This article contains supplementary material available via the Internet at http://www.interscience.wiley.com/jpages/1552-4922/suppmat.

Received 14 March 2007; Revision Received 19 July 2007; Accepted 11 September 2007

Grant sponsors: Toyama Medical Bio-Cluster Project, Hiroshima Bio-Cluster Project of the Ministry of Education, Culture, Sports, Science and Technology, Japan

the specific antibodies to rabies (4), but only 0.01% of B-cells produce antibodies to hepatitis B virus surface (HBs) antigen (HBs-Ag) after vaccination (5). To analyze the response of individual B-cells to an antigen, it is necessary to analyze a large number of cells at the level of a single cell by cytomics.

To meet the demand for this type of analysis, we have developed an analysis system using a microwell array chip that has a large number of microwells whose size and shape just fit a single cell (6). By applying an individual cell to each microwell, an array of live cells was prepared, and the cellular responses of individual cells, such as alteration of intracellular Ca^{2+} concentration, were monitored using a fluorescent scanner that was modified to scan cells. Because the position of each cell was fixed, we could repeatedly analyze the cellular response of the same cell. Consequently, a cell microarray system, a combination of a microwell array chip and a cell scanner, enabled us to analyze the time course of the cellular responses of a large number of cells at the single-cell level. Recently, Deutsch et al. (7) and Biran and Walt (8) reported live cell arrays for analyzing a large number of cells at the single-cell level. These cell arrays were chiefly designed for analyzing cells, and the retrieval of objective cells from wells without disturbing surrounding cells might be difficult. In contrast, our cell microarray system was aimed at not only analyzing cells but also retrieving objective cells from an array.

Hepatitis B virus (HBV) infection is one of the world's major health problems, especially in East Asia. It causes self-limiting acute hepatitis, but the infection may often become chronic, causing hepatic cirrhosis and hepatic cell cancer. Vaccines based on recombinant DNA technology have been developed and applied for the protective immunization of humans. Such vaccination, however, does not always provoke a sufficient and rapid antibody response, and HB immunoglobulins (HBIg) have been employed in combination with HB vaccination (9). HBIg injection, however, poses some serious problems such as contamination by unknown infectious agents and lack of a continuous supply in hospitals. Human monoclonal antibodies (MoAb) represent an alternative to HBIg therapy.

We have applied our cell microarray system to detect human B-cells that respond to HBs-Ag from the peripheral blood of HBs-Ag-vaccinated volunteers, and obtained HBs-Ag-specific MoAb that neutralized HBV to prevent its infection of human hepatocytes. Our technology may contribute to the production of human MoAb not only for HBV but also for other various human health-threatening infectious agents, such as AIDS and SARS as well as microbes that might be used for bioterrorism.

Materials and Methods

Microwell Array Chip

A microwell array chip was manufactured using micromachining (microelectromechanical system) techniques (10) (Supplementary Fig. 1). Briefly, a thin film of silicon dioxide was grown on a silicon surface by thermal oxidation (11), a photoresist was coated on the thin film, and microwell patterns were transferred from a mask via photolithography using a Karl Suss MA6 Mask Aligner (SUSS MicroTec AG, Garching, Germany). Then, the exposed oxidized silicon surface was etched by silicon deep reactive ion etching (12) to form microwells using buffered hydrofluoric acid as a solvent of silicon dioxide. Then, a fluorocarbon polymer was formed on the photoresist and the sidewall of a microwell using plasma-enhanced chemical vapor deposition (13). Finally, the photoresist was removed and the fluorocarbon polymer was lifted off the surface of the silicon chip.

Preparation of Lymphocytes for Intracellular Calcium Analysis

To analyze the efficiency of the cell microarray system, we used splenocytes of MD4 transgenic mice (C57BL/6-Tg(IghelMD4)4Ccg/J from The Jackson Laboratory, Bar Harbor, ME), of which the transgene encodes mouse antibody (HyHEL10) for hen egg lysozyme (HEL) (14). The splenocytes were prepared and loaded with 0.1 μM CellTracker orange (Ex, 535 nm; Em, 585 nm; Invitrogen, Carlsbad, CA) and 0.1 μM Fluo-4 (Ca^{2+}-dependent fluorophore; Ex, 473 nm; Em, 532 nm; Invitrogen), as previously described (6). The procedure for the transgenic mice experiments was approved by the Committee for Recombinant DNA Experiments (#19-9) and Animal Experiments at the University of Toyama (# 2006-Med-33). The mice were examined to determine whether they were transgenic by staining peripheral blood lymphocytes with biotinylated HEL and PE-conjugate of streptavidin. The mice with HEL-specific Ab-expressing B-cells were used for the experiments.

For preparation of human B-cells, peripheral blood lymphocytes were isolated from healthy donors according to the standard Ficoll–Hypaque method (Lymphosepal; IBL, Takasaki, Japan). B-cells were purified using MACS (Miltenyi Biotec K.K., Tokyo, Japan), and loaded with CellTracker orange and Fluo-4.

Cell Microarray Analysis and Antibody Preparation

Cell microarray analysis was performed as previously described (6). Briefly, cells were loaded onto a microwell array

chip and the Fluo-4 fluorescence of the individual cells before stimulation was measured by scanning the chip with a cell scanner (CRBIO IIe-FITC, Hitachi Software, Tokyo, Japan) that was modified from a DNA chip scanner (CRBIO IIe) by changing a 635-nm laser to a 473-nm laser and whose minimum resolution was improved to 2.5 μm. The cells were then stimulated with antigen at room temperature in air by exchanging the buffer on the chip with buffer containing antigen, and the cellular Fluo-4 fluorescence was measured with the scanner after stimulation. The Fluo-4 fluorescence intensities of the individual cells before and after antigen-stimulation were plotted in a scatter diagram with TIC-Chip Analysis software (Hitachi Software). Cells whose fluorescence increased more than fivefold were retrieved as antigen-activated B-cells from each well with a micromanipulator (TransferMan NK2, Eppendorf, Hamburg, Germany). To prepare an antibody from a retrieved B-cell, antibody cDNA was amplified with single-cell RT-PCR, inserted into an expression vector, and transfected into 293T cells to obtain a supernatant containing the antibody as previously described (6) (see supplementary Fig. 2).

ELISA for Detection of Anti-HBs Antibody

Maxisorp 96-well plates (Nunc, Roskilde, Denmark) were coated with 50 μl/well of 10 μg/ml HBs-Ag (Kaketsuken, Kumamoto, Japan) in phosphate buffered saline (PBS) and then blocked with 3% casein in PBS. After washing, cell culture supernatant containing the antibody was added to the plates and incubated for 15 min at room temperature. The binding of human antibody to the coated antigen was detected using alkaline phosphatase-labeled anti-human immunoglobulins and *p*-nitrophenylphosphate. The optical absorbance was measured at 414 nm with an ELISA reader (Labsystem Japan, Tokyo, Japan). To confirm the antigen-specificity of an antibody, a competitive binding assay was performed. Briefly, in the ELISA assay described earlier, 0.4, 2.0, 10 μg/ml soluble HBs-Ag was added together with the culture supernatants that contained the anti-HBs antibody.

Estimation of the Epitope

For the estimation of the anti-HBs antibody epitope, competitive ELISA was performed with mouse anti-HBs MoAb (anti-d, Institute of Immunology, Takasaki, Japan; anti-a and anti-r, provided from T. Nakashima, Kaketsuken) whose epitopes were already determined. Briefly, to Maxisorp 96-well plates coated and blocked as described earlier was added 50 μl/well of diluted mouse anti-HBs MoAb as a competitor, followed by 50 μl/well of 20 ng/ml sample antibody. After 15-min incubation at room temperature, binding of the sample antibody was assessed as described earlier.

In Vivo Neutralizing Activity Assay

To examine the HBV neutralization activity of the antibodies, chimeric mice having human hepatocytes were used. The chimeric mice were produced by transplanting human hepatocytes into albumin enhancer/promoter-driven urokinase-type plasminogen activator-transgenic SCID mice (uPA/SCID mice) (15). The transplanted human hepatocytes were shown to be infected with hepatitis C virus and HBV (16,17). Human serum albumin produced from the transplanted human hepatocytes was in the range of 1.3–5.2 mg/ml, indicating that 32–66% of the liver cells were replaced with those of human origin. Serum of a chronic hepatitis B patient was used for the preparation of HBV. The titer of the virus was determined by quantitative PCR of the virus genome as previously described (16). Primers for PCR are listed in supplementary Table 1. For the test of neutralization activity, the patient serum that contained 1.0×10^6 copies of the virus was mixed with 30 μg of either control or anti-HBs antibodies, incubated at room temperature for 30 min, and then intravenously injected into mice. For the estimation of neutralization activity, virus titers in the sera of the infected mice were measured every 2 weeks as described earlier.

Results

A Live Cell Microarray System for Lymphocytes

A general view of the live cell microarray system that we have developed is shown in Figure 1A. The cell suspension is prepared from blood or tissue, and loaded with fluorophore whose intensity alters with cell conditions, such as Ca^{2+} concentration, membrane potential, and pH. Then, the live cells are applied onto the microwell array chip that contains an array of 234,000 wells in which only a single cell can be trapped. Cells are stimulated with or without stimulants such as infectious reagents, and the alteration in fluorescence is monitored with a cell scanner. The signal of each cell is analyzed on a single-cell basis. Because the position of each cell on the chip is fixed, the signal of any one cell can be repeatedly analyzed. Then, individual cells of interest are identified, retrieved, analyzed at the molecular level, and used for protein engineering.

To prepare a live cell microwell array chip for lymphocytes, we manufactured a microwell array chip as shown in supplementary Figure 1. The microwell array chip was fabricated on a silicon substrate by using a micromachining technique (10). A scanning electron micrograph of the chip is shown in Figure 1B. To smoothly apply single cells to and retrieve them from the microwells, we made the surface of the microwell array chip hydrophilic by growing silicon dioxide using a thermal oxidation process (11), and the bottom and the sidewall hydrophobic by coating with a thin fluorocarbon film. To prepare the live lymphocyte array, the lymphocytes were separated, loaded with fluorescent dye, and applied to the microwell array chip. The chip was covered with a cover glass to prevent drying. The array of the live lymphocytes was examined under a fluorescence microscope (Fig. 1C). To raise the array rate and to efficiently hold the single lymphocytes in the microwells, the sizes (diameter and depth) as well as shapes of the microwells were optimized (Fig. 1D and data not shown). When the depth of the microwells is less than 10 μm, the lymphocytes in the microwells were easily moved out during a washing step. As a result, cylindrical microwells with a diameter of 10 μm and a depth of more than 12 μm were most suitable for raising the array rate and efficiently holding the

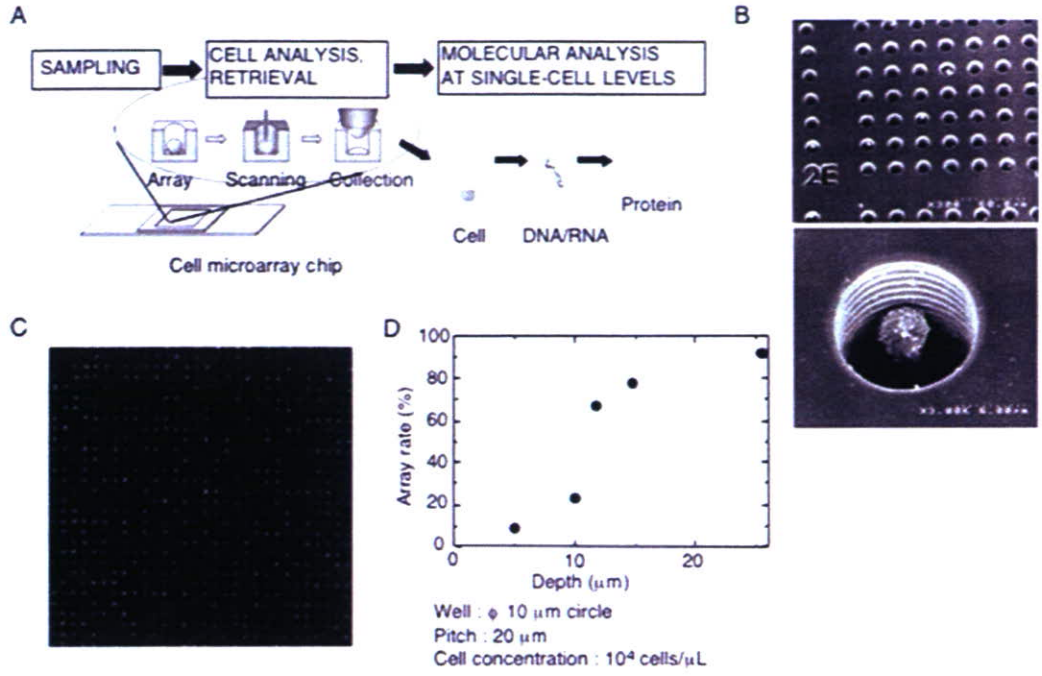

Figure 1. A cell microarray system. (**A**) General view of the cell microarray system. Lymphocytes are spread in the microwell array chip and their cellular responses are monitored. The responding cells are then retrieved and used for molecular analysis and protein engineering at the single-cell level. (**B**) Scanning electron micrograph images. Wide view (top) and single-well view (bottom) are shown. The lymphocyte became smaller due to the fixation process. (**C**) Picture of live cell array. Lymphocytes are loaded with CellTracker orange, arrayed on the chip, and observed under a fluorescence microscope. (**D**) Relationship between well depth and cell array rate. 50 μl of 10^4 cells/μl lymphocytes is applied on a chip. The percentage of wells that trap lymphocytes after any untrapped cells are washed away is the array rate.

lymphocytes. The relationship between the spacing of the microwells and number of arrayed cells was also investigated. When the spaces between the microwells increased, the number of arrayed cells decreased (data not shown). Based on these data, a microwell array chip with wells at a pitch of 20 μm, a diameter of 10 μm, and a depth of 15 μm were used in the following experiment.

Detection of Antigen-Stimulated Activated B-Cells on the Chip, and Comparison of Efficiency with a Flow Cytometry System

We first assessed the efficiency of the cell microarray system using mouse B-cells prepared from MD4 mice that carry a transgene encoding antibody (HyHEL10) to HEL (14). It has been reported that when MD4 B-cells, which express HyHEL10 antibody on their cell surface, are stimulated with HEL, the cells are activated and their intracellular Ca^{2+} concentrations increase (18). We prepared MD4 B-cells, loaded them with a fluorescent Ca^{2+} indicator, Fluo-4 (19), and arrayed them on the chip. The fluorescence of the cells was monitored with a cell scanner that was equipped with a 473-nm laser to activate and monitor the fluorescence of Fluo-4. Before stimulation, the fluorescence of each cell was measured and analyzed with the cell scanner at the single-cell level (left panel of Fig. 2A). Then, MD4 splenocytes were stimulated with HEL and their fluorescence was monitored with the cell scanner (right panel of Fig. 2A). The cell scanner scans the total cell area between 30 and 90 s after the stimulation. Because the address of each cell on the chip is fixed, we could compare the fluorescence intensity of each cell before and after stimulation by using analysis software. As shown in Figure 2B, dots corresponding to cells whose fluorescence was unchanged after the stimulation were located on the "$y = x$" line. When B-cells were activated with HEL, the intracellular Ca^{2+} level was increased, and the fluorescence intensity was augmented by a factor of about five. Dots corresponding to such activated B-cells shifted upward and were discriminated from those of the unstimulated cells. If cells with a twofold increase in Fluo-4 fluorescence were considered as positive cells, 49% of the cells were positive. If cells with a fourfold increase in fluorescence were considered as positive cells, 17% of the cells became positive. The earlier results demonstrate that a combination of a microwell array chip and a cell scanner (cell microarray system) can monitor the activation of about 200,000 individual cells by monitoring the alteration of intracellular Ca^{2+} concentration. We then compared the efficiency of the cell microarray system for detection of activated MD4 B-cells with that of a flow cytometry system (Figs. 2B and 2C). Splenocytes prepared from MD4 transgenic mice or normal mice were prepared and loaded with Fluo-4. These cells were mixed

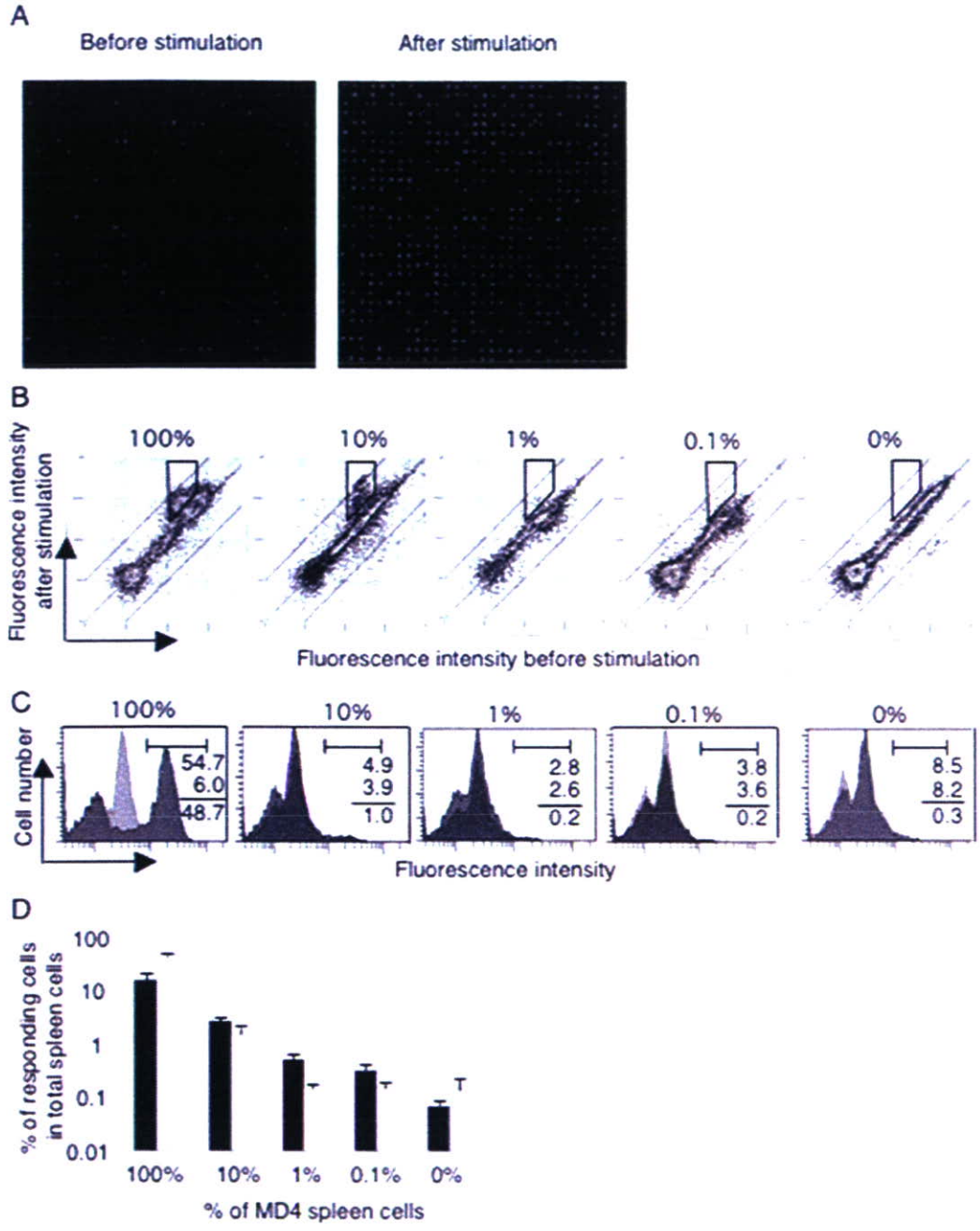

Figure 2. Analysis of B-cells with the cell microarray system. (**A**) Scanned image of intracellular Ca^{2+} signals of B-cells arrayed on a microwell array chip before and after stimulation. MD4 B-cells were loaded with Fluo-4, stimulated with HEL on the chip, and scanned with a cell scanner before (left) and after (right) stimulation. (**B**) Scattered plots of Fluo-4-intensities of MD4 B-cells before and after HEL-stimulation. Various percentages (100%, 10%, 1%, 0.1%, and 0%) of MD4 splenocytes were prepared by mixing with normal splenocytes, loaded with Fluo-4, arrayed on a chip, and stimulated with HEL (10 μg/ml). The cells were scanned with a cell scanner and Fluo-4 fluorescence signals before (*x*-axis) and after (*y*-axis) stimulation were calculated and plotted on scatter diagrams. Each spot corresponds to one cell. Cells whose fluorescence was not altered correspond to dots on the blue line ($y = x$). Dots of the activated cells shifted upward. Cells with more than a fourfold increase in Fluo-4 fluorescence intensity were considered as responding cells and enclosed with a box. (**C**) The same MD4 B-cell preparation was analyzed with a flow cytometer. Fluorescence histograms of lymphocytes before (light gray) and after (dark gray) stimulation are shown. The percentages of fluorescence positive cells after stimulation, before stimulation, and their differences are shown from top to bottom in each panel. (**D**) Comparison of cell microarray system and flow cytometry for detecting activated cells. The percentages of HEL-activated MD4 B-cells detected with either the cell microarray system (black column) or flow cytometry (open column) were calculated as in Figures 2B or 2C. The results represent the average of three independent experiments, and error bars represent the standard deviation.

with various ratios such as 100%, 10%, 1%, and 0.1% MD4 spleen cells. Cell preparation was applied on the microwell array chip and stimulated with HEL, and the fluorescence of each B-cell before and after stimulation with HEL was measured with the cell scanner (Fig. 2B). To compare the detection efficiency with flow cytometry, the spleen cell mixture was stimulated with HEL and the alteration of fluorescence intensity was also measured with a flow cytometer (Fig. 2C). Figure 2D shows the percentages of positively detected cells with the cell microarray system or flow cytometry. We selected cells with a more than fourfold increase in Fluo-4 fluorescence after stimulation with antigen as antigen-activated B-cells in the cell microarray analysis. For the flow cytometry analysis, we calculated the positive cells by subtracting the percentage of positive cells before stimulation from the percentage of positive cells after stimulation. As noted, we observed the cell population with high Fluo-4 fluorescence intensity before stimulation, which was detected not only in the cell microarray system but also in flow cytometry (Figs. 2A–2C). The cell microarray system could distinguish the false positive cells and antigen-stimulated cells whose Fluo-4 fluorescence signals increased after antigen stimulation because each cell address was fixed on the chip and the fluorescence levels of the same cell before and after stimulation could be compared. In contrast, flow cytometry could not make the same distinction because it could not repeatedly monitor the signals of specific cells. Therefore, the detection of antigen-stimulated B-cells of low frequency with flow cytometry was hampered by the false positive cells, but was not hampered with the cell microarray system (Fig. 2D). In the case of the microwell array chip, the percentage of positively detected cells was linearly decreased with the dilution of MD4 spleen cells and about 0.2% of the positive cells were still detected in 0.1% of the MD4 spleen cells. The background of the positive cells was around 0.08% (Figs. 2B and 2D). In contrast, an analysis of less than 1% of the MD4 spleen cells with the flow cytometry system could not show the difference because of the false positive cells (Figs. 2C and 2D).

Application to the Detection of HBV-Specific B-Cells and Generation of MoAb with Neutralizing Activity of HBV Infection

Because the efficacy of detecting single cells using the cell microarray system was verified, we tried to apply the system to detect human antigen-specific B-cells and prepare human antigen-specific MoAb. Volunteers were vaccinated with recombinant HBs-Ag according to the vaccination protocol. Seven days after the last vaccination, B-cells were prepared from peripheral blood, loaded with Fluo-4, applied on a microwell array chip, and stimulated with HBs-Ag on the chip. The fluorescence intensity of individual cells before and after stimulation was measured with the cell scanner. As shown in Figure 3A, the fluorescence intensity of the minor cell population increased with the HBs-Ag stimulation. Figure 3B shows the percentage of activated B-cells before and after the boost of HBs-Ag in the volunteers. The HBs-Ag boost almost doubled the percentage of positive cells (Fig. 3B). We then retrieved the positive cells from the microwells using a micro-

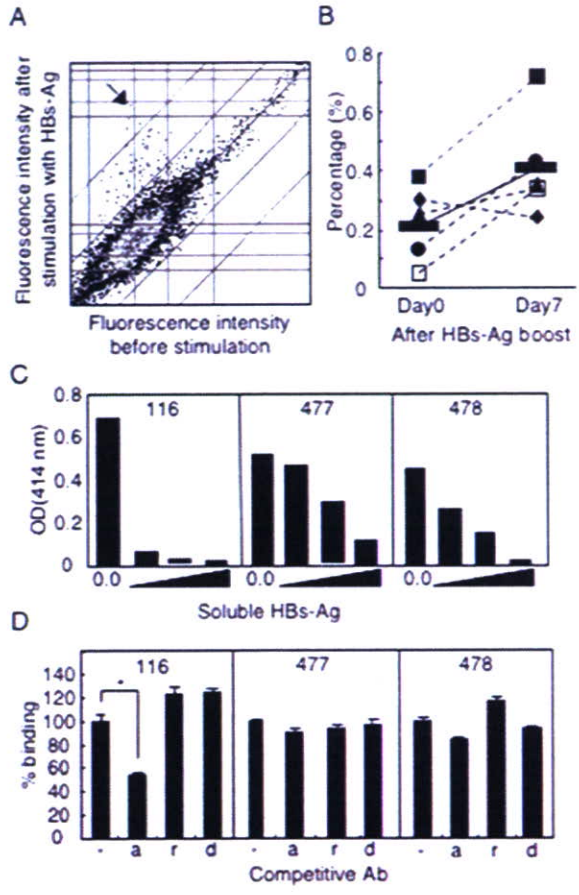

Figure 3. Detection of HBs-Ag-specific B-cells and production of HBs-Ag-specific antibodies with the cell microarray system. (**A**) Scatter diagram of peripheral blood B-cells before and after stimulation with HBs-Ag (100 μg/ml). (**B**) Percentage of HBs-Ag-stimulated B-cells in peripheral blood B-cells before and after application of HBs vaccine. Black bars show the average. (**C**) Antigen specificity of anti-HBs antibody. The antigen specificity of 116, 477, and 478 antibodies were examined with a competitive binding assay. Binding of the antibodies to plate-coated HBs-Ag was examined in the presence of various doses (0.4, 2, 10 μg/ml) of soluble HBs-Ag. Representative data of three to four independent experiments are shown. (**D**) Determination of epitope of anti-HBs antibodies. Binding of the antibodies to plate-coated HBs-Ag was examined in the presence of competitive MoAb to specific epitopes (a, r, d) on HBs-Ag. $^*P < 0.01$.

manipulator and the heavy chain and light chain variable regions of cDNA were augmented from a single B-cell by RT-PCR. The cDNAs were inserted into expression vectors, which were then cotransfected to 293T cells derived from human embryonic kidney (Supplementary Fig. 2). The culture supernatants of the 293T cells were collected and the specific binding activity of the antibody was analyzed with ELISA. Sixty-six IgM and 12 IgG cDNAs were cloned from 377 individual B-cells. In 12 IgG cDNAs, 3 IgG (116, 477, and 478) were found to bind HBs-Ag specifically (Fig. 3C). It has been reported

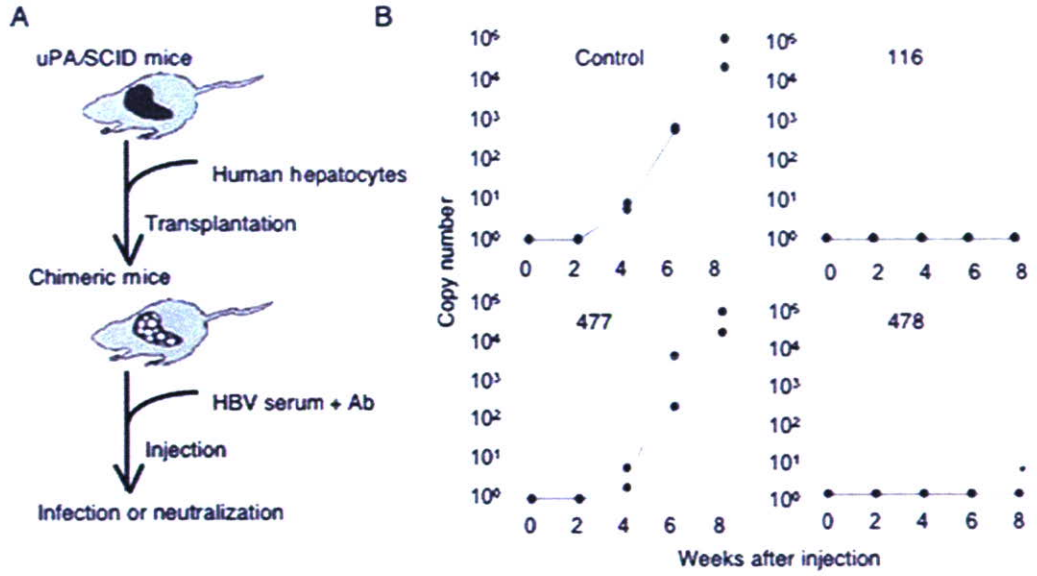

Figure 4. Assessment of virus neutralization activity of anti-HBs antibodies. (**A**) Scheme of protocol for examination of HBV neutralization activity with human hepatocyte-chimeric mice. (**B**) Effect of anti-HBs antibodies on HBV infection. A mixture of HBV and anti-HBs antibodies or control antibody was injected i.v. into human hepatocyte-chimeric mice, and the virus titer in mouse serum was determined by a quantitative viral genome PCR analysis ($n = 2$). *One mouse died during the 8-week study. Representative data of two independent experiments are shown.

that there are three major epitopes, a, d/y, and r/w on HBs-Ag (20). The epitopes recognized by the IgG antibodies were examined by competition with epitope-specific antibodies. As shown in Figure 3D, 116 was found to bind epitope "a" since the "a"-epitope-specific antibody inhibited the binding of 116 to HBs-Ag. Because the other antibodies (477 and 478) were not competed with the antibodies used, we could not determine their epitopes. The affinities of the 116, 477, and 478 antibodies were determined with surface plasmon resonance. It was revealed that their K_d were 1.2×10^{-9} M, 5.9×10^{-7} M, and 1.4×10^{-7} M, respectively (data not shown).

Finally, we tested to see if these antibodies inhibit the infection of HBV in human hepatocytes. We used chimeric mice to which human hepatocytes had been transplanted (16,17) (Fig. 4A). We injected i.v. 1.0×10^6 copies of HBV into the mice in the presence of control, 116, 477, or 478 antibody. Infection and replication of the virus was determined by detecting the viral genomic DNA in the serum with PCR. Figure 4B shows that the virus DNA was detected 4 weeks later in the chimeric mice injected with the virus together with either the control or 477 antibody. In contrast, the virus DNA was not detected in the serum of the mice that had been injected with the virus together with either the 116 or 478 antibody, demonstrating that these antibodies showed the neutralizing activity.

Discussion

We described a live cell microarray chip, which arrays a large number of nonadherent cells, such as lymphocytes, by accommodating them in microwells whose size and shape are optimized to trap only a single cell. By stimulating individual cells on the chip and monitoring cell signals such as intracellular Ca^{2+} levels before and after antigen-stimulation, we could efficiently detect the activation of individual cells. Since the cell scanner repeatedly scans 234,000 cells in about 2 min/scan, it can follow the history of cell signals every 2 min. We verified that this system is more suitable for analyzing the "cytome" (21,22) when compared with microscopy and flow cytometry-based systems, which cannot analyze the history of the cellular responses of large numbers of cells at the level of single cells. Using a novel microwell array chip, we could successfully analyze human B-cells of volunteers who had been vaccinated with HBs-Ag and generate some MoAb that were able to neutralize the HBV infection of human hepatocytes.

Regarding the viability of cells on the chip, Fluo-4 fluorescence intensities of about 90% of the B-cells were increased by stimulation with anti-IgM antibodies (data not shown). As about 90% of B-cells express IgM on the cell surface, the result indicates that most of the B-cells were alive on the chip during the assay. The diameter and depth of the microwells were critical in preparing an array of single cells (data not shown and Fig. 1D). When the diameter or the depth was insufficient, array rate was decreased, and when they were too great, fluorescence signals of two to three cells were observed under a fluorescence microscope (data not shown). In this study, well size and shape were optimized to those of lymphocytes, so that we observed single fluorescence signals from each well, as shown in Figure 1C.

Recently, Deutsch et al. reported a live cell array to perform cell-based assays on thousands of individual cells (7).

Microcavities were arrayed in a honeycomb-like structure and could monitor cell responses, such as reactive oxygen species generation, on a single-cell basis (23). Biran and Walt reported an optical imaging fiber-based single live cell array for analyzing yeast or bacterial cells (8). These cell arrays were specifically designed for analyzing cells, and it may be difficult to retrieve objective cells from wells without disturbing the surrounding cells. In our cell microarray, each well was separated so that we could retrieve single objective cells without disturbing the surrounding cells.

Therapeutic MoAb have been developed into beneficial and profitable medical products in molecule-targeted therapeutics (24). To produce fully human MoAb, the bacteriophage display method and the Epstein-Barr virus method have been developed (25,26). The bacteriophage display method requires the preparation of a large-scale bacteriophage library ($\sim 10^{11}$ clones) to get an antigen-specific antibody, making the method difficult to perform in ordinary laboratories. With the availability of a cell scanner, microwell array chips, and a micromanipulator, our system is easy to operate and can be performed in any laboratory that is so equipped. Recently, we have developed a prototype of equipment that can automatically perform processes from cell application to cell retrieval, making the system more appropriate as a procedure for producing human MoAb. Concerning the Epstein-Barr virus method, the virus can transform human B-cells to produce human antibodies. Because only a part of the B-cells can be transformed with the virus, the method cannot efficiently screen antigen-specific B-cells. Our system can screen antigen-specific B-cells directly from freshly separated human peripheral blood lymphocytes and produce antigen-specific MoAb.

We showed that the human anti-HBs antibodies that we had produced exhibited neutralization activity to inhibit HBV infection of human hepatocytes. To demonstrate the neutralization activity, we constructed an estimation system using chimeric mice transplanted with human hepatocytes (16). As HBV can infect only fresh human hepatocytes or that of primates, but not that of rodents, examination of neutralization activity has mostly been performed on chimpanzees (27), which is very expensive and might be contrary to the principles of animal welfare. The use of chimeric mice described in this study is simple and not expensive when compared with the conventional method using chimpanzees because mice are small, easy to breed, and much less expensive.

In conclusion, we have demonstrated and described a system for analyzing a large number of cells on a single-cell basis, which can be applied to "cytomics" analysis, and to the production of human MoAb directly from human peripheral blood B-cells. The system might be applicable for analyzing T-cells to detect antigen-specific T-cells and for cloning TCR cDNAs. By changing the chip design, we are able to analyze various kinds of cells such as hybridoma, hepatocytes, and nerve cells as well as lymphocytes. Our technology should expand the horizons of cell analysis as well as the generation of human MoAb for antibody-based therapeutics and diagnosis of hepatitis virus infection.

Acknowledgments

We thank M. Suzuki, I. Maruyama, H. Nakazato, and Y. Shimizu for their helpful contributions to our discussions, T. Nakashima for providing anti-HBs MoAb, S. Hirota for technical assistance, and K. Hata for secretarial assistance.

Literature Cited

1. Valet G, Leary JF, Tarnok A. Cytomics—New technologies: Towards a human cytome project. Cytometry Part A 2004;59A:167–171.
2. Brehm-Stecher BF, Johnson EA. Single-cell microbiology: Tools, technologies, and applications. Microbiol Mol Biol Rev 2004;68:538–559.
3. Abbas AK, Lichtman AH. Cellular and Molecular Immunology. Philadelphia, PA: Saunders; 2003.
4. Ueki Y, Goldfarb IS, Harindranath N, Gore M, Koprowski H, Notkins AL, Casali P. Clonal analysis of a human antibody response. Quantitation of precursors of antibody-producing cells and generation and characterization of monoclonal IgM, IgG, and IgA to rabies virus. J Exp Med 1990;171:19–34.
5. Shokrgozar MA, Shokri F. Enumeration of hepatitis B surface antigen-specific B lymphocytes in responder and non-responder normal individuals vaccinated with recombinant hepatitis B surface antigen. Immunology 2001;104:75–79.
6. Yamamura S, Kishi H, Tokimitsu Y, Kondo S, Honda R, Rao SR, Omori M, Tamiya E, Muraguchi A. Single-cell microarray for analyzing cellular response. Anal Chem 2005;77:8050–8056.
7. Deutsch M, Deutsch A, Shirihai O, Hurevich I, Afrimzon E, Shafran Y, Zurgil N. A novel miniature cell retainer for correlative high-content analysis of individual untethered non-adherent cells. Lab Chip 2006;6:995–1000.
8. Biran I, Walt DR. Optical imaging fiber-based single live cell arrays: A high-density cell assay platform. Anal Chem 2002;74:3046–3054.
9. Szmuness W, Stevens CE, Oleszko WR, Goodman A. Passive-active immunisation against hepatitis B: Immunogenicity studies in adult Americans. Lancet 1981;1(8220, Part 1):575–577.
10. Madou MJ. Fundamentals of Microfabrication: The Science of Miniaturization. Boca Raton, FL: CRC; 2002.
11. Atalla MM. Semiconductor surfaces and films; the Si-SiO_2 system. In: Gatos HC, editor. Properties of Elemental and Compound Semiconductors, Vol. 5. New York: Interscience; 1960. pp 163–181.
12. Laermer F, Schilp A. Method of anisotopically etching silicon patent. U.S. Patent 5,501,893; 1996.
13. Coburn JW, Winters HF. Plasma etching—A discussion of mechanisms. J Vac Sci Technol 1979;16:391.
14. Goodnow CC, Crosbie J, Adelstein S, Lavoie TB, Smith-Gill SJ, Brink RA, Pritchard-Briscoe H, Wotherspoon JS, Loblay RH, Raphael K, Trent RJ, Basten A. Altered immunoglobulin expression and functional silencing of self-reactive B lymphocytes in transgenic mice. Nature 1988;334:676–682.
15. Tateno C, Yoshizane Y, Saito N, Kataoka M, Utoh R, Yamasaki C, Tachibana A, Soeno Y, Asahina K, Hino H, Asahara T, Yokoi T, Furukawa T, Yoshizato K. Near completely humanized liver in mice shows human-type metabolic responses to drugs. Am J Pathol 2004;165:901–912.
16. Tsuge M, Hiraga N, Takaishi H, Noguchi C, Oga H, Imamura M, Takahashi S, Iwao E, Fujimoto Y, Ochi H, Chayama K, Tateno C, Yoshizato K. Infection of human hepatocyte chimeric mouse with genetically engineered hepatitis B virus. Hepatology 2005;42:1046–1054.
17. Mercer DF, Schiller DE, Elliott JF, Douglas DN, Hao C, Rinfret A, Addison WR, Fischer KP, Churchill TA, Lakey JR, Tyrrell DL, Kneteman NM. Hepatitis C virus replication in mice with chimeric human livers. Nat Med 2001;7:927–933.
18. Cooke MP, Heath AW, Shokat KM, Zeng Y, Finkelman FD, Linsley PS, Howard M, Goodnow CC. Immunoglobulin signal transduction guides the specificity of B cell-T cell interactions and is blocked in tolerant self-reactive B cells. J Exp Med 1994;179:425–438.
19. Gee KR, Brown KA, Chen WN, Bishop-Stewart J, Gray D, Johnson I. Chemical and physiological characterization of fluo-4 Ca(2+)-indicator dyes. Cell Calcium 2000;27:97–106.
20. Bancroft WH, Mundon FK, Russell PK. Detection of additional antigenic determinants of hepatitis B antigen. J Immunol 1972;109:842–848.
21. Bonn D. Biocomplexity: Look at the whole, not the parts. Lancet 2001;357:288.
22. Freeman WJ, Kozma R, Werbos PJ. Biocomplexity: Adaptive behavior in complex stochastic dynamical systems. Biosystems 2001;59:109–123.
23. Zurgil N, Shafran Y, Afrimzon E, Fixler D, Shainberg A, Deutsch M. Concomitant real-time monitoring of intracellular reactive oxygen species and mitochondrial membrane potential in individual living promonocytic cells. J Immunol Methods 2006;316:27–41.
24. Reichert JM, Rosensweig CJ, Faden LB, Dewitz MC. Monoclonal antibody successes in the clinic. Nat Biotechnol 2005;23:1073–1078.
25. Rosen A, Persson K, Klein G. Human monoclonal antibodies to a genus-specific chlamydial antigen, produced by EBV-transformed B cells. J Immunol 1983;130:2899–2902.
26. Smith GP. Filamentous fusion phage: Novel expression vectors that display cloned antigens on the virion surface. Science 1985;228:1315–1317.
27. Sawada H, Iwasa S, Nishimura O, Kitano K. Efficient production of anti-(hepatitis B virus) antibodies and their neutralizing activity in chimpanzees. Appl Microbiol Biotechnol 1995;43:445–451.

ヒト肝細胞キメラマウス

Chimeric mice with human hepatocytes

立野知世[1)]・森川良雄[2)]・吉里勝利[1, 3)]

Key Words
ヒト肝細胞，キメラマウス，移植，ヒト化モデル動物

Abstract

私達はマウスの肝臓のほとんどをヒト肝細胞で置換させる技術を開発した。ヒト肝細胞は免疫不全でかつ肝障害を持つマウスの肝臓に生着し，増殖することができる。このヒトの肝細胞を持つキメラマウスはヒト肝臓における肝細胞の性質を保持していたことから，医薬候補品のヒトにおける薬物代謝，薬効，および毒性を予測するための新たなツールとして期待される。

■ はじめに

医薬品の開発には，マウス，ラット，イヌ，およびサルなどの多くの動物が使われているが，薬効や毒性には種差があることが知られており，動物実験結果がヒトの臨床試験結果を必ずしも反映していない。そこで，私達はマウスの肝臓のほとんどがヒト肝細胞で置換されたヒト肝細胞キメラマウス（以下，キメラマウス）を開発した。

■1. uPA/SCIDマウスの作製と性質

ウロキナーゼプラスミノーゲンアクチベータトランスジェニック（uPA）マウスとSCIDマウスを掛け合わせてuPA/SCIDマウスを作製した[1)]。uPA遺伝子にはアルブミンエンハンサープロモーターが接続されてあるため，肝細胞においてのみuPAが高発現し血液中に分泌される。uPA/SCIDマウスの肝細胞は，uPAの発現により萎縮しており増殖することができないが，肝臓におけるHGFの活性が高いことが知られている。また，UPA/SCIDマウスは重度免疫不全の性質を持つため，異種であるヒト肝細胞はこのマウスの肝臓に生着することができる。現在，このuPA/SCIDマウスにSCIDマウスを繰り返し戻し交配することにより，このマウスの背景遺伝子を均一化することに努めている。

■2. ヒト肝細胞キメラマウスの作製

生後8日目のマウスの尾からDNAを抽出し，遺伝子検査によりuPAとSCIDの遺伝子をホモ接合型で持つマウスを選択した。生後3週目のホモ接合型uPA/SCIDマウスの脇腹を約5 mm切開し脾臓を引き出し，27Gの注射針を用いて約1 x 10^6個のヒト肝細胞を移植した（図A）[1)]。通常，移植には米国から輸入した子供の凍結保存肝細胞を融解して用いている。ドナーによってマウス肝臓への生着や増殖のしやすさが異なるため，キメラマウス作製に適したドナー細胞を選択する必要がある。また，大人の肝細胞より子供の肝細胞の方が高置換率のキメラマウスを高頻度に得ることができることがわかってきた。通常ヒト肝細胞の凍結保存チューブから約1 x 10^7個の生存肝細胞を得ることができるため，凍結チューブ1本から約10匹のキメラマウスを作製することができる。

■3. ヒト肝細胞キメラマウスの性質

子供のドナー肝細胞をuPA/SCIDマウスに移植すると，マウス血中のヒトアルブミン濃度は対数的に増加し，移植後60日頃にはプラトーに達する（図B）。マウスの肝臓切片を作製し，ヒト特異的サイトケラチン8/18 (hCK8/18) 抗体で染色すると，ヒト肝細胞のみ染め分けることができる（図C）。肝臓切片あたりのhCK8/18陽性領域の面積の割合

[1)] Chise Tateno, [2)] Yoshio Morikawa,
[1, 3)] Katsutoshi Yoshizato
[1)] 広島県産業科学技術研究所　知的クラスター創成事業　吉里プロジェクト
[2)] （株）フェニックスバイオ
[3)] 広島大学大学院理学研究科生物科学専攻

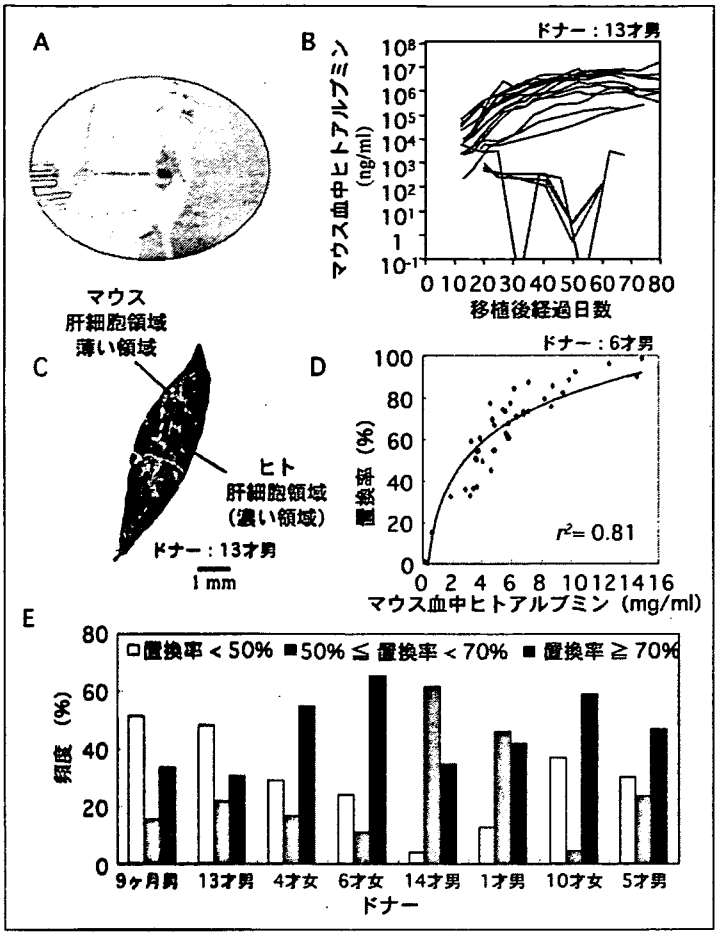

図 キメラマウスの作製と性状。A. uPA/SCIDマウスへのヒト肝細胞の移植。B. キメラマウス血中ヒトアルブミン濃度の推移。C. キメラマウス肝臓切片のhCK8/18染色像。D. キメラマウス血中ヒトアルブミン濃度と置換率の相関。E. 各ドナーで作製したキメラマウス置換率の割合。

を求め，キメラマウスの置換率とした。マウス血中ヒトアルブミン濃度と置換率には相関がある。ドナー肝細胞によって多少異なるが，マウス血中ヒトアルブミン濃度が6 mg/ml以上のキメラマウスの置換率はほぼ70%以上となる（図D）[1]。また，子供の肝細胞を移植した場合，移植したマウスのうち約半分のマウスが置換率70%以上となる（図E）。

現在では，キメラマウス作製技術は（株）フェニックスバイオに移転されている。（株）フェニックスバイオでは，現在置換率70%のキメラマウスを年間1000匹以上生産することができ，さらに増産を計画している。これらのマウスは医薬品開発のための受託試験に利用されている。

キメラマウスの肝臓における薬物代謝や解毒に関わる酵素であるシトクロームP450，第2相酵素，そしてトランスポーター遺伝子やタンパク質は，ヒトの肝臓と近いレベルで発現していることが示されている[1~3]。ヒトB型，C型肝炎ウィルスはヒトやチンパンジーの*in vivo*の肝臓にしか感染しないため，これまで，ウィルスの感染メカニズムや抗ウィルス剤を開発するための実験系が存在しなかった。キメラマウスは，ヒトB型，C型肝炎ウィルスを感染させることが可能であることも示されている[4]。さらに，キメラマウスは，肝臓をターゲットとした遺伝子治療ベクターの有効性や安全性を確かめる実験モデルとしても有用であることが示された[5]。

■ おわりに

これまで，*in vivo*のヒト肝臓における遺伝子およびタンパク質発現，酵素活性に対する医薬候補品などの影響を調べることは不可能であった。また，ヒトにおける医薬候補品の薬物代謝，薬効，および肝毒性を動物実験により予測してきたが，充分なものとはいえなかった。キメラマウスは，*in vivo*における医薬候補品などのヒトにおける薬物代謝，薬効，および肝毒性を予測することができる，画期的なモデル動物になりえると考えている。現在のところ，キメラマウスは医薬候補品のヒトにおける薬物動態を予測する系や，肝炎ウィルスに対する抗ウィルス薬のスクリーニング系として実際に利用されている。現在，キメラマウスが薬効や肝毒性を調べるためのツールとしても有効かどうか，データを積み重ねているところである。

さらに，キメラマウスを用いることによって，これまで調べることが困難であった，*in vivo*におけるヒト肝細胞の増殖，分化，老化などのメカニズムについて明らかにすることが可能となり，臨床における肝臓疾患治療への一助となることを期待している。

文　献

1) Tateno C,Yoshizane Y,Saitou N,*et al.*Am J Pathol 165:901-912, 2004
2) Katoh M, Matsui T, Okamura H, *et al.*Drug Metab Dispos 33:1333-1340, 2005
3) Nishimura M, Yoshitsugu H, Yokoi T, *et al.* Xenobiotica 35:877-890, 2005
4) Tsuge M, Hiraga N, Takaishi H, *et al.* Hepatology 42:1046-1054, 2005
5) Emoto K, Tateno C, Hino H, *et al.* Human GeneTherapy 16:1168-1174, 2005

Hindawi Publishing Corporation
EURASIP Journal on Bioinformatics and Systems Biology
Volume 2007, Article ID 47214, 11 pages
doi:10.1155/2007/47214

Research Article

Gene Systems Network Inferred from Expression Profiles in Hepatocellular Carcinogenesis by Graphical Gaussian Model

Sachiyo Aburatani,[1] Fuyan Sun,[1] Shigeru Saito,[2] Masao Honda,[3] Shu-ichi Kaneko,[3] and Katsuhisa Horimoto[1]

[1] *Biological Network Team, Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-42 Aomi, Koto-ku, Tokyo 135-0064, Japan*
[2] *Chemo & Bio Informatics Department, INFOCOM CORPORATION, Mitsui Sumitomo Insurance Surugadai Annex Building, 3-11, Kanda-Surugadai, Chiyoda-ku, Tokyo 101-0062, Japan*
[3] *Department of Gastroenterology, Graduate School of Medical Science, Kanazawa University, 13-1 Takara-machi, Kanazawa, Ishikawa 920-8641, Japan*

Received 28 June 2006; Revised 27 February 2007; Accepted 1 May 2007

Recommended by Paul Dan Cristea

Hepatocellular carcinoma (HCC) in a liver with advanced-stage chronic hepatitis C (CHC) is induced by hepatitis C virus, which chronically infects about 170 million people worldwide. To elucidate the associations between gene groups in hepatocellular carcinogenesis, we analyzed the profiles of the genes characteristically expressed in the CHC and HCC cell stages by a statistical method for inferring the network between gene systems based on the graphical Gaussian model. A systematic evaluation of the inferred network in terms of the biological knowledge revealed that the inferred network was strongly involved in the known gene-gene interactions with high significance ($P < 10^{-4}$), and that the clusters characterized by different cancer-related responses were associated with those of the gene groups related to metabolic pathways and morphological events. Although some relationships in the network remain to be interpreted, the analyses revealed a snapshot of the orchestrated expression of cancer-related groups and some pathways related with metabolisms and morphological events in hepatocellular carcinogenesis, and thus provide possible clues on the disease mechanism and insights that address the gap between molecular and clinical assessments.

1. INTRODUCTION

Hepatitis C virus (HCV) is the major etiologic agent of non-A non-B hepatitis, and chronically infects about 170 million people worldwide [1–3]. Many HCV carriers develop chronic hepatitis C (CHC), and finally are afflicted with hepatocellular carcinoma (HCC) in livers with advanced-stage CHC. Thus, the CHC and HCC cell stages are essential in hepatocellular carcinogenesis.

To elucidate the mechanism of hepatocellular carcinogenesis at a molecular level, many experiments have been performed from various approaches. In particular, recent advances in techniques to monitor simultaneously the expression levels of genes on a genomic scale have facilitated the identification of genes involved in the tumorigenesis [4]. Indeed, some relationships between the disease and the tumor-related genes were proposed from the gene expression analyses [5–7]. Apart from the relationship between tumor-related genes and the disease at the molecular level, the information about the pathogenesis and the clinical characteristics of hepatocellular carcinogenesis has accumulated steadily [8, 9]. However, there is a gap between the information about hepatocellular carcinogenesis at the molecular level and that at more macroscopic levels, such as the clinical level. Furthermore, the relationships between tumor-related genes and other genes also remain to be investigated. Thus, an approach to describe the perspective of carcinogenesis from measurements at the molecular level is desirable to bridge the gap between the information at the two different levels.

Recently, we have developed an approach to infer a regulatory network, which is based on graphical Gaussian modeling (GGM) [10, 11]. Graphical Gaussian modeling is one of the graphical models that includes the Boolean and Bayesian models [12, 13]. Among the graphical models, GGM has the simplest structure in a mathematical sense; only the inverse

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 ith gene in the jth 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 ith 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 $C(\tau)$. τ 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 χ^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 ≤ 0.05, or the probability value corresponding to dev2 ≤ 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 τ 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.

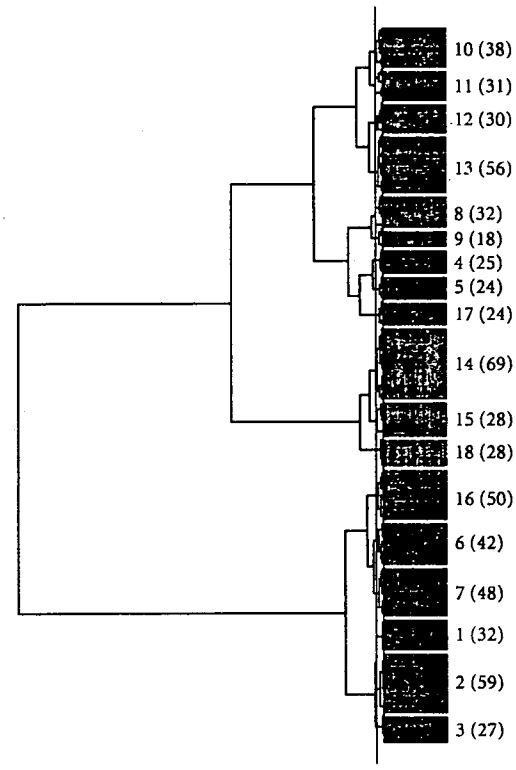

FIGURE 1: *Dendrogram of genes and profiles.* The dendrogram was constructed by hierarchical clustering with the metric of the Euclidian distances between the correlation coefficients and the UPGMA. The blue line on the dendrogram indicates the cluster boundary estimated automatically by ASIAN. The gene 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,

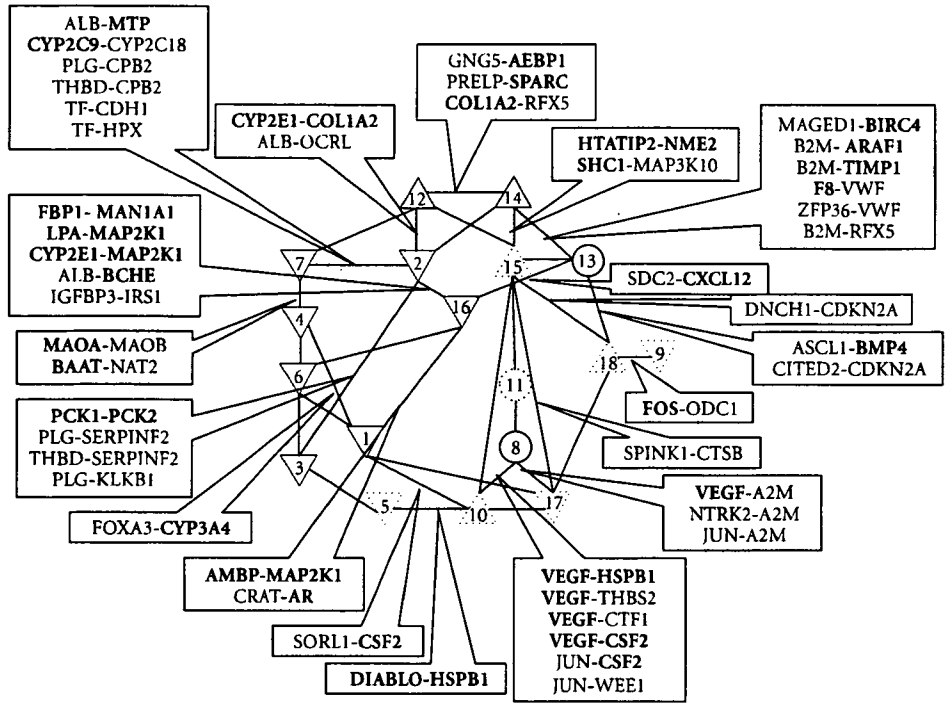

FIGURE 2: *Network between clusters, together with a projection of biological knowledge about the gene interactions.* The clusters are indicated by triangles and circles, in which the cluster numbers correspond to those in Figure 1, and the edges between the clusters are associations with 5% significance 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

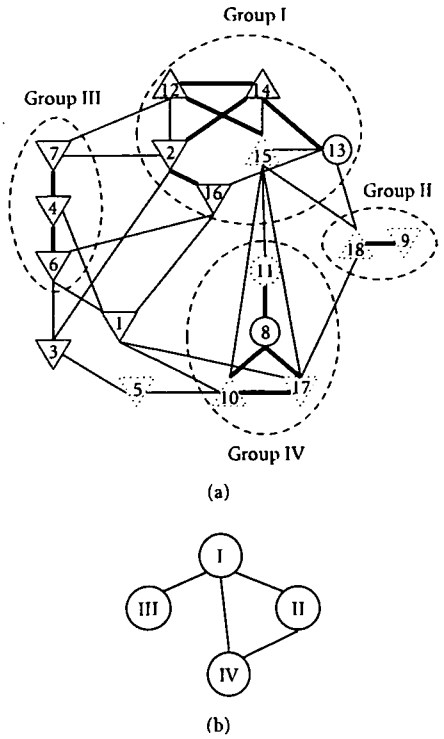

FIGURE 3: *Orchestration of gene systems.* (a) The association with 1% significance probability is indicated by a bold line, and the clusters with 1% significance association are naturally divided into four groups, which are enclosed by broken lines. (b) The connections between the groups are drawn schematically,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