Table 2. List of genes related oxidative stress and neuronal disease

Gene name	Discription	Summary of biological function
Apbb	amyloid beta (A4) precursor protein-binding	binds to the intracellular domain of the Alzheimer's disease beta-amyloid precursor protein
App	amyloid beta (A4) precursor protein	This protein may relate in absorption of Cu.
Bace1	Beta-site APP-cleaving enzyme 1	Amyloid beta peptide is generated by proteolytic cleavage of amyloid precursor protein (APP) by g-secretase and this protease. Also called b-secretase.
Bcl2	B-cell CLL/lymphoma 2	This gene encodes an integral outer mitochondrial membrane protein that blocks the apoptotic death of cells.
Ccs	Copper chaperone for superoxide dismutase	Copper chaperone for superoxide dismutase specifically delivers Cu to superoxide dismutase 1 and may activate superoxide dismutase 1 through direct insertion of the Cu cofactor.
Cdk5	Cyclin-dependent kinase 5	The product of this gene binds to APP and transfer it to b-secretase under the oxidative stress.
Gpx1	Glutathione peroxidase 1	Glutathione peroxidase functions in the detoxification of hydrogen peroxide, and is one of the most important antioxidant enzymes
Ide	Insulin-degrading enzyme	belong to a protease family responsible for intercellular peptide signalling. It degrades the intracellular domain of the amyloid precursor
Psen1	Presenilin 1	A subunit of g-secretase. It plays essential role in noramal cleavage of the amyloid precursor protein (APP).
Sod1	Superoxide dismutase 1, soluble	The protein binds copper and zinc ions and is one of two isozymes responsible for destroying free superoxide radicals in the cells. Mutations in this gene have been implicated as causes of familial amyotrophic lateral sclerosis.
Ubb	Ubiquitin B	Polyubiquitin precursor with a final amino acid after the last repeat. Aberrant form of this protein has been noticed in patients with Alzheimer's and Down syndrome.
Ube2l3	Ubiquitin-conjugating enzyme E2L 3	A member of the E2 ubiquitin-conjugating enzyme family.
Ubqln1	Ubiquilin 1	This ubiquilin has been shown to modulate accumulation of presenilin proteins, and it is found in lesions associated with Alzheimer's and
Uchl1	Ubiquitin carboxyl-terminal esterase L1 (ubiquitin thiolesterase)	UCHL1 is a member of a gene family whose products hydrolyze small C-terminal adducts of ubiquitin to generate the ubiquitin monomer. It is present in all neurons.

Figure 1. Differential gene expression of copper accumulation-response genes with a genome informatics approach

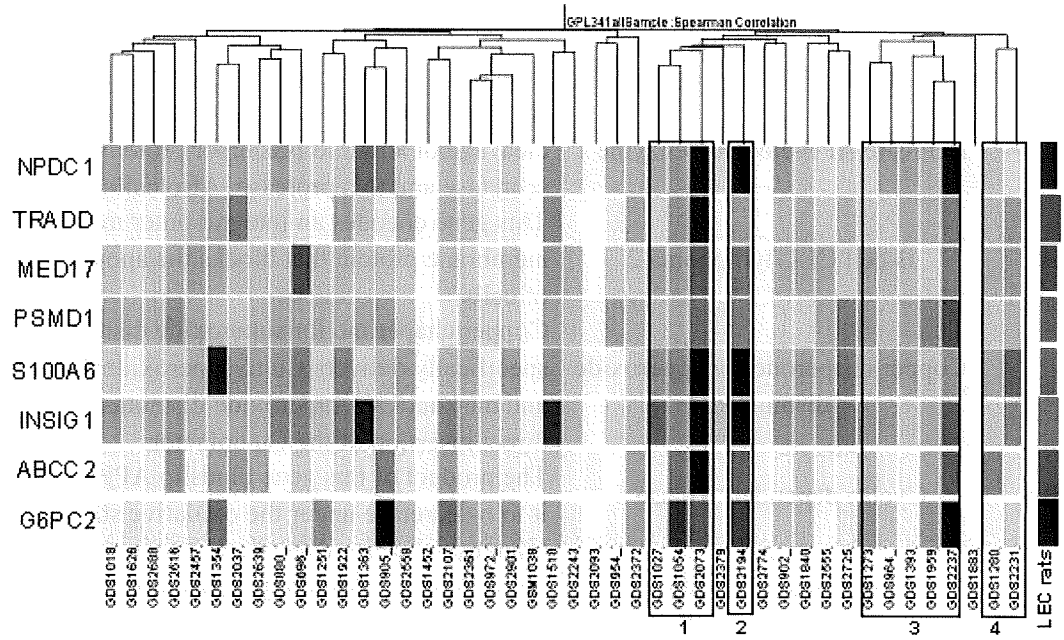


Figure 2. Prediction of gene interaction network for oxidative stress regulation in the heavy metal exposure in HepG2 cells

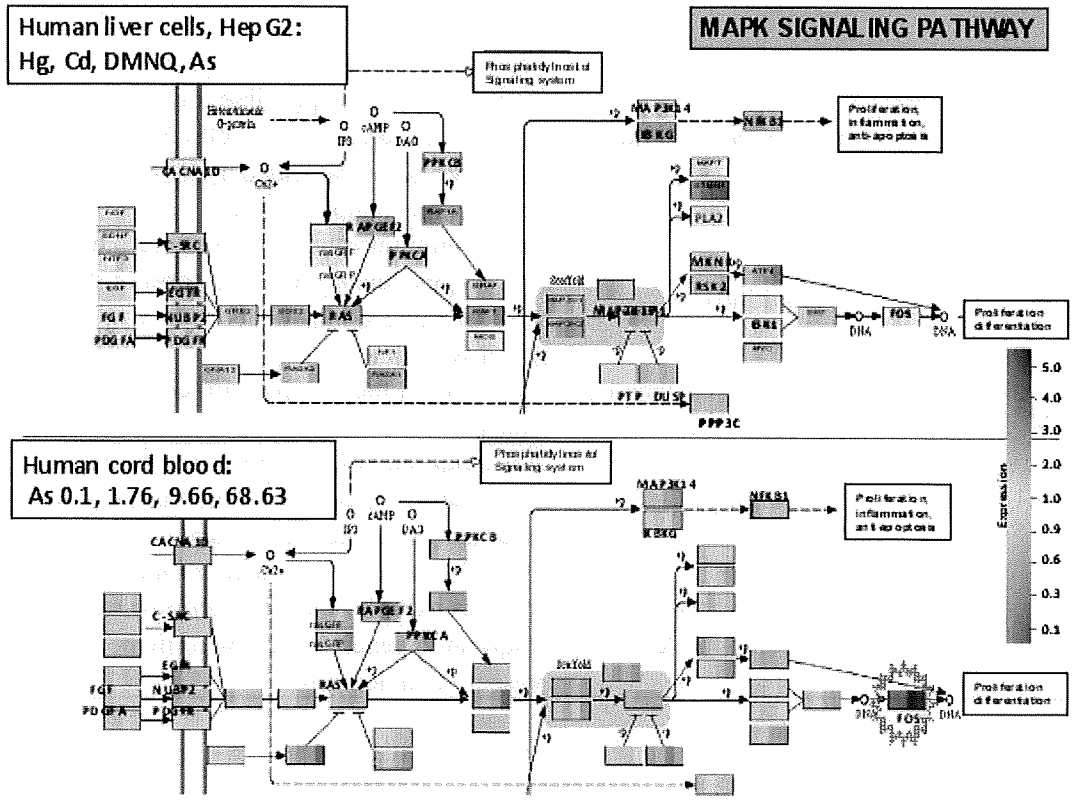


Figure 3. Genetic networks of oxidative stress 11 genes in data selected from the human cord blood study

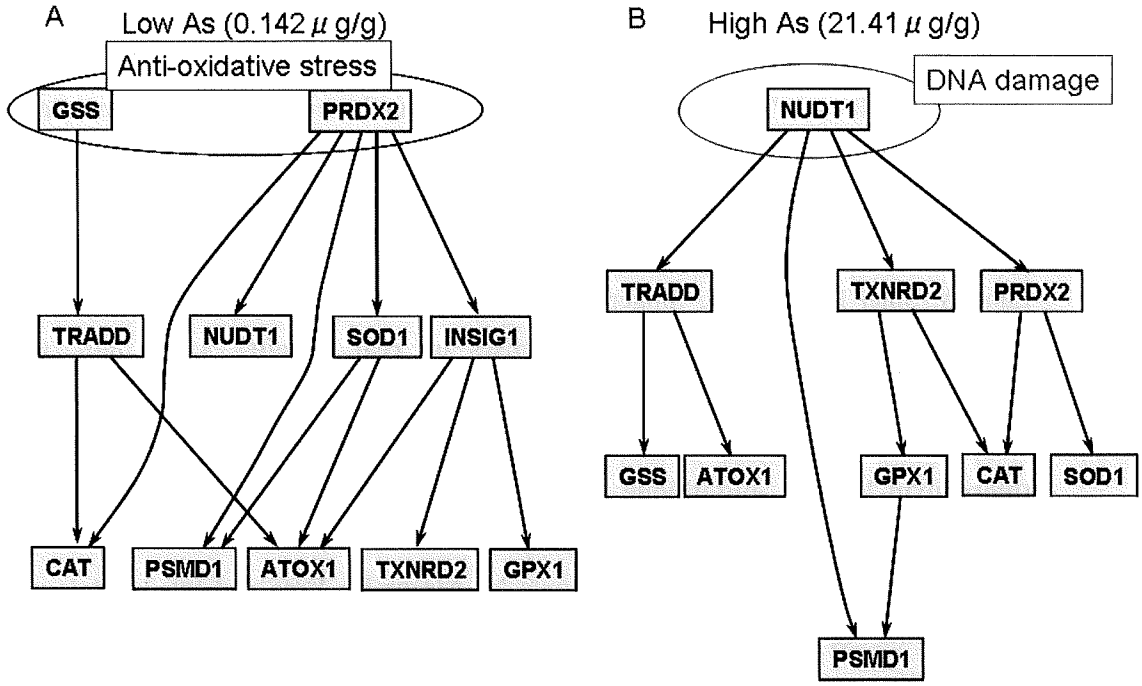
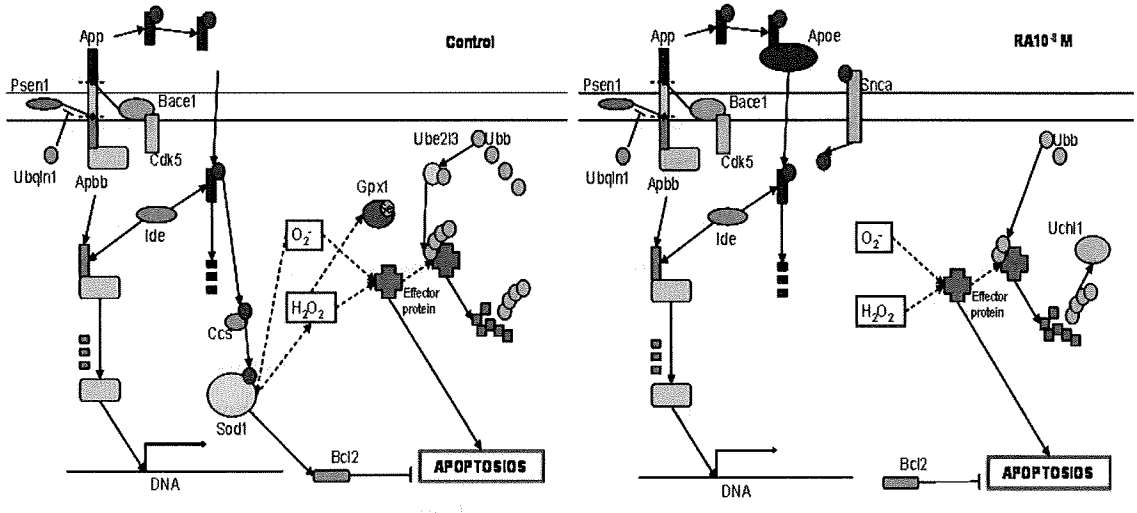


Figure 4. Gene Expression Signature influenced with RA exposure during development of ES to neuronal line ages



Refferences

1. Gibb S: **Toxicity testing in the 21st century: a vision and a strategy**. *Reprod Toxicol* 2008, **25**(1):136-138.
2. Woods CG, Heuvel JP, Rusyn I: **Genomic profiling in nuclear receptor-mediated toxicity**. *Toxicol Pathol* 2007, **35**(4):474-494.
3. Mendrick DL: **Genomic and genetic biomarkers of toxicity**. *Toxicology* 2008, **245**(3):175-181.
4. Luhe A, Suter L, Ruepp S, Singer T, Weiser T, Albertini S: **Toxicogenomics in the pharmaceutical industry: hollow promises or real benefit?** *Mutat Res* 2005, **575**(1-2):102-115.
5. Hansen JM, Zhang H, Jones DP: **Differential oxidation of thioredoxin-1, thioredoxin-2, and glutathione by metal ions**. *Free Radic Biol Med* 2006, **40**(1):138-145.
6. Valko M, Rhodes CJ, Moncol J, Izakovic M, Mazur M: **Free radicals, metals and antioxidants in oxidative stress-induced cancer**. *Chem Biol Interact* 2006, **160**(1):1-40.
7. Bau DT, Wang TS, Chung CH, Wang AS, Wang AS, Jan KY: **Oxidative DNA adducts and DNA-protein cross-links are the major DNA lesions induced by arsenite**. *Environ Health Perspect* 2002, **110 Suppl 5**:753-756.
8. Kawai Y, Furuhata A, Toyokuni S, Aratani Y, Uchida K: **Formation of acrolein-derived 2'-deoxyadenosine adduct in an iron-induced carcinogenesis model**. *J Biol Chem* 2003, **278**(50):50346-50354.
9. Chiu HJ, Fischman DA, Hammerling U: **Vitamin A depletion causes oxidative stress, mitochondrial dysfunction, and PARP-1-dependent energy deprivation**. *Faseb J* 2008, **22**(11):3878-3887.
10. Knott L, Hartridge T, Brown NL, Mansell JP, Sandy JR: **Homocysteine oxidation and apoptosis: a potential cause of cleft palate**. *In Vitro Cell Dev Biol Anim* 2003, **39**(1-2):98-105.
11. Nebert DW, Petersen DD, Fornace AJ, Jr.: **Cellular responses to oxidative stress: the [Ah] gene battery as a paradigm**. *Environ Health Perspect* 1990, **88**:13-25.
12. Cheng Y, Chang LW, Cheng LC, Tsai MH, Lin P: **4-Methoxyestradiol-induced oxidative injuries in human lung epithelial cells**. *Toxicol Appl Pharmacol* 2007, **220**(3):271-277.
13. Portier CJ, Toyoshiba H, Sone H, Parham F, Irwin RD, Boorman GA: **Comparative analysis of gene networks at multiple doses and time points in livers of rats**

exposed to acetaminophen. *Altex* 2006, **23 Suppl**:380-384.

14. Toyoshiba H, Sone H, Yamanaka T, Parham FM, Irwin RD, Boorman GA, Portier CJ: **Gene interaction network analysis suggests differences between high and low doses of acetaminophen**. *Toxicol Appl Pharmacol* 2006, **215**(3):306-316.
15. Toyoshiba H, Yamanaka T, Sone H, Parham FM, Walker NJ, Martinez J, Portier CJ: **Gene interaction network suggests dioxin induces a significant linkage between aryl hydrocarbon receptor and retinoic acid receptor beta**. *Environ Health Perspect* 2004, **112**(12):1217-1224.
16. Yamanaka T, Toyoshiba H, Sone H, Parham FM, Portier CJ: **The TAO-Gen algorithm for identifying gene interaction networks with application to SOS repair in E. coli**. *Environ Health Perspect* 2004, **112**(16):1614-1621.
17. Nair J, Sone H, Nagao M, Barbin A, Bartsch H: **Copper-dependent formation of miscoding etheno-DNA adducts in the liver of Long Evans cinnamon (LEC) rats developing hereditary hepatitis and hepatocellular carcinoma**. *Cancer Res* 1996, **56**(6):1267-1271.
18. Sone H, Li YJ, Ishizuka M, Aoki Y, Nagao M: **Increased mutant frequency and altered mutation spectrum of the lacI transgene in Wilson disease rats with hepatitis**. *Cancer Res* 2000, **60**(18):5080-5086.
19. Sone H, Wakabayashi K, Kushida H, Enomoto K, Mori M, Takeichi N, Tsuda H, Sugimura T, Nagao M: **Hepatocellular carcinoma induction in LEC rats by a low dose of 2-amino-3,8-dimethylimidazo[4,5-f]quinoxaline**. *Jpn J Cancer Res* 1996, **87**(1):25-29.
20. Jia G, Takahashi R, Zhang Z, Tsuji Y, Sone H: **Aldo-keto reductase 1 family B7 is the gene induced in response to oxidative stress in the livers of Long-Evans Cinnamon rats**. *Int J Oncol* 2006, **29**(4):829-838.
21. Jia G, Tohyama C, Sone H: **DNA damage triggers imbalance of proliferation and apoptosis during development of preneoplastic foci in the liver of Long-Evans Cinnamon rats**. *Int J Oncol* 2002, **21**(4):755-761.

Chapter 5

High-Performance Gene Expression Module Analysis Tool and Its Application to Chemical Toxicity Data

Wataru Fujibuchi, Hyeryung Kim, Yoshifumi Okada, Takeaki Taniguchi, and Hideko Sone

Summary

Gene clustering is one of the main themes of data mining approaches in bioinformatics. Although it has the power to analyze gene function, interpretation of the results becomes increasingly difficult when the number of experiments (samples) exceeds hundreds or more. A new type of clustering called "biclustering," where genes and experiments are coclustered in a large-scale of gene expression data, has been extensively studied in the last decade. We have developed "SAMURAI," an original program that detects all the biclusters or "gene modules" whose genes have similar expression patterns to query profile using the ultrafast data mining algorithm called Linear-time Closed itemset Miner (LCM). Using chemical toxicity dataset from J&J rat liver experiments, we compiled an exhaustive dictionary of gene modules by searching datasets of gene modules with each chemical exposure experiment as query. Through the module analysis, we found that our program can detect up/down-regulated gene sets that significantly represent particular GO functions or KEGG pathways, thereby unraveling reactions and mechanisms common to different toxicochemical treatments of hepatocytes.

Key words: Gene expression module, Biclustering, Chemical toxicity, Data mining, Linear time common itemset miner, Common reaction and mechanism

1. Introduction

Microarrays or other high-throughput gene expression analysis systems provide extensive information on gene expression differences under various experimental conditions, such as cell type, developmental stage, and reaction to stimulus. Recent easy access to such experimental techniques has promoted the accumulation of gene expression data in public gene expression data

Hisashi Koga (ed.), *Reverse Chemical Genetics*, Methods in Molecular Biology, vol. 577
DOI 10.1007/978-1-60761-232-2_5,

repositories, such as GEO and ArrayExpress *(1, 2)*. Among available tools to analyze such large-scale data, a promising method called "biclustering" *(3)* has emerged and has been widely studied *(4–9)* for its ability to mine datasets containing hundreds of experiments. Biclustering detects common gene expression patterns or "gene expression motifs" that are represented in any combination of experiments. A subset of genes that contain a common expression pattern in a subset of experiments is called a "gene module."

We have developed a high-performance biclustering method that has high calculation speed and high biological evaluation accuracy. Our biclustering system called "SAMURAI (System for Assembling Modules by Ultra Rapid Algorithm on Itemsets)" *(10)* can search thousands of microarray data for gene modules in several seconds in most cases. In addition, the detected gene modules show surprisingly high accuracy of matching to known gene function groups in the study of 2,988 disease microarray data provided by the Critical Assessment of Microarray Data Analysis meeting *(11)*. Here we applied this system to a chemical toxicity dataset obtained from J&J rat liver experiments and compiled an exhaustive dictionary of gene modules by searching the dataset with each chemical exposure experiment as query. From the resultant gene modules, we found that there are a total of 92,100 modules of which 10,805 (11.7%) represent known functions (GO or KEGG) at a high significance threshold ($p < 1e-5.5$ and $p < 1e-4$, respectively).

2. Materials and Methods

To obtain and analyze gene modules from large-scale gene expression data, we first normalize and convert them into a unified file format. After obtaining formatted data, we discretize them to certain degrees of expression to find common expression patterns in a limited search space. Then, to reduce search space effectively and to detect gene modules in real time, we perform "query-and-database" search. In this approach, given query gene expression data by a user, all of the discretized values not common to the query in the database are erased, thus reducing database size extensively. This process is critical to the calculation of the following module detection process, giving exhaustive (not partial) results.

After fully reducing search space by discretizing and erasing data unrelated to query, we perform rapid data mining where common gene expression patterns that are preserved in all combinations of (maximum) experiments are exhaustively retrieved.

However, the gene modules retrieved in this step have rigid patterns with no relaxation (i.e., noise), which contrasts to real data containing noise and biological flexibility. Thus, in the next step, these "core modules" are compared with each other and merged into bigger modules containing noise. As the final output, we obtain each module consisting of a subset of genes and a subset of experiments. By scrutinizing both gene functions and experimental relationships in each module, we can formulate new and interesting hypotheses on gene functions and experimental groups.

2.1. Rat Chemical Exposure Dataset and Normalization

1. Download gene expression data of chemical effects on rat liver by J & J from the Web site: http://cebs.niehs.nih.gov/. To retrieve the data, select the subject "J&J Hepatotoxicant Library" in the "Display All Studies" page or limit data by the organization name "Johnson and Johnson" in the "Study Characteristics" page. There are 133 toxicochemical groups containing 964 microarray experiments. Go to "View Selected Microarray Data by Studies/Experiments" from the bottom of page, select the dataset, and go to "View Details about Selected Experiment(s)." Then, download microarray data files after entering "Click to Download" page. As the file size is 117.3 MB, downloading will take 5 min to hours depending on the user's network conditions.
2. Unzip downloaded files and select files to analyze. In this study, we select only experiments that have both chemical exposure done and control data collected on the same day. The number of such experiments is 298. Among 9,215 probes in the dataset, we delete low-abundance genes that show no expression in all of the 298 experiments and select only 7,614 probes. Every pair of gene expression values is transformed into log-fold-change abundance by subtracting control values from the chemical exposure after taking $\log_2$ and subtracting the median value in each array (*see* **Note 1**). Finally, the log-fold-change values are normalized to Z-score by $(x - \text{mean})/\text{SD}$.
3. To perform gene functional analysis in later process, check if each probe has a link to UniGene database. Only 5,832 probes have links to UniGene database. Then, to remove probe redundancies, take the average if multiple probes correspond to the same UniGene ID. As a result, we obtain 2,497 averaged probes that have links to UniGene database for 298 chemical exposure experiments.

2.2. Formatting Database by Discretization

1. Convert normalized dataset into rank-ordered discrete data within each experiment. To do this, select one experiment and sort genes by expression value. Then, put all the genes into 10,000 bins of the same size and assign every gene a rank value equal to the number of bins (1st–10,000th from low to high) that it belongs to.

2. Select one gene and make a distribution of rank values for all the 298 experiments. Then, set a discretization parameter and thresholds. In this study, we set the degree of discretization at 3 (±1, 0) and the thresholds at 3% or 5% from each side of top and bottom (i.e., 6% or 10% in total) in rank value distribution. Using these parameters, we assign discrete values to rank values.

2.3. Query Data and Database Compression

1. Given query gene expression data, discretize gene expression values using the same procedure as that employed in the above database formatting process. Use the same degree and thresholds to discretize query data as that used in the database discretization.
2. Compare discretized query to discretized database at the gene level. Then, delete all discrete values that differ from the query in the database. In addition, delete all zero values in the database. This procedure compresses the database to an extremely small size (*see* **Note 2**).

2.4. LCM for Data Mining of Core Gene Modules

1. LCM *(12, 13)* is an ultrafast algorithm for data mining that has been used to retrieve maximum common itemsets from a large list of itemsets called a transaction database (*see* **Note 3**). To apply this algorithm, assign item names to all the existing combinations of gene names and discrete values in the gene expression database. For example, we give item names "a – 1" and "a – 2" to the case that gene "a" has values of both "–1" and "–2" in the database. This procedure converts each gene expression experiment datum into an itemset list that symbolizes gene expression status consisting of gene names and their discrete values.
2. Write itemset list to a file in the format of one experiment in one line. Then, download the LCM program from the Web site: http://research.nii.ac.jp/~uno/codes.htm. Run the LCM program with the itemset list file with parameters of minimum size of gene modules to extract: m genes × n experiments. For example, the input command to run LCM on a Linux machine is: % lcm CqI –l m [input_file] n [output_file] (*see* **Note 3**).

2.5. Merging Redundant Modules

1. Raw gene modules (core modules) extracted by LCM are expected to be highly redundant. To reduce almost the same or quite similar gene modules in output data, merge them if they meet conditions specified by the following procedure. First, sort gene modules by size. Then, select the largest module and merge it with other modules one by one from large to small ones. Suppose we are merging modules A and B. If A and B share genes g1 and g2 and experiments e1 and e2 but A has another gene g3 and B has another experiment e3, the merged module will have genes g1, g2, and g3 and experiments e1, e2,

and e3 by adding missing gene (g3) and experiment (e3) to B and A, respectively. However, if any gene or experiment of the merged module contains inconsistent values, such as missing or different discrete values, the percentages of inconsistencies in each line and row of the merged module should be checked. If the inconsistency in every line and row is less than the threshold (in this study, 0.4 and 0.5 are adopted), execute the merge and replace the larger module with it. (Delete the smaller one.) Repeat this "check and merge" process for this larger module until it reaches the smallest one.

2. Repeat the above "check and merge" process from the next largest module to the smallest module. Once the smallest module is reached, sort the modules by size again and perform "check and merge" from the largest module to the smallest one for a new list of modules.
3. Repeat the above "sort, check, and merge" until no merge happens. Then, output the final (merged) modules to a file (*see* **Note 4**). The whole process from formatting data to merging modules is illustrated in **Fig. 1**. Here an example of two degrees (high and low expressions, or +1 and −1) of discretization is shown.

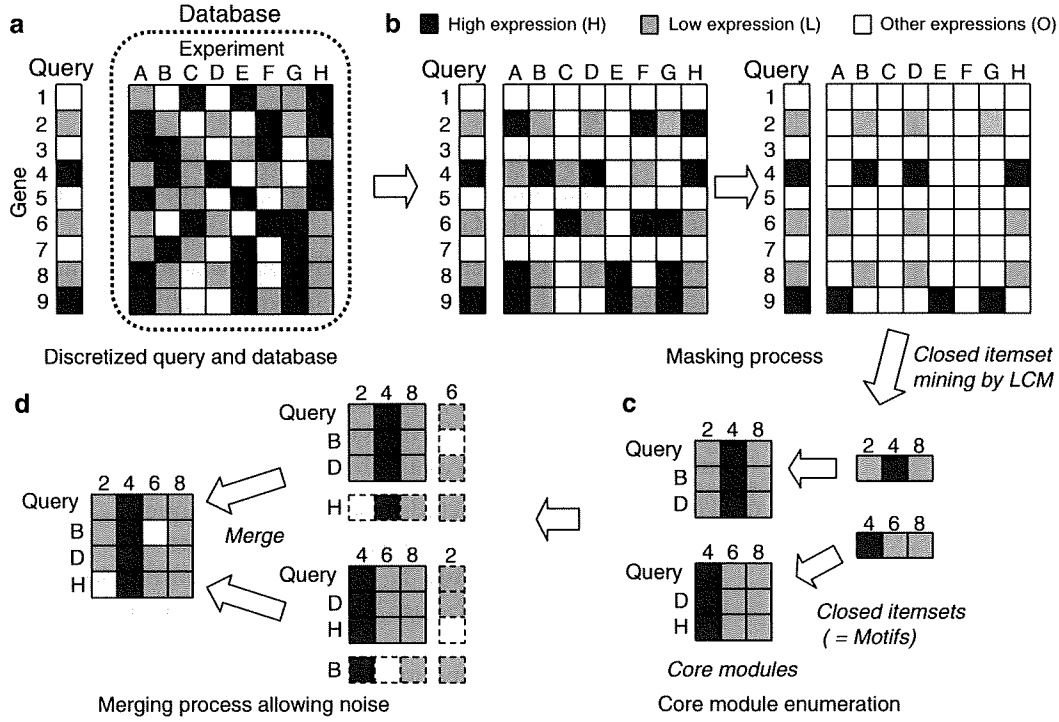


Fig. 1. Whole scheme of gene module search from gene expression database. The system consists of four steps (**a**) discretization of query and database, (**b**) database masking, (**c**) enumeration of core modules, and (**d**) module merging. In this example, the query and the database are discretized into only three degrees "High (+1)," "Low (−1)," and "Other expressions (0)".

2.6. Evaluation by GO and KEGG

Once the final set of modules is obtained, it is necessary to check the biological validities of those modules to verify if the selected parameters (discretization degree, noise threshold, module size, etc.) work properly. To approach this, compare each gene module with known biological functions or pathways to investigate if genes in a single module are statistically "enriched" for a particular category of functions. Here we describe the method of performing categorical enrichment analysis based on Gene Ontology (GO) functions and KEGG biological pathways.

1. Download necessary files from two FTP sites (a) gene2unigene and gene2go.gz files from ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/ for gene name conversion and GO term information, respectively; and (b) gene_map.tab for KEGG pathway map index information and rno_gene_map.tab files for rat specific pathway gene list from ftp://ftp.genome.ad.jp/pub/kegg/pathway/.
2. Take one gene expression module. Convert UniGene IDs into EntrezGene IDs via the cross-reference list in gene2unigene file (*see* **Note 5**).
3. Assign GO function terms to the converted (Entrez-) genes. Obtain four parameters (a) U, the number of EntrezGenes found in the input gene expression data (2,497 UniGenes); (b) M, the number of EntrezGenes assigned to each GO term; (c) k, the number of EntrezGenes in each gene module; and (d) m, the number of EntrezGenes in each GO term found in each module.
4. Assign KEGG pathway map names to the converted (Entrez-) genes. Obtain the four parameters for each pathway map in the same way as the GO terms.
5. Evaluate each GO term and KEGG pathway map names for each gene module with standard hypergeometric distribution statistics by giving m as positives out of k samples and M known positives among the total population of U. To critically assess p value threshold, shuffle gene names (UniGene IDs) in the input dataset and do the same statistical analysis for each module, and use them as a null-distribution model.
6. The numbers of extracted modules for 298 query chemicals under two different discretization thresholds (3% and 5%) are summarized in **Table 1**. Two different noise ratios in the module merging process are also tested in both data.

2.7. Analysis of Common Reactions Among Chemicals by Gene Modules

The main objective in gene module analysis of reverse chemical genomics is to find new functional relationships between different chemicals. Once a set of gene modules annotated by GO and KEGG is obtained, check the common function names for every combination of query chemicals with a significant p value threshold.

Table 1
Numbers of obtained modules under various parameter conditions

	Noise ratio	
Discretization	**0.4**	**0.5**
3%	2,859	2,088
5%	92,100	61,852

1. First, assign *p* values to each gene module by GO and KEGG categorical enrichment analysis as described in 2.6. Choose the most significant (the smallest) *p* value among various functional candidates for a single gene module. Then, plot each module by its (−log) *p* value, as shown in **Fig. 2a**.
2. Plot the same module in which gene IDs are shuffled by its *p* value, as shown in **Fig. 2b**. Compare the two plots. Set the *p* value threshold at the critical point where raw modules are still observed but gene-shuffled modules disappear. Here, we arbitrarily choose 1e − 5.5 and 1e − 4 as the thresholds for GO and KEGG, respectively, for the modules extracted by the parameters of 5% discretization and 0.4 of noise ratio.
3. Take every pair of query chemicals and check if they share common function or pathway names at the above threshold. Store the results.
4. Find biological hypotheses that can explain the obtained relationships among two or more chemicals. For example, in our data, some modules obtained from a query of Benzbromarone are found to affect the pathway of "Biosynthesis of unsaturated fatty acids." This pathway was also found with queries of Fenbufen, Clofibrate, and Dichloroacetate. Three genes are involved in the obtained modules: acyl-CoA thioesterase 3 (Rn.11326), acyl-CoA thioesterase 7 (Rn.6024), and acyl-Coenzyme A oxidase 1, palmitoyl (Rn.31796), all of which are important in unsaturated fatty acid metabolism. This result and information from literature suggest that these chemicals could affect the rat liver in a similar manner that is related to carcinogenesis via PPARα-derived oxidative reaction. **Figure 3b** is an example of the graphical output of these modules produced by the web-based GUI version of the SAMURAI system.

2.8. Usage of SAMURAI System

To enhance the above analysis in a coordinated way, we have developed a gene module extraction and evaluation system called SAMURAI. The free trial version of the program coded in java/C++/Perl is available from http://samurai.cbrc.jp/download/SAMURAI-Progressive/ (*see* **Note 6**). Here we describe briefly the usage of the system.

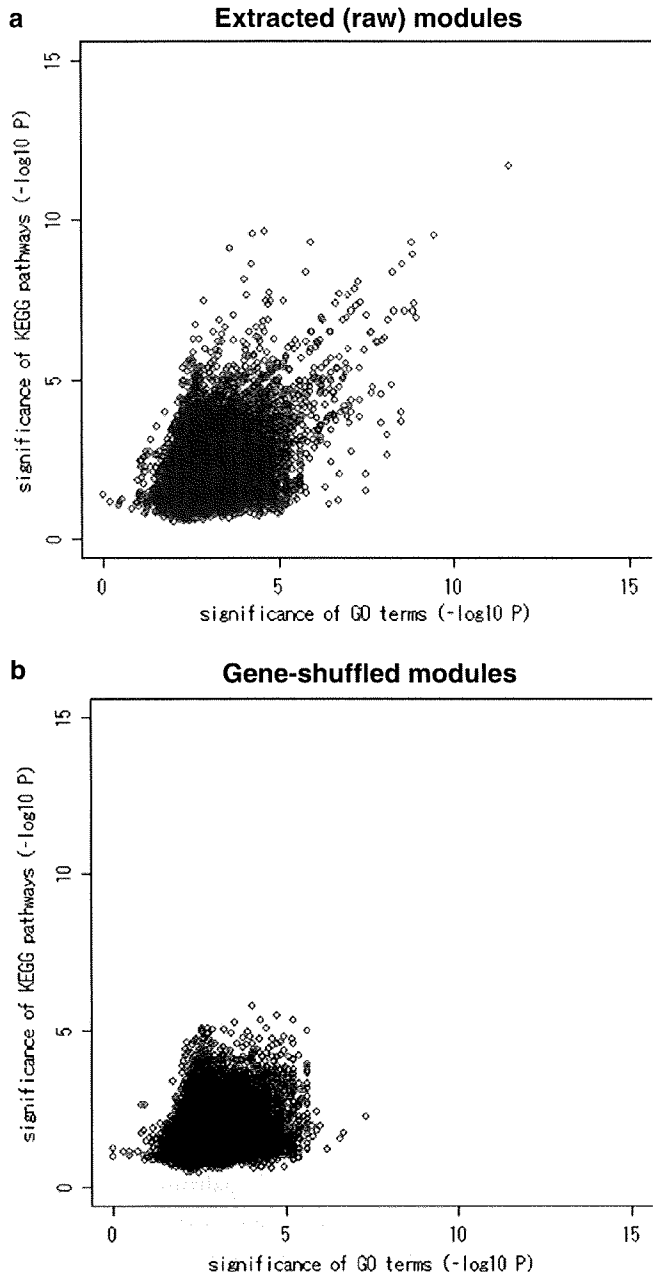


Fig. 2. Analysis of significant gene modules by GO and KEGG function groups. Each gene module (with 5% discretization threshold and 0.4 noise ratio) is evaluated with GO terms and KEGG pathway maps based on hypergeometric distribution statistics and plotted by the most significant *p* values. A total of 92,100 obtained gene modules are plotted in (**a**), and the same gene modules in which their gene ids are shuffled are plotted in (**b**). *p* values are dramatically increased due to randomness in (**b**). There is a significant correlation between *p* values of GO and KEGG in only (**a**). We arbitrarily choose 1e – 5.5 and 1e – 4 as the threshold for GO and KEGG, respectively.

1. Download SAMURAI program from the above site. Select SAMURAI-P program. Uncompress and untar the frozen file on your Linux machine. Go to the expanded directory and type "make compile" to compile all the programs in the system.

2. Transform your gene expression dataset into the "CellMontage" format *(14)* where a gene expression profile consists of a single description line, starting with ">," followed by one or more data lines. The data consist of elements separated by a space or a new line. Each element consists of a UniGene identifier and an expression value separated by a colon. See **Fig. 3a** for example data.

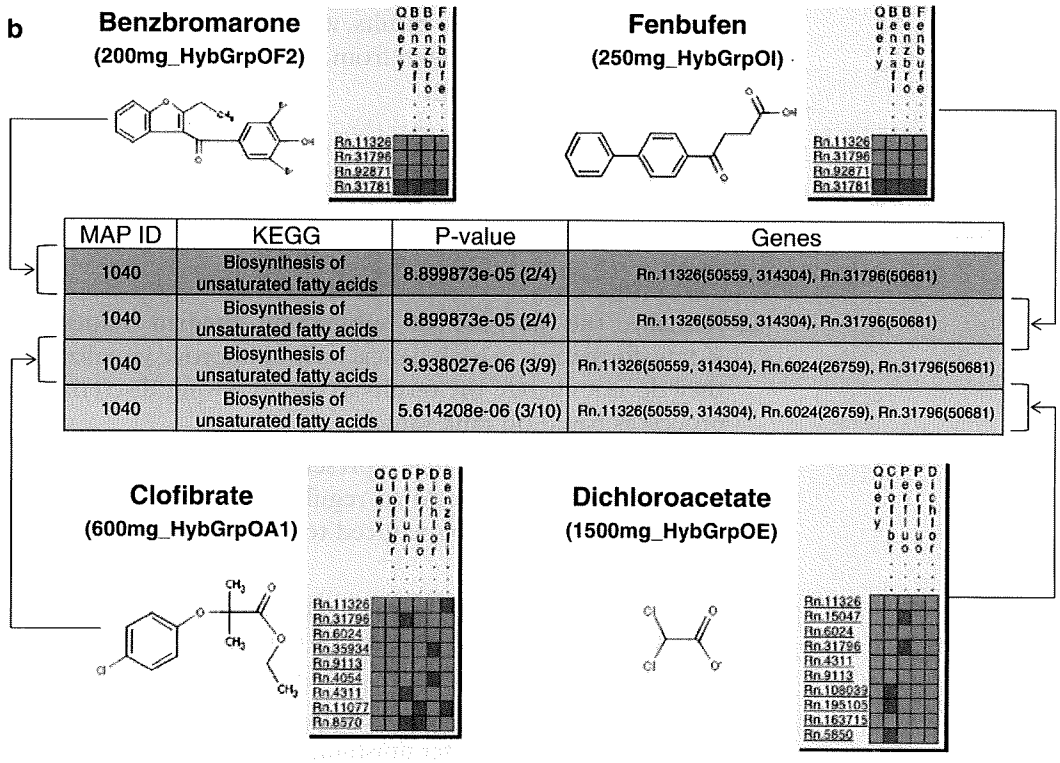
a

An example of partial profile in CellMontage format, where only three genes and their expression values are shown.

```
>Clofibrate_600mg_HybGrpOA2
Rn.98209:0.26226233    Rn.53257:0.550381825
Rn.94195:-0.285033381
```

A gene expression profile consists of a single description line, starting with ">", followed by one or more data lines. The data consist of elements, separated by a space or a new line. Each element consists of a UniGene identifier and an expression value separated by a colon.

b



MAP ID	KEGG	P-value	Genes
1040	Biosynthesis of unsaturated fatty acids	8.899873e-05 (2/4)	Rn.11326(50559, 314304), Rn.31796(50681)
1040	Biosynthesis of unsaturated fatty acids	8.899873e-05 (2/4)	Rn.11326(50559, 314304), Rn.31796(50681)
1040	Biosynthesis of unsaturated fatty acids	3.938027e-06 (3/9)	Rn.11326(50559, 314304), Rn.6024(26759), Rn.31796(50681)
1040	Biosynthesis of unsaturated fatty acids	5.614208e-06 (3/10)	Rn.11326(50559, 314304), Rn.6024(26759), Rn.31796(50681)

Fig. 3. Examples of input and output of SAMURAI-GUI system. An example of analysis with queries tnts (**a**) the CellMontage format as input and (**b**) graphical output of modules that share common biological function. Four example modules are searched by Benzbromarone, Fenbufen, Clofibrate, and Dichloroacetate. KEGG pathway functions for each module with the first and the second most significant *p* values are only shown.

3. Format data as described in 2.2 "Formatting Database by Discretization." Use the "formatdb" command. To discretize your data with the thresholds, as exemplified in 2.2, type "%formatdb dataset_file 0.03 (or 0.05)." This command takes a while and creates both the discretized data "dataset_file.db" and its gene index file "dataset_file.idx."
4. Run the LCM algorithm to extract core modules and merge them by typing: "% java –Xmx2000m xSamurai –M –i dataset_file.idx –d dataset_file.db –q query_file –n noise_threshold –s minimum_module –r result_dir."

   The query_file must also be written in the CellMontage format. The "noise_threshold" must be in the range of [0,1]. The "minimum_module" parameter sets the minimum size of gene modules to output. The "result_dir" parameter indicates the directory to write the results of final gene modules.
5. To perform module evaluations with GO and KEGG, execute the command: "%EA = GO_KEGG_test GO_KEGG_test/assignKEGG.pl dataset_file.idx P result_dir/*."

   The "GO_KEGG_test" is the directory for GO/KEGG evaluation package (*see* **Note** 7).
6. The GUI-based web version for multi-CPU calculation and module visualization in color is also available from a commercial site. To view free test results, visit: http://samurai.cbrc.jp/ and try the "Module search from a large-scale database" Web page (*again, see* **Note** 6).

3. Notes

1. Actually, the purpose of subtracting median values and Z-transformation in each data is only to improve visualization; they do not change the results as the following discretization process is based on rank values.
2. With an average query that represents only 10% of its discrete values are active (up- or down-regulated) genes, it is estimated that the database will be reduced to ($0.05 \times 0.05 \times 2 =$) 0.5% of its original size.
3. LCM program usually takes several seconds to finish, but sometimes takes several minutes. A faster program is currently available from the developers' Web site.
4. Many versions of merging processes are available. For example, implementing merge after finishing all the "checks" of module pairs is an option. Take the best approach depending on the situation (computational resource, purpose, etc.).

5. The conversion of genes from UniGene into EntrezGenes generates often more than single (one-to-one) correspondences.
6. The full license and web-enhanced server versions are available from HPC Solutions Inc. (http://www.hpc-sol.co.jp/).
7. Before running GO/KEGG evaluation script, do not forget to download necessary data files, including gene2unigene, gene2go.gz, and KEGG pathway maps. To do this, add your species code to "getKEGGmaps.pl" script in the "GO_KEGG_test" directory and then type "make compile" at the top directory.

Acknowledgments

We would like to thank Dr. Takeaki Uno at National Institute of Informatics for advice and kindly providing the LCM program for free use.

References

1. Barrett, T., Suzek, T.O., Troup, D.B., Wilhite, S.E., Ngau, W.C., Ledoux, P., Rudnev, D., Lash, A.E., Fujibuchi, W., and Edgar, R. (2005) NCBI GEO: mining millions of expression profiles – database and tools. *Nucleic Acids Res.* **33** (Database issue), D562–D566.
2. Brazma, A., Parkinson, H., Sarkans, U., Shojatalab, M., Vilo, J., Abeygunawardena, N., Holloway, E., Kapushesky, M., Kemmeren, P., Lara, G.G., Oezcimen, A., Rocca-Serra, P., and Sansone, S.A. (2003) ArrayExpress – a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* **31**, 68–71.
3. Cheng, Y., and Church, G. (2000) Biclustering of expression data. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, 93–103.
4. Tanay, A., Sharan, R., and Shamir, R. (2002) Discovering statistically significant biclusters in gene expression data. *Bioinformatics* **18**, S136–S144.
5. Ben-Dor, A., Chor, B., Karp, R., and Yakhini, Z. (2002) Discovering local structure in gene expression data: the order-preserving submatrix problem. *Proceedings of the 6th Annual International Conference on Computational Biology, ACM Press, New York, NY, USA*, 49–57.
6. Murali, T.M., and Kasif, S. (2003) Extracting conserved gene expression motifs from gene expression data. *Pac. Symp. Biocomput.* **8**, 77–88.
7. Ihmels, J., Bergmann, S., and Brkai, N. (2004) Defining transcription modules using large-scale gene expression data. *Bioinformatics* **20**, 1993–2003.
8. Wu, C.J., and Kasif, S. (2005) GEMS: a web server for biclustering analysis of expression data. *Nucleic Acids Res.* **33**, W596–W599.
9. Prelic, A. et al. (2006) A systematic comparison and evaluation of biclustering methods for gene expression data. *Bioinformatics* **22**, 1122–1129.
10. Okada, Y., and Fujibuchi, W. (2007) Mining a Large-scale Microarray Database for Similar Gene Expression Modules to Find Distant Relationships between Down Syndrome and Huntington's Disease. *Proceedings of Critical Assessment of Microarray Data Analysis 07, Valencia, Spain.*
11. http://camda.bioinfo.cipf.es/camda07/
12. Uno, T., Asai, T., Uchida, Y., and Arimura, H, (2004) An efficient algorithm for enumerating closed patterns in transaction databases, *Lecture Notes in Artificial Intelligence* **3245**, 16–31.
13. Uno, T., Kiyomi, M., and Arimura, H, (2002) LCM ver.2: Efficient Mining Algorithms for Frequent/Closed/Maximal Itemsets, *IEEE ICDM'04 Workshop FIMI'04* **126**.
14. Fujibuchi, W., Kiseleva, L., Taniguchi, T., Harada, H. and Horton, P. (2007) CellMontage: Similar Expression Profile Search Server, *Bioinformatics* **23**, 3103–3104.

Down-regulation of cIAP2 enhances 5-FU sensitivity through the apoptotic pathway in human colon cancer cells

Hideaki Karasawa,[1] Koh Miura,[1,4] Wataru Fujibuchi,[2] Kazuyuki Ishida,[3] Naoyuki Kaneko,[1] Makoto Kinouchi,[1] Mitsunori Okabe,[1] Toshinori Ando,[1] Yukio Murata,[1] Hiroyuki Sasaki,[1] Kazuhiro Takami,[1] Akihiro Yamamura,[1] Chikashi Shibata[1] and Iwao Sasaki[1]

[1]Division of Biological-Regulation and Oncology, Department of Surgery, Tohoku University Graduate School of Medicine, 1-1 Seiryo-machi, Aoba-ku, Sendai, Miyagi, Japan 980-8574; [2]Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, 2-42 Aomi, Koto-ku, Tokyo, Japan 135-0064; [3]Department of Pathology, Tohoku University Hospital, 1-1 Seiryo-machi, Aoba-ku, Sendai, Miyagi, Japan 980-8574

(Received August 31, 2008/Revised December 28, 2008/Accepted January 7, 2009/Online publication March 2, 2009)

Currently 5-fluorouracil (5-FU) plays a central role in the chemotherapeutic regimens for colorectal cancers and thus it is important to understand the mechanisms that determine 5-FU sensitivity. The expression profiles of human colon cancer cell line DLD-1, its 5-FU-resistant subclone DLD-1/FU and a futher 21 types of colon cancer cell lines were compared to identify the novel genes defining the sensitivity to 5-FU and to estimate which population of genes is responsible for 5-FU sensitivity. In the hierarchical clustering, DLD-1 and DLD-1/FU were most closely clustered despite over 100 times difference in their 50% inhibitory concentration of 5-FU. In DLD-1/FU, the population of genes differentially expressed compared to DLD-1 was limited to 3.3%, although it ranged from 4.8% to 24.0% in the other 21 cell lines, thus indicating that the difference of 5-FU sensitivity was defined by a limited number of genes. Next, the role of the *cellular inhibitor of apoptosis 2* (*cIAP2*) gene, which was up-regulated in DLD-1/FU, was investigated for 5-FU resistance using RNA interference. The down-regulation of cIAP2 efficiently enhanced 5-FU sensitivity, the activation of caspase 3/7 and apoptosis under exposure to 5-FU. The immunohistochemistry of cIAP2 in cancer and corresponding normal tissues from colorectal cancer patients in stage III revealed that cIAP2 was more frequently expressed in cancer tissues than in normal tissues, and cIAP2-positive patients had a trend toward early recurrence after fluorouracil-based chemotherapy. Although the association between drug sensitivity and the IAP family in colorectal cancer has not yet been discussed, cIAP2 may therefore play an important role as a target therapy in colorectal cancer. (*Cancer Sci* 2009; 100: 903–913)

5-fluorouracil (5-FU) is an anticancer drug that has been mainly used in the treatment of colorectal cancers. Recently, 5-FU has been combined with oxaliplatin or irinotecan as the first-line treatment for advanced colorectal cancers and these have significantly improved the response rates to 40–50% and prolonged overall survival.[1,2] Furthermore, novel biological agents including monoclonal antibodies such as cetuximab, which is an antibody against epidermal growth factor receptor (EGFR), and bevacizumab, which is an antibody against vascular endothelial growth factor, have been shown to provide additional clinical benefit for patients with metastatic colorectal cancers.[3–5] However, there are still a large number of patients who do not benefit from the present treatments because of anticancer drug resistance. Elucidating the mechanisms by which 5-FU resistance arises in colorectal cancer therefore remains an important issue for either overcoming or predicting such resistance.

5-FU is an analog of uracil and is rapidly incorporated into the cells using the same transport system as uracil.[6] Subsequently, 5-FU is converted into active metabolites which disrupt the action of thymidylate synthetase (TS) and RNA synthesis. TS and 5-FU-metabolizing enzymes such as dihydropyrimidine dehydrogenase (DPD) and thymidine phosphorylase (TP) have been analyzed to elucidate 5-FU resistance.[7] However the resistance to 5-FU has not been sufficiently explained by the metabolic pathway of 5-FU alone, because multiple factors participate in chemoresistance.[8] Recently, complementary DNA (cDNA) microarray technology has been used to identify novel genes regulating 5-FU resistance, and the potential biomarkers of 5-FU resistance other than pyrimidine metabolism-related enzymes have been proposed.[9,10]

Apoptosis is found to be one of the primary mechanisms of the cytotoxic effect of chemotherapeutic agents and inhibition of the apoptotic pathway is one of the factors that may be responsible for drug resistance.[11–13] In the process of apoptosis, the caspase cascade plays a central role,[14,15] and the inhibitor of apoptosis protein (IAP) family is thought to prevent apoptosis through direct caspase and pro-caspase inhibition (primarily caspase 3 and 7). The IAPs have been described to be abnormally regulated in various types of cancers,[16,17] and recently they have been regarded as therapeutic targets of cancer.[18,19] Although the association between IAPs and drug resistance has been discussed in cancers of some organs such as lung, pancreas and kidney,[20–22] it has not been fully analyzed in human colorectal cancer.

The present study compared the messenger RNA (mRNA) expression profiles between the human colon cancer cell line DLD-1 and its 5-FU-resistant subclone DLD-1/FU by cDNA microarray to investigate the novel genes regulating 5-FU resistance. To estimate which population of genes are responsible for regulating 5-FU sensitivity or resistance, the expression profiles of DLD-1 and DLD-1/FU were also compared to another 21 types of colon cancer cell lines. Next, the role of the cellular IAP 2 (cIAP2) gene, which is most highly expressed among genes of the IAP family in DLD-1/FU, was investigated using RNA interference (RNAi) on the sensitivity to 5-FU, the activation of caspase 3/7, and apoptosis in human colon cancer cells. Finally, to identify the association between cIAP2 expression and 5-FU resistance in human primary colorectal cancer, immunohistochemistry for

[4]To whom correspondence should be addressed.
E-mail: k-miura@surg1.med.tohoku.ac.jp

Abbreviations: 5-FU, 5-fluorouracil; TS, thymidylate synthetase; DPD, dihydropyrimidine dehydrogenase; TP, thymidine phosphorylase; IAP, inhibitor of apoptosis protein; RNAi, RNA interference; FBS, fetal bovine serum; MTS, 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfenyl)-2H-tetrazolium, inner salt; IC_{50}, inhibitory concentration 50%; SD, standard deviation; RT-PCR, reverse transcription polymerase chain reaction; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; siRNA, small interfering RNA; DW, distilled water; PI, propidium iodide; FITC, fluorescein isothiocyanate; CTNNA1, catenin alpha 1; HBEGF, heparin-binding EGF-like growth factor; PLA2G2A, phospholipase A2 group IIA; OPRT, orotate phosphoribosyl transferase; EGFR, epidermal growth factor receptor; NCBI, National Center for Biotechnology Information.

doi: 10.1111/j.1349-7006.2009.01112.x

cIAP2 was analyzed on cancer and corresponding normal tissues from colorectal cancer patients with curative operations followed by fluorouracil-based adjuvant chemotherapies.

Materials and Methods

Cell lines and reagents. The human colon cancer cell lines (Clone A, COLO205, COLO320, CX-1, DLD-1, HCT-8, HCT-15, HCT116, HT-29, KM12C, LoVo, LS174T, LS180, MIP101, RKO, SW48, SW480, SW620, SW948, SW1116, T84 and WiDr-TC) were either provided from the Cell Resource Center for Biomedical Research, Institute of Development, Aging and Cancer, Tohoku University (Sendai, Japan), or purchased from RIKEN BioResource Center (Tsukuba, Japan) and the American Type Culture Collection (Manassas, VA, USA). The 5-FU-resistant subclone DLD-1/FU was generously provided from Dr M. Fukushima (Taiho Pharmaceutical, Co. Ltd, Tokyo, Japan). The DLD-1/FU cell line was originally derived from the DLD-1 cell line by continuous *in vitro* exposure of DLD-1 cells to increasing concentrations of 5-FU through a number of successive passages, as described earlier.[23] The cells were cultured in either the recommended medium: RPMI1640 (Sigma-Aldrich, St. Louis, MO, USA), L-15 Medium Leibovitz (Sigma-Aldrich), 1 × McCoy 5A (MP Biomedicals, Solon, OH, USA), or D-MEM/F-12 (Invitrogen, Carlsbad, CA, USA), containing 10% heat-inactivated fetal bovine serum (FBS; Sigma-Aldrich, St Louis, MO, USA) and 1% penicillin-streptomycin (Invitrogen). 5-FU was kindly provided by Kyowa Hakko Kogyo (Tokyo, Japan).

Cell proliferation assay and drug cytotoxicity assay. Cells were seeded in 100-mm culture plate at a density of 5×10^4 cells/plate for the cell proliferation assay. On the indicated days, the viable number of cells was determined using a hemocytometer under a light microscope using the trypan blue exclusion method.

Drug-induced cytotoxicity was assessed by the 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfenyl)-2H-tetrazolium, inner salt (MTS) assay using the CellTiter 96 Aqueous One Solution Proliferation Assay (Promega, Madison, WI, USA). First, 50 μL of cell suspension was seeded in 96-well plates at a density of 5×10^3 cells/well. All plates were incubated for 24 h at 37°C in a humidified 5% CO_2 atmosphere. Subsequently, 10 dilutions of 5-FU were prepared in growth medium. After incubation, 50 μL of growth medium with diluted 5-FU or growth medium only (as a control) was distributed in 96-well plates. The plates were incubated for 72 h at 37°C. Following incubation, the drugs were removed. Then, fresh medium with MTS was added to each well and the cultures were incubated for 2 h at 37°C. The absorbance of formazan at 490 nm, considered to be directly proportional to the number of living cells in the culture,[24] was measured using a plate reader Multiskan JX (Thermo Fisher Scientific, Yokohama, Japan). The cytotoxic effect of 5-FU was assessed by the 50% inhibitory concentration (IC_{50}: inhibitory drug concentration that results in 50% cell survival) value. With the approximation formula obtained from the straight-line portion of the graph showing the cell survival rate for each dilution of drug, the IC_{50} value was calculated for subsequent analysis.

Total RNA isolation and reverse transcription. Total RNA was extracted from cell lines using the RNeasy Mini Kit (Qiagen, Valencia, CA, USA). The quality and quantity of the extracted total RNA were confirmed by electrophoresis on 1.2% denaturing agarose gels. cDNA was synthesized using a SuperScript III First-Strand Synthesis System for Reverse Transcription – Polymerase Chain Reaction (RT-PCR) Kit (Invitrogen) from 5 μg of total RNA. All the processes were carried out according to the manufacturer's instructions.

cDNA microarray analysis. The CodeLink Uniset Human 20KI Expression Bioarray (GE Healthcare Bio-Sciences, Piscataway, NJ, USA) was used for the cDNA microarray analysis. cRNA synthesis was carried out following the manufacturer's instructions. All of the following reagents were included in the CodeLink Expression Assay Reagent Kit (GE Healthcare Bio-Sciences). First-strand cDNA was generated from 2 μg of total RNA using reverse transcriptase and T7 oligo(dT) primer. Subsequently, second-strand cDNA was produced using *Escherichia coli* DNA polymerase mix and RNase H. The resultant double-stranded cDNA was purified on QIAquick PCR Purification Kit (Qiagen) and cRNA as a probe for microarray was generated by *in vitro* transcription reaction using T7 RNA polymerase and biotin-11-UTP (Perkin Elmer, Boston, MA, USA). cRNA was purified on the RNeasy Mini Kit (Qiagen), quantified by spectrophotometry and 10 μg was then fragmented by heating at 94°C for 20 min in the presence of magnesium ions. The fragmented cRNA was hybridized overnight at 37°C in hybridization buffer to each array in an Innova 4080 Shaking Incubator (New Brunswick Scientific, Edison, NJ, USA) for 18 h. After hybridization, the arrays were washed in 0.75 × TNT buffer (1 × TNT: 0.10 M Tris-HCl [pH 7.6], 0.15 M NaCl, 0.05% Tween-20) at 46°C for 1 h followed by incubation with streptavidin-Cy5 at room temperature for 30 min in the dark. The arrays were then washed in 1 × TNT twice for 5 min each followed by a rinse in 0.05% Tween-20 in water and then dried. Glass slides were scanned using a GenePix 4000 A Scanner (Axon Instruments, Union City, CA, USA). The grids of the image spots were adjusted and their signals were analyzed using the CodeLink System Software (GE Healthcare Bio-Sciences). For each cell line, the experiments were performed either two or three times, independently for all of the steps through cell culture to microarray experiment, to confirm the reproducibility of experiments. To compare the gene expression values from the various experiments, the following array-based normalization was applied. In each array experiment, a set of values was log-transformed and Z-normalized by the mean and the standard deviation (SD) calculated from the 5th to 95th percentiles of all non-marker genes. An unsupervised hierarchical clustering analysis was applied using Cluster 3.0 and Java TreeView programs.[25,26]

RT-PCR. Gene-specific primer sets were designed using the Primer Express Software ver. 2.0 (Applied Biosystems, Foster City, CA, USA). Real-time RT-PCR using cDNA of cell lines was carried out using the Power SYBR Green PCR Master Mix (Applied Biosystems) following the manufacturer's instructions. Triplicate cDNA of each cell line was applied to 96-well reaction plates. Thermal cycling was carried out in the following steps: one cycle of 95°C for 10 min; then 40 cycles of 95°C for 15 s, 60°C for 1 min. *Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)* was used as an endogenous control. The mRNA expression level of *cIAP2* was normalized to that of the *GAPDH* in the corresponding sample.

Western blotting. Human colon cancer cells were harvested with trypsin/EDTA (edetic acid) and phosphated-buffered saline (PBS)-washed cell pellets were treated with EBC lysis buffer (1% NP-40, 40 mM Tris-HCl pH 8.0, 100 mM NaCl). Electrophoresis was performed using 4–20% Tris-Glycine Gels (Invitrogen) and proteins were electro-transferred onto Sequi-Blot PVDF (polyvinylidene difluoride) Membrane (Bio-Rad Laboratories, Hercules, CA, USA). The membranes were probed with the following primary antibodies: rabbit polyclonal anticIAP2 antibody (Santa Cruz Biotechnology, Santa Cruz, CA, USA), or rabbit polyclonal anti-α actin antibody (AbCam, Cambridge, UK) as a control, at 4°C overnight. The membranes were washed and subsequently incubated with horseradish peroxidase-coupled donkey antirabbit antibody (Santa Cruz Biotechnology) for 60 min. All proteins were visualized using SuperSignal West Pico Chemiluminescent Substrate (PIERCE, Rockford, IL, USA).

RNA interference (RNAi). Human colon cancer cells were trypsinized and plated into 6-well plates. After 24 h, the cells were transfected with 50 nM ON-TARGETplus *SMART*pool cIAP2 siRNA (small interfering RNA; Dharmacon, Lafayette, CO, USA) by using DharmaFECT siRNA Transfection Reagent 4 (Dharmacon) according to the manufacturer's instructions. For control experiments, ON-TARGETplus siCONTROL Non-targeting siRNA

doi: 10.1111/j.1349-7006.2009.01112.x

(Dharmacon) was used under the same conditions. Between 48 and 120 h after transfection, gene silencing was examined with real-time RT-PCR and Western blotting.

Caspase 3/7 assay and flow cytometry analysis with the annexin V/propidium iodide (PI) staining. Caspase 3 and 7 activation assays were performed using the Caspase-Glo 3/7 Assay Kit (Promega) according to the manufacturer's instructions. Briefly, the cells were seeded in 96-well plates at a density of 5×10^3 cells/well. After the cells were treated with 5-FU or distilled water (DW), Caspase-Glo 3/7 Reagent (100 μL) was added to each well. The plate was then incubated at room temperature for 1 h and the luminescence of each sample was measured with a Centro LB960 96-well Luminometer (Berthold Technologies, Natick, MA, USA). Because the luminescence of this assay was proportional to the cell number with the preliminary experiments, caspase 3/7 activity was assessed by the luminescence compensated by cell number.

In addition, apoptosis was measured by the Annexin V-FITC Apoptosis Detection Kit I (BD Biosciences, San Jose, CA, USA). Cells were treated with annexin V-fluorescein isothiocyanate (FITC) and PI according to the manufacturer's protocol and were analyzed by multicolor flow cytometry using FACS Calibur with Cell-Quest Software (BD Biosciences).

Immunohistochemistry. Primary colorectal cancer and corresponding normal tissue specimens were obtained from 40 colorectal cancer patients with locoregional lymph node metastasis in stage III according to the TNM (tumor–node–metastasis) classification. All patients had undergone curative operations at Tohoku University Hospital, from 1999 to 2004, and were treated with fluorouracil-based adjuvant chemotherapies (5-FU/leucovorin, doxifluridine or uracil/tegafur). Written informed consent was obtained from all patients. The median age of the patients was 61.0 years (range 30–76). Nineteen patients were female and 21 were male. The median postoperative follow-up time was 70.5 months (range 13.6–110.7).

Formalin-fixed paraffin-embedded tissue sections (3 μm thick) were deparaffinized in xylene and rehydrated in graded alcohol dilutions. Endogenous peroxidase activity was blocked by 3% hydrogen peroxidase for 10 min at room temperature. Antigen retrieval was performed using an autoclave in 0.01 M citrated buffer (pH 6.0) at 121°C for 5 min. For the reduction of non-specific staining, the sections were exposed to 10% rabbit serum for 30 min. They were incubated at 4°C overnight with goat polyclonal anticIAP2 antibody (R & D Systems, Minneapolis, MN, USA) at 1:200 dilution. Secondary antibody reaction was performed using biotinylated rabbit antigoat antibody (Dako, Copenhagen, Denmark) at 1:800 dilution for 30 min at room temperature and peroxidase-conjugated streptavidin (Nichirei Bioscience, Tokyo, Japan) was used according to the manufacturer's instructions. The reacted sections were visualized using 3,3′-diaminobenzidine solution (1 mM 3,3′-diaminobenzidine, 50 mM Tris-HCl [pH 7.6] and 0.006% H_2O_2) and counterstained with hematoxylin for nuclear staining. Immunoreactivity for cIAP2 expression was graded as positive if >10% of cancer or normal epithelial cells were stained and as negative if <10% of cells were stained.

Statistical analysis. Any statistical significance of the mRNA expression level, caspase activity and time to recurrence was determined using either Student's *t*-test or Mann–Whitney *U*-test. The Chi-square test was carried out to test the association between cIAP2 expression and cancer recurrence. The computer program Statcel2 Software (OMS Publishing, Saitama, Japan) and Microsoft Excel 2007 (Microsoft, Redmond, WA, USA) were used for statistical analysis. Values of $P < 0.05$ were considered to be statistically significant.

Results

The growth rates in DLD-1 and DLD-1/FU and cytotoxicity of 5-FU. The growth rates of parental DLD-1 cells and its 5-FU-resistant subclone DLD-1/FU were equivalent (Fig. 1A). However, 5-FU sensitivity of DLD-1/FU was quite different from that of DLD-1 and the IC_{50} of 5-FU in DLD-1/FU was higher than that in DLD-1 by over 100 times: 55.1 ± 10.7 and 0.488 ± 0.013 μg/mL, respectively (Figs 1B and 2B). Among the 21 cell lines other than DLD-1 and DLD-1/FU, the IC_{50} of 5-FU varied from 0.00998 to 9.96 μg/mL. SW620, HCT-8 and SW480 showed high IC_{50} of 5-FU: 9.96, 4.60 and 4.33 μg/mL, respectively; and WiDr-TC, HCT116 and SW48 showed low IC_{50} of 5-FU: 0.00998, 0.0499 and 0.149 μg/mL, respectively (Fig. 2B).

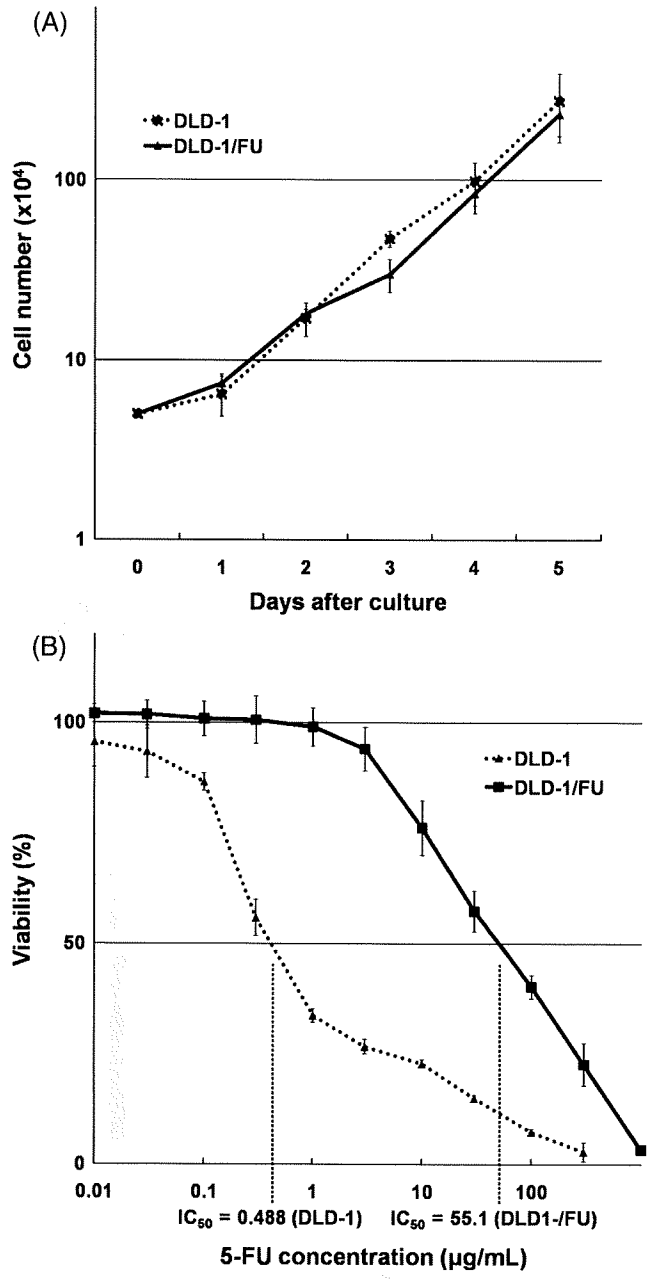


Fig. 1. (A) Proliferation assays of DLD-1 and DLD-1/FU cells. The cell numbers were determined on the indicated days using trypan blue counting. The data represent the means ± SD of triplicate cultures. (B) Cytotoxicity of 5-FU and 50% inhibitory concentration in DLD-1 and DLD-1/FU cells. The MTS assay was performed 72 h after treatment with 5-FU. The data represent the means ± SD of three independent experiments.

cDNA microarray analysis and isolation of genes regulating 5-FU resistance. On average 13 472 (67.4%) out of the 19 982 probes in the microarray experiments with the 23 colon cancer cell lines

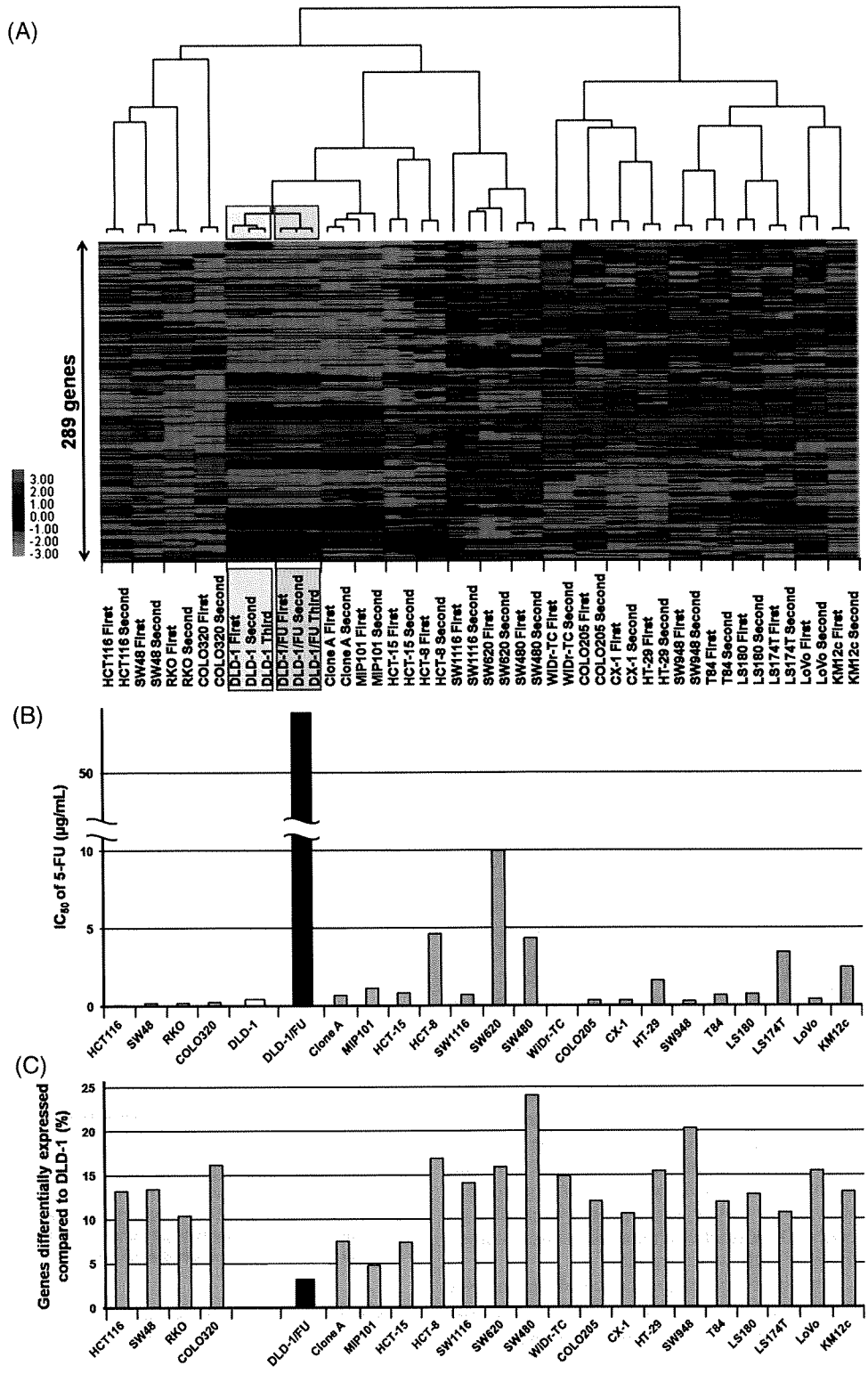


Fig. 2. (A) A dendrogram and its image plot of the hierarchical clustering analysis of 23 colon cancer cell lines including DLD-1 and DLD-1/FU. A cohort of 289 genes out of 13 472 genes, the expression ratios of which varied by SDs of >1.25, were filtered using program Cluster 3.0. (B) The 50% inhibitory concentration of 5-FU in 23 colon cancer cell lines ranged from 0.00998 to 55.1 μg/mL. The data represent the means of two or three independent experiments. (C) Percentages of differentially expressed genes compared to DLD-1. The population of genes with more than a two-fold change in comparison to DLD-1 ranged from 3.3% to 24.0%.

were assessed as 'expressed genes' that gave larger expression values than the 'raw threshold' which was calculated from a set of bacterial marker genes included in the CodeLink microarray as experimental controls.

To evaluate the reproducibility of the microarray experiments and to assess the similarities or differences of gene expression among the colon cancer cells, an unsupervised clustering analysis was performed using 289 genes out of 13 472 genes, the expression

doi: 10.1111/j.1349-7006.2009.01112.x