (Fig. 3C) compared with the mice fed the standard chow, and this suggested that the Ath+HF diet causes hepatic insulin resistance. Furthermore, as shown in Table 2, the homeostasis model assessment of insulin resistance (HOMA-IR) was significantly higher in the mice fed the Ath+HF diet than in the control mice. The expression of messenger RNA (mRNA) for phosphoenolpyruvate carboxykinase, the rate-controlling enzyme of

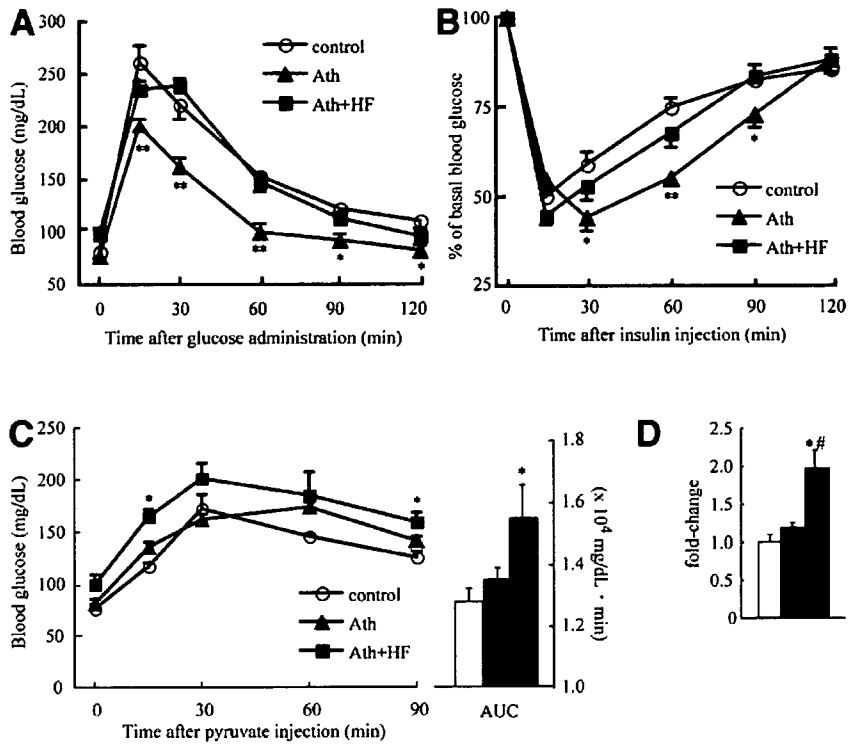


Fig. 3. Evaluation of glucose tolerance and insulin sensitivity. (A) GTT and (B) ITT after 12 weeks on standard chow (n = 4), the Ath diet (n = 5), or the Ath+HF diet (n = 5). (C) Pyruvate challenge test after 6 weeks on standard chow (n = 4), the Ath diet (n = 4), or the Ath+HF diet (n = 4). The area under the curve (AUC) of the blood glucose levels during the pyruvate challenge test was calculated. (D) mRNA levels of phosphoenolpyruvate carboxykinase genes in the livers of mice fed standard chow (white bar; n = 3), the Ath diet (gray bar; n = 3), or the Ath+HF diet (black bar; n = 3) after 12 weeks. The gene expression was normalized with eukaryotic 18S ribosomal RNA. The degree of change in the gene expression was based on the mean expression levels in control mice. The values represent the means ± the SEM. $*P < 0.05$ and $**P < 0.01$ versus the control group. $\#P < 0.05$ versus the Ath group.

gluconeogenesis for which the expression is negatively regulated by insulin, was significantly higher in the mice fed the Ath+HF diet than in the control mice (Fig. 3D). These results suggest that the Ath+HF diet causes hepatic insulin resistance.

***Ath Diet Induces Steatosis, Fibrosis, and Cellular Ballooning of the Liver.*** Figure 4 shows the time course of histological changes in the livers of mice fed the Ath or Ath+HF diet. The Ath diet induced progressive steatosis, inflammation, and fibrosis in a time-dependent manner from 6-24 weeks. Moreover, cellular ballooning, an important histological feature for the diagnosis of human NASH, was observed in the Ath group after 24 weeks. The addition of a high-fat component to the Ath diet accelerated the development of steatosis, inflammation, and fibrosis. Furthermore, before the Ath group, cellular ballooning was already observed in the Ath+HF group after 12 weeks. The hepatic hydroxyproline content, an indicator of collagen accumulation in the liver, increased significantly in the mice fed the Ath diet and increased further in the mice fed the Ath+HF diet (Fig. 4C). Therefore, the Ath diet induces steatohepatitis, and the addition of a high-fat component exacerbates the histological severity of steatohepatitis and hepatic insulin resistance.

***High-Fat Component Further Enhances the Activation of Hepatic Stellate Cells (HSCs) with the Ath Diet.*** The major sources of collagen and other extracellular matrix proteins in liver fibrosis are HSCs.[23] In response to stimuli such as oxidative stress and inflammatory cytokines, HSCs become activated and transform into proliferative fibrogenic cells.[24] We performed an immunohistochemical analysis of α-SMA, an activated HSC marker, at different times. Representative photomicrographs of liver sections stained with the anti–α-SMA antibody are shown in Fig. 5A. We quantified the areas in the liver sections positive for α-SMA morphometrically in the 3 groups at different times as described (Fig. 5A, lower panel). The activation of HSCs was promoted in the livers of mice fed the Ath diet in a time-dependent manner from 6-24 weeks and was further accelerated by the addition of a high-fat component to the Ath diet.

To evaluate oxidative stress causing HSC activation, we assayed proteins modified with 4-HNE, which is a major aldehyde end product of membrane lipid peroxidation due to oxidative stress (Fig. 5B). In concert with the increase in α-SMA–positive cells, 4-HNE–modified proteins accumulated in the livers of mice fed the Ath diet and further accumulated in those of mice fed the Ath+HF diet. In addition to 4-HNE–modified proteins, hepatic protein carbonyls, another marker of oxidative stress, also increased with the Ath and Ath+HF diets (Fig. 5C). These results are consistent with the observation that the Ath+HF diet induced more severe inflammation and fibrosis than the Ath diet.

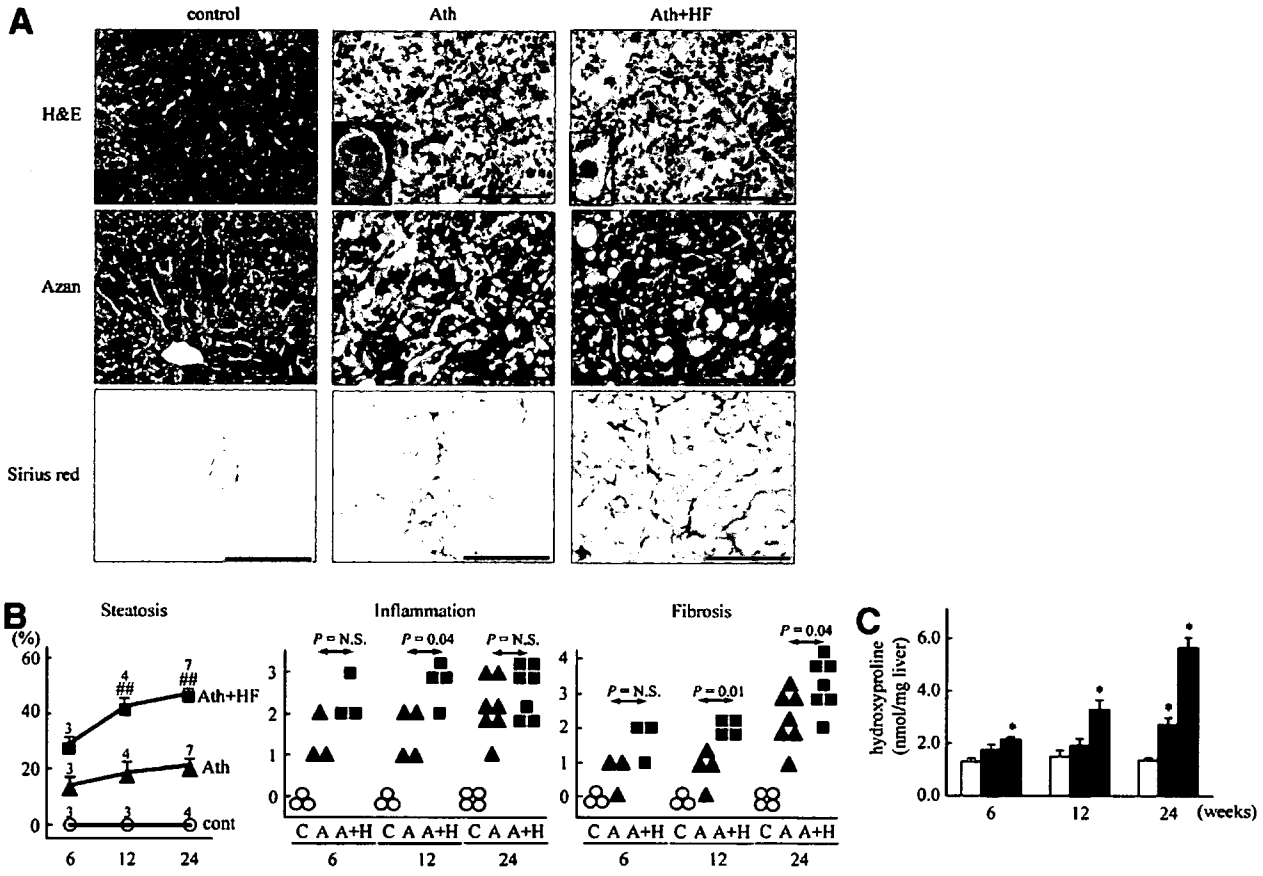


Fig. 4. Representative liver histology, scoring, and occurrence of hepatocyte ballooning. (A) Liver sections were stained with H&E, Azan, and Sirius red after 6, 12, and 24 weeks. The arrows indicate infiltration of the inflammatory cells in the hepatic parenchyma. The characteristic initial pattern of fibrosis in steatohepatitis is collagen deposition, as identified by blue and red staining. The original magnification was ×200. The scale bars represent 10 $\mu$m. Ballooning hepatocytes were seen only in the Ath and Ath+HF groups (shown in the inset). (B) The absolute percentage of the macrovesicular fat droplet area in the H&E-stained area was determined to evaluate steatosis. The values represent the means ± the SEM. The number of animals per group is indicated just above the points. ##$P < 0.01$ versus the Ath group. Inflammation and fibrosis scores were assigned in a blinded fashion to H&E-stained samples for inflammation and to Azan-stained samples for fibrosis. The criteria for each score are described in the Materials and Methods section. Differences in the inflammation and fibrosis histological scores between the Ath and Ath+HF groups were compared with the Mann-Whitney U test. The control, Ath, and Ath+HF groups are labeled C (white circles), A (gray triangles), and A+H (black squares), respectively. (C) The hydroxyproline content was determined in the livers of mice fed standard chow (white bars; n = 3), the Ath diet (gray bars; n = 3), or the Ath+HF diet (black bars; n = 3) at 6, 12, and 24 weeks. The values represent the means ± the SEM. **$P < 0.01$ versus the control group. ##$P < 0.01$ versus the Ath group.

***Gene Expression in the Livers of Mice Fed the Ath Diet.*** To address the molecular basis of Ath diet–induced steatohepatitis, we performed a microarray analysis, using livers at early (6 weeks) and precirrhosis stages (24 weeks) in the development of steatohepatitis. We screened 103 pathways determined with GenMAPP and extracted the metabolic pathways significantly altered in the livers of the mice fed the Ath and Ath+HF diets (Table 3). In the livers of the mice fed the Ath diet, genes involved in the inflammatory response and p38 mitogen-activated protein kinase (MAPK) signaling pathway were up-regulated significantly, whereas genes involved in fatty acid $\beta$-oxidation were down-regulated significantly in the early stage (6 weeks), and this was followed by coordinated up-regulation of the genes involved in fibrogenesis, such as the transforming growth factor $\beta$ (TGF-$\beta$) signaling pathway, in the late stage (24 weeks). Adding the high-fat component to the Ath diet accelerated the up-regulation of the genes involved in inflammation (electron-transport chain, p38 MAPK signaling pathway, and Fas pathway and stress induction) and fibrogenesis (TGF-$\beta$ signaling pathway and matrix metalloproteinases). Of these pathways altered in the models, we present the expression levels of representative genes involved in lipid metabolism, inflammation, oxidative stress, and fibrogenesis in Fig. 6 and Supplementary Table 2.

In the livers of mice fed the Ath diet, the expression of genes for fatty acid synthesis, such as sterol regulatory

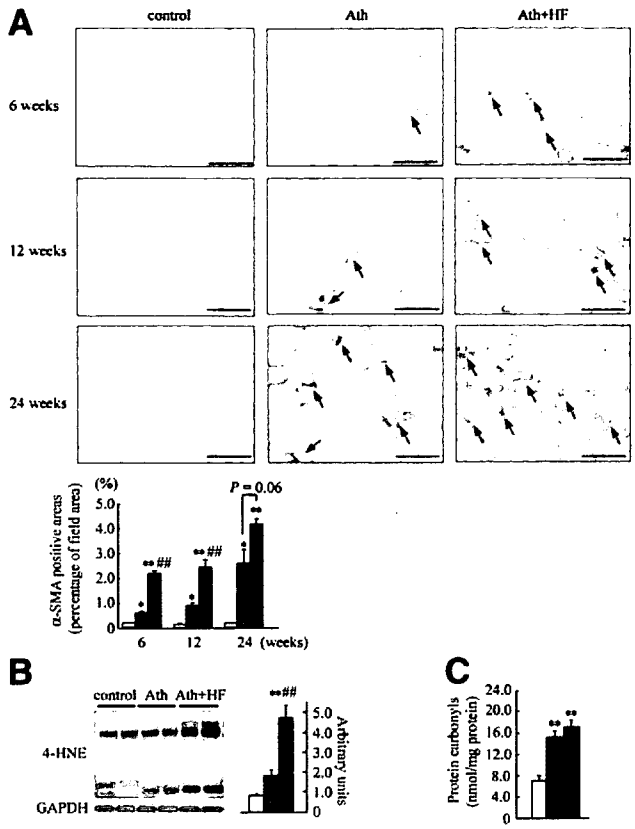


Fig. 5. Activation of HSCs and oxidative stress in the livers of mice fed the Ath or Ath+HF diet. (A) Hepatic α-SMA-positive cells (indicated by arrows) were detected by immunohistochemical staining at 6, 12, or 24 weeks. The original magnification was ×200. The scale bars represent 10 μm. The α-SMA-positive area was quantified morphometrically in the liver sections of mice fed standard chow (white bar; n = 3), the Ath diet (gray bar; n = 3), or the Ath+HF diet (black bar; n = 3) at different times, as described in the Materials and Methods section. (B) Western blot of 4-HNE-modified proteins in the liver after 24 weeks. The hepatic content of 4-HNE-modified proteins was quantified in mice fed standard chow (white bar; n = 4), the Ath diet (gray bar; n = 4), or the Ath+HF diet (black bar; n = 4), as described in the Materials and Methods section. (C) Hepatic protein carbonyls were determined in the mice fed standard chow (white bar; n = 3), the Ath diet (gray bar; n = 4), or the Ath+HF diet (black bar; n = 4) after 24 weeks, as described in the Materials and Methods section. The values represent the means ± the SEM. *$P < 0.05$ and **$P < 0.01$ versus the control group. #$P < 0.05$ and ##$P < 0.01$ versus the Ath group.

element binding protein 1c (SREBP-1c), a transcriptional regulator of fatty acid synthesis,[25] and fatty acid synthase (FAS), was coordinately up-regulated. In contrast, the expression levels of genes for the mitochondrial fatty acid β-oxidation pathway were coordinately repressed in concert with a decrease in the expression of peroxisome proliferator-activated receptor α (PPARα), a transcriptional up-regulator of fatty acid β-oxidation in the liver.[26] It is recognized that mitochondrial β-oxidation and the levels of carnitine palmitoyltransferase 1a (CPT-1a) and PPARα expression are increased compensatively in the livers of patients with NAFLD[27,28] and obese-diabetic (ob/ob) mice with severe steatosis of the liver.[29] Therefore, although the levels of PPARα and CPT-1a mRNA expression in the Ath+HF group were higher than those in the Ath group, it may not have been enough to metabolize the excessive fatty acids from the high-fat component and intrahepatic fatty acid synthesis.

It is believed that oxidative stress due to the generation of reactive oxygen species (ROS) or decreased antioxidant defenses is directly involved in the development of steatohepatitis.[30] The expression levels of genes for the reduced-form nicotinamide adenine dinucleotide phosphate (NADPH) oxidase complex, an important source of ROS,[31] were coordinately elevated in mice fed the Ath diet and further up-regulated in mice fed the Ath+HF diet.

The Ath diet has previously been reported to induce the expression of genes for inflammation.[32,33] Our results further demonstrate that inflammatory cytokines, such as tumor necrosis factor α (TNF-α), chemokines, and their receptors, are up-regulated in mice fed the Ath diet.

The Ath diet also induced genes involved in collagen accumulation, especially after 24 weeks. At 6 weeks, the expression levels of collagen genes were higher in the Ath+HF group than in the Ath group (Fig. 6). In addition, the expression levels of genes for TGF-β and plasminogen activator inhibitor 1 (PAI-1), key inducers of fibrogenesis, were dramatically up-regulated in the Ath+HF group compared with the Ath group at 24 weeks. These results support the finding that the Ath+HF diet induces more rapid progression of steatohepatitis than the Ath diet.

## Discussion

Whether cholesterol, TG, or FFA contributes to the development of NASH remains controversial.[34] Because the feeding of cholesterol and cholic acid, which are the main components of the Ath diet, leads to the additive accumulation of cholesterol in the liver, the main pathology in Ath diet–induced steatohepatitis is caused by cholesterol-induced toxicity.[35]

In this study, we have shown that Ath diet–induced steatohepatitis with atherosclerosis is a better experimental model of human NASH for the following reasons: (1) this model seems to be a more physiological dietary model of NASH than existing animal models, which require genetic defects, chemical agents such as carbon tetrachloride, or the depletion of nutrients, such as the MCD diet–induced model; (2) the liver pathology involves steatohepatitis with cellular ballooning, a necessary histological feature defining human NASH; (3) the addition of a high-fat component to the Ath diet causes hepatic insulin

**Table 3. Biological Pathways of Liver Genes Regulated by the Ath or Ath+HF Diets After 6 or 24 Weeks**

| Pathway Name | Number of Genes Changed | Number of Genes Measured | Z Score | Permuted *P* Value |
|---|---|---|---|---|
| **Ath diet** | | | | |
| ***Up-regulated at 6 weeks*** | | | | |
| Inflammatory Response | 23 | 41 | 3.22 | < 0.01 |
| DNA replication Reactome | 21 | 41 | 2.56 | 0.010 |
| Cell Cycle-G1 to S control Reactome | 32 | 68 | 2.56 | 0.016 |
| G1 to S cell cycle Reactome | 32 | 68 | 2.56 | 0.016 |
| RNA transcription Reactome | 20 | 40 | 2.36 | 0.036 |
| p38 MAPK signaling | 15 | 28 | 2.38 | 0.037 |
| ***Down-regulated at 6 weeks*** | | | | |
| Amino Acid Metabolism | 23 | 45 | 2.94 | < 0.01 |
| Cholesterol Biosynthesis | 11 | 15 | 3.56 | < 0.01 |
| Complement and Coagulation Cascades | 29 | 59 | 3.04 | < 0.01 |
| Mitochondrial fatty acid betaoxidation | 11 | 16 | 3.28 | < 0.01 |
| Blood Clotting Cascade | 11 | 18 | 2.77 | 0.012 |
| Unsaturated Fatty Acid Beta Oxidation | 5 | 6 | 2.78 | 0.014 |
| Biogenic Amine Synthesis | 8 | 14 | 2.13 | 0.042 |
| Krebs-TCA Cycle | 14 | 29 | 2.03 | 0.045 |
| ***Up-regulated at 24 weeks*** | | | | |
| mRNA processing binding Reactome | 196 | 438 | 5.91 | < 0.01 |
| TGF Beta Signaling Pathway | 62 | 124 | 4.37 | < 0.01 |
| Translation Factors | 27 | 49 | 3.50 | < 0.01 |
| Complement Activation Classical | 9 | 15 | 2.34 | 0.021 |
| ***Down-regulated at 24 weeks*** | | | | |
| GPCRDB Other | 52 | 147 | 3.58 | < 0.01 |
| Small ligand GPCRs | 11 | 19 | 3.61 | < 0.01 |
| Synthesis and Degradation of Ketone Bodies | 4 | 4 | 3.66 | < 0.01 |
| Mitochondrial fatty acid betaoxidation | 9 | 16 | 3.16 | < 0.01 |
| Cholesterol Biosynthesis | 8 | 15 | 2.79 | < 0.01 |
| Metabotropic glutamate pheromone | 6 | 10 | 2.78 | 0.020 |
| **Ath + HF diet** | | | | |
| ***Up-regulated at 6 weeks*** | | | | |
| Electron Transport Chain | 35 | 82 | 4.93 | < 0.01 |
| mRNA processing binding Reactome | 120 | 434 | 3.64 | < 0.01 |
| Translation Factors | 20 | 49 | 3.48 | < 0.01 |
| p38 MAPK signaling pathway | 13 | 28 | 3.36 | < 0.01 |
| Unsaturated Fatty Acid Beta Oxidation | 4 | 6 | 2.78 | 0.018 |
| Matrix Metalloproteinases | 9 | 24 | 2.03 | 0.034 |
| TGF Beta Signaling Pathway | 35 | 124 | 2.08 | 0.039 |
| Fas pathway and stress induction | 41 | 149 | 2.07 | 0.042 |
| ***Down-regulated at 6 weeks*** | | | | |
| Focal adhesion | 56 | 186 | 3.51 | < 0.01 |
| Steroid Biosynthesis | 8 | 12 | 4.06 | < 0.01 |
| Complement and Coagulation Cascades | 20 | 59 | 2.70 | < 0.01 |
| G Protein Signaling | 26 | 83 | 2.62 | 0.013 |
| Calcium regulation in cardiac cells | 41 | 145 | 2.54 | 0.014 |
| Cholesterol Biosynthesis | 7 | 15 | 2.60 | 0.016 |
| ***Up-regulated at 24 weeks*** | | | | |
| Translation Factors | 21 | 49 | 3.993 | < 0.01 |
| mRNA processing binding Reactome | 121 | 437 | 4.055 | < 0.01 |
| p38 MAPK signaling pathway | 12 | 28 | 3.016 | < 0.01 |
| TGF Beta signaling pathway | 35 | 124 | 2.077 | 0.039 |
| ***Down-regulated at 24 weeks*** | | | | |
| Amino Acid Metabolism | 19 | 45 | 3.891 | < 0.01 |
| Urea cycle and metabolism of amino groups | 10 | 20 | 3.472 | < 0.01 |
| Striated muscle contraction | 16 | 42 | 3.080 | < 0.01 |
| Steroid Biosynthesis | 6 | 12 | 2.689 | 0.015 |
| Small ligand GPCRs | 8 | 19 | 2.513 | 0.020 |
| Glycolysis and Gluconeogenesis | 14 | 41 | 2.402 | 0.023 |

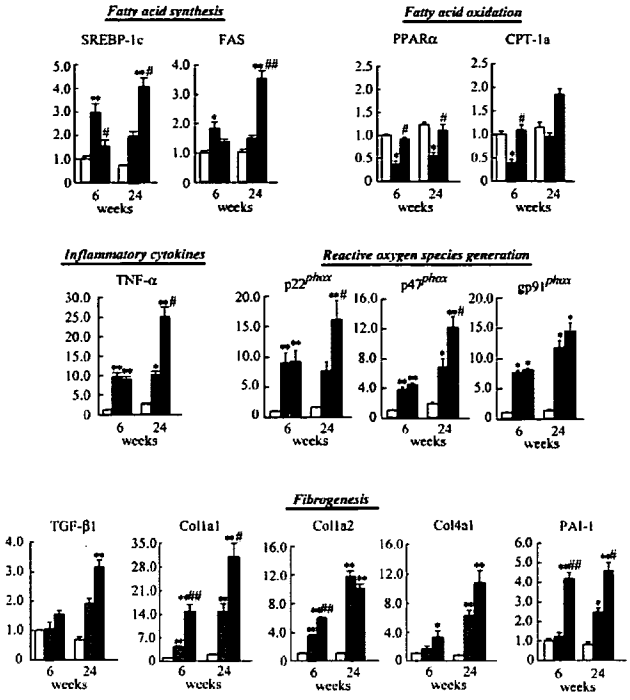


Fig. 6. Quantitative real-time PCR for representative genes involved in steatohepatitis. The mRNA levels of genes for SREBP-1c, FAS, PPARα, CPT-1a, TNF-α, $p22^{phox}$, $p47^{phox}$, $gp91^{phox}$, TGF-β1, procollagen type I alpha 1 (Col1a1), procollagen type I alpha 2 (Col1a2), procollagen type IV alpha 1 (Col4a1), and PAI-1 in the livers of mice fed standard chow (n = 3), the Ath diet (n = 3), or the Ath+HF diet (n = 3) were quantified with real-time PCR after 6 and 24 weeks. The RNA samples used for real-time PCR were the same as those used for the microarray analysis. The gene expression was normalized with eukaryotic 18S ribosomal RNA. The degree of change in the gene expression was based on the mean expression levels of control mice at 6 weeks. The values represent the means ± SEM. $*P < 0.05$ and $**P < 0.01$ versus the control group. $\#P < 0.05$ and $\#\#P < 0.01$ versus the Ath group.

resistance and promotes oxidative stress, the activation of HSCs, and all components of the liver pathology of NASH (steatosis, inflammation, fibrosis, and cellular ballooning); and (4) there is a molecular signature indicative of lipid-induced oxidative stress in the liver, which may play a causal role in the development of steatohepatitis.

To diagnose human NASH, cellular ballooning, in addition to simple steatosis and inflammatory cell infiltration, is one of the most important pathological features.[36] However, ballooning degeneration has scarcely been determined in the existing animal models, including mice fed the MCD diet. We believe that our study is the first to report that cellular ballooning is frequently induced in the livers of mice fed the Ath diet.

Recently, we proved that insulin resistance accelerates the pathological development of steatohepatitis experimentally.[11] In this study, on the basis of the results of the pyruvate challenge test and HOMA-IR, we concluded that the Ath+HF diet causes hepatic insulin resistance. It is known that the excessive accumulation of FFAs caused by the overexpression of lipoprotein lipase[37] and an increase in SREBP-1c–regulated lipogenesis[38] leads to impaired tyrosine phosphorylation of IRS-1 and IRS-2, resulting in hepatic insulin resistance. Furthermore, the up-regulation of SREBP-1c–regulated lipogenesis contributes to the development of insulin resistance via the down-regulation of IRS-2 in the liver.[39,40] Indeed, in our study, the induction of lipoprotein lipase and SREBP-1c and the repression of IRS-2 were detected in the livers of mice fed the Ath diet (Fig. 7). Moreover, the up-regulation of stearoyl–coenzyme A desaturase 1, an enzyme that catalyzes the synthesis of monounsaturated fatty acids, might contribute to lipid accumulation and insulin resistance in the liver, as reported in skeletal muscle.[41] Therefore, the cholesterol-induced and TG-induced alteration of fatty acid metabolism may cause hepatic insulin resistance in this model of steatohepatitis.

Another possible cause of the liver pathology in our model is lipid-induced oxidative stress and its downstream events, as we identified an accumulation of 4-HNE and protein carbonyls, the activation of stellate cells, and hepatic inflammation with cell ballooning. In this study, in addition to cholesterol, the accumulation of TG and FFAs by the addition of a high-fat component accelerated oxidative stress, possibly via the up-regulation of genes involved in the generation of ROS, such as the NADPH oxidase complex, and the down-regulation of genes for antioxidant enzymes. While we were preparing this article, Mari et al.[35] reported that the mitochondrial loading of free cholesterol, but not TG and FFA, decreases mitochondrial glutathione and sensitizes it to the TNF-α–mediated apoptosis of hepatocytes. Therefore, the different kinds of accumulated lipids may cause oxidative stress in the liver additively in different ways. In

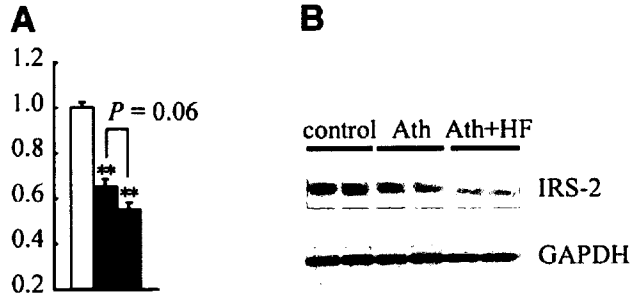


Fig. 7. The Ath and Ath+HF diets decreased the mRNA and protein levels of IRS-2 in the liver. (A) mRNA levels of the IRS-2 genes in the livers of mice fed standard chow (white bar; n = 3), the Ath diet (gray bar; n = 3), or the Ath+HF diet (black bar; n = 3) after 12 weeks. The values represent the means ± the SEM. $*P < 0.05$ versus the control group. $\#P < 0.05$ versus the Ath group. (B) Western blot of IRS-2 in the livers of mice fed the standard chow, Ath diet, or Ath+HF diet after 12 weeks.