菅野 純	Percellome Project による毒性トランスクリプトミクスの新しい試み	細胞工学	26(1)		
	毒性の高精細解析に向けてのトキシコゲノミクス、	医学のあゆみ	218 No.12	1035-6	2006

The Journal of Toxicological Sciences,
Vol.31, No.5, 433-448, 2006

UTILIZATION OF A ONE-DIMENSIONAL SCORE FOR SURVEYING CHEMICAL-INDUCED CHANGES IN EXPRESSION LEVELS OF MULTIPLE BIOMARKER GENE SETS USING A LARGE-SCALE TOXICOGENOMICS DATABASE

Naoki KIYOSAWA[1], Kouji SHIWAKU[1], Mitsuhiro HIRODE[1], Ko OMURA[1], Takeki UEHARA[1], Toshinobu SHIMIZU[1], Yumiko MIZUKAWA[2], Toshikazu MIYAGISHIMA[1], Atsushi ONO[1], Taku NAGAO[3] and Tetsuro URUSHIDANI[1,2]

[1]*Toxicogenomics Project, National Institute of Biomedical Innovation, 7-6-8 Asagi, Ibaraki, Osaka 567-0085, Japan*

[2]*Department of Pathophysiology, Faculty of Pharmaceutical Sciences, Doshisha Women's College of Liberal Arts, Kodo, Kyotanabe, Kyoto 610-0395, Japan*

[3]*National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan*

(Received August 18, 2006; Accepted September 11, 2006)

ABSTRACT — A large-scale toxicogenomcis database has now been constructed in the Toxicogenomics Project in Japan (TGP). To facilitate the analytical procedures for such large-scale microarray data, we developed a simple one-dimensional score, named TGP1 which expresses the trend of the changes in expression of biomarker genes as a whole. To evaluate the usefulness of the TGP1 score, microarray data of rat liver and rat hepatocytes deposited in the TGP database were scored for three biomarker gene sets, i.e., carcinogenesis-related, PPARα-regulated and glutathione depletion-related gene sets. The TGP1 scoring system gave reasonable results, i.e., the scores for carcinogenesis-related genes were high in omeprazole-, chlorpromazine-, hexachlorobenzene-, sulfasalazine- and Wy-14,643-treated rat livers, that for PPARα-regulated genes were high in clofibrate-, Wy-14,643-, gemfibrozil-, benzbromarone- and aspirin-treated rat livers as well as rat hepatocytes, and for glutathione deficiency-related genes were high in omeprazole-, bromobenzene-, acetaminophen- and coumarin-treated rat liver. We concluded that the TGP1 score is useful for surveying the expression changes in multiple biomarker gene sets for a large-scale toxicogenomics database, which would reduce the time of doing conventional multivariate statistical analysis. In addition, the TGP1 score can be applied to screening of compatible biomarker gene sets between rat liver and rat hepatocytes, like PPARα-regulated gene sets, which will allow us to develop an appropriate in vitro system for drug safety assessment in vivo.

KEY WORDS: Toxicogenomics, Database, Scoring system, Liver, Rat

INTRODUCTION

Toxicogenomics has been considered to be a promising methodology for understanding the molecular mechanisms of toxicity, and this has been proven by a number of studies (Kiyosawa *et al.*, 2004b; Sehata *et al.*, 2004; Ito *et al.*, 2006). Toxicogenomics research requires a high-quality microarray database that covers a sufficient number of well-studied compounds, and such databases are being developed for both public and commercial use (Boverhof and Zacharewski, 2006). The Toxicogenomics Project in Japan (TGP) has been conducted by the collaborative research of 15 pharmaceutical companies, the National Institute of Health Science and the National Institute of Biomedical Innovation, and is a five-year project started from 2002 (Takashima *et al.*, 2006; Urushidani and Nagao, 2005). The TGP database is a toxicogenomics-oriented data-

Correspondence: Tetsuro URUSHIDANI (E-mail: turushid@dwc.doshisha.ac.jp)

base that consists of microarray data of rat liver, rat hepatocytes and human hepatocytes, as well as traditional toxicological data for 150 compounds. The large-scale TGP database is expected to be useful for characterizing the chemical-induced molecular dynamics and for evaluating and predicting the potential risk of chemical toxicity at early stages of drug development.

Specific gene sets whose mRNA levels are correlated with certain toxicological phenotypes, or toxicogenomics biomarkers, are useful for evaluating the toxicological significance from microarray data. For example, previous reports demonstrated that certain gene sets could be applied for evaluating a chemical-induced glutathione deficiency or risk of acetaminophen-type hepatotoxicity (Kiyosawa *et al.*, 2004a), a chemical-induced phospholipidosis not only in rodent but also in the human case (Sawada *et al.*, 2005), and evaluation of the carcinogenic potential of the chemicals (Ellinger-Ziegelbauer *et al.*, 2004). Such candidate biomarkers have been rapidly accumulated. The significance of the toxicogenoimics database will grow synergistically with accumulation of toxicogenomics biomarker information.

When using the toxicogenomics database, one of the major obstacles is its gigantic data size. Furthermore, it becomes more complex as the toxicogenomics biomarker knowledge accumulates. Multivariate statistical techniques such as hierarchical clustering, k-means clustering, self-organizing map or principal component analysis are usually applied for analyzing the microarray data (Kaminski and Friedman, 2002; Draghici, 2003). Although such techniques provide the general characteristics of gene expression profiles rather intuitively, they are not suitable for high-throughput analysis when a series of biomarker gene sets is to be analyzed at the same time. One simple solution is to circumvent the time-consuming analytical procedure by utilizing a one-dimensional score, which reflects the level of gene expression changes for certain biomarker gene sets. The calculated scores for many samples can be presented in the multiple biomarker gene sets simultaneously, so it becomes easy to capture the toxicological endpoints that should be focused on for further detailed analysis. In the present study, we have developed a score that reflects the level of chemical-induced gene expression changes for certain biomarker gene sets in our large-scale TGP database. The score is calculated from the fold change value of each gene, calculated by dividing the mean signal of the chemical-treated group by that of the vehicle-treated group, and therefore is easy to calculate and interpret results. The usefulness of the presented score was verified in three biomarker gene sets.

MATERIALS AND METHODS

Chemicals

Thirty-eight chemicals used for the data analysis are listed in Table 1, in which the chemical name, manufacturer, highest dosage and vehicle used in the study are described. In our standard protocol, all the drugs were administered in three dose levels, i.e., high, middle (1/3 of the high dose), and low (1/10 of the high dose). In the present study, however, the data of the high dose are exclusively shown for simplicity. In certain cases (Fig. 3), the results of other dose levels are also exhibited in order to show dose-dependency.

Animal treatment

Six-week old male Sprague-Dawley rats (Charles River Japan, Inc., Kanagawa, Japan) consisted of five animals per group and were used for the in-life study. The detailed study information is the same as in the literature (Takashima *et al.*, 2006). Rats were sacrificed at 24 hr after a single dosing of chemicals. Liver was removed at necropsy, and soaked in RNAlater® (Ambion Inc., Austin, TX, USA) to prevent RNA degradation. We affirm that experimental protocols for both the animal and hepatocyte studies were reviewed and approved by the Ethics Review Committee for Animal Experimentation of the National Institute of Health Science.

Hepatocytes treatment

Hepatocytes were isolated from 6-week-old male Sprague-Dawley rats under sodium pentobarbital (120 mg/kg, ip) and anesthetized by a modified two-step collagenase perfusion method. The liver was perfused via the portal vein for 10 min with divalent cation-free, EGTA (0.5 mM)-supplemented HEPES buffered Hank's balanced salt solution followed by a 10-min perfusion with HEPES (10 mM)-buffered normal Hank's balanced salt solution containing soybean trypsin inhibitor (Sigma, T-2011, 0.05 g/L) and collagenase (WAKO 034-10533, 0.5 g/L) at the flow rate of 10 - 30 ml/min. The isolated cells were washed three times by 50 g for 1 min to obtain a parenchymal cell-enriched pellet. Hepatocytes were not used when their viability assessed by trypan blue exclusion was lower than 70%. The cells were seeded into collagen-coated six-well plates (BD BioCoat™ Collagen I Cellware,

Table 1. Chemicals.

ID	Chemicals (abbreviation)	Manufacturer	*in vivo*		*in vitro*	
			Dosage (mg/kg)	Vehicle	Concentration (μM)	Vehicle
001	Acetaminophen (APAP)	Sigma-Aldrich	1000	MC	10000	Medium
002	Isoniazid (INAH)	Sigma-Aldrich	200	MC	10000	Medium
003	Carbon tetrachloride	Sigma-Aldrich	300	OIL	10000	DMSO
004	Phenobarbital	Sigma-Aldrich	100	MC	10000	Medium
005	Valproate sodium (VPA)	Wako Pure Chemical Industries, Ltd.	450	MC	5000	Medium
006	Clofibrate	Wako Pure Chemical Industries, Ltd.	300	OIL	300	DMSO
007	Methotrexate	Wako Pure Chemical Industries, Ltd.	100	MC	3000	Medium
008	Rifampicin	Wako Pure Chemical Industries, Ltd.	200	MC	70	DMSO
009	Alpha-naphthy lisothiocyanate (ANIT)	Kanto Chemical Co., Inc.	15	OIL	200	DMSO
010	Allyl alcohol	Tokyo Chemical Industry Co., Ltd	30	OIL	70	Medium
011	Phenylbutazone	Sigma-Aldrich	200	MC	400	DMSO
012	Omeprazole (OPZ)	Wako Pure Chemical Industries, Ltd.	1000	MC	600	DMSO
013	Ethionine (ET)	Tokyo Chemical Industry Co., Ltd	250	MC	10000	Medium
014	Aspirin (ASA)	Wako Pure Chemical Industries, Ltd.	450	MC	3000	DMSO
015	Indomethacin	Sigma-Aldrich	5	MC	200	DMSO
016	Chlorpromazine (CPZ)	Wako Pure Chemical Industries, Ltd.	45	MC	20	DMSO
017	Thioacetamide	Sigma-Aldrich	45	MC	10000	Medium
018	Carbamazepine	Sigma-Aldrich	300	MC	300	DMSO
019	Diclofenac sodium (DFNa)	CAYMAN / Tokyo Chemical Industry Co., Ltd.	10	MC	400	DMSO
020	Nitrofurantoin	ICN	100	MC	125	DMSO
021	Benzbromarone (BBr)	Sigma-Aldrich	200	MC	100	DMSO
022	Hexachlorobenzene (HCB)	Tokyo Chemical Industry Co., Ltd	300	OIL	30	DMSO
023	Diazepam	Wako Pure Chemical Industries, Ltd.	250	MC	250	DMSO
024	Cyclophosphamide	Sigma-Aldrich	15	MC	2000	Medium
025	Methapyrilene hydrochloride	Sigma-Aldrich	100	MC	600	Medium
026	Phenytoin	Tokyo Chemical Industry Co., Ltd	600	MC	300	DMSO
027	Coumarin (CMA)	Tokyo Chemical Industry Co., Ltd	150	OIL	300	DMSO
028	Allopurinol	Sigma-Aldrich	150	MC	140	DMSO
029	Propylthiouracil	Tokyo Chemical Industry Co., Ltd	100	MC	4000	Medium
030	Wy-14,643 (WY)	Tokyo Chemical Industry Co., Ltd	100	OIL	150	DMSO
031	Gemfibrozil (GFZ)	Sigma-Aldrich	300	OIL	100	DMSO
032	Bromobenzene (BBz)	Tokyo Chemical Industry Co., Ltd	300	OIL	200	DMSO
033	Amiodarone hydrochloride	Sigma-Aldrich	200	MC	7	DMSO
034	Sulfasalazine (SS)	Sigma-Aldrich	1000	MC	150	DMSO
035	Cimetidine	Tokyo Chemical Industry Co., Ltd	1000	MC	300	DMSO
042	Glibenclamide	Sigma-Aldrich	1000	OIL	20	DMSO
045	Perhexiline maleate	Sigma-Aldrich	150	MC	15	DMSO
046	Azathioprine	ICN	30	MC	75	DMSO

BD Bioscience) at a density of 1×10^6 cells/well in 2 ml HMC Bulletkit medium (CAMBREX) supplemented with 10% fetal bovine serum. Following an attachment period of 3 hr, the medium was replaced and kept overnight before drug exposure at 37°C in an atmosphere of 5% CO_2. The test compounds were added to the medium directly or as a 1000× stock solution in dimethylsulfoxide (DMSO). After 2, 8 and 24 hr-exposure, cells were dissolved with RLT buffer (Qiagen) and collected for expression profiling. GeneChip analysis was performed in a duplicated manner for each time and concentration point.

Microarray analysis

The detailed information is described in the previous literature (Takashima *et al.*, 2006). Briefly, 3 liver samples out of 5 samples collected in the animal were used for analysis. Total RNA was isolated using RNeasy Mini Kit with Bio Robot 3000 (Qiagen, Inc., Valencia, CA, USA), and 5 μg of the total RNA was used for cDNA synthesis using T7-(dT)24 oligonucleotide primer (Affymetrix, Inc., Santa Clara, CA, USA) and SuperScript Choice System (Invitrogen, Carlsbad, CA, USA). Biotin-labeled cRNA was synthesized using BioArray High Yield RNA Transcription Labeling Kit (Enzo Diagnostics, Farmingdale, NY, USA). The hybridization cocktail was prepared with 10 μg of fragmented cRNA, and hybridized to RAE 230A GeneChip array (Affymetrix, Inc.) at 45°C for 18 hr. The hybridized GeneChip array was washed and stained by streptavidin-phycoerythrin using Fluidics Station 400 (Affymetrix, Inc.) and scanned by GeneArray Scanner (Affymetrix, Inc.). The scanned image files were analyzed using Microarray Suite ver. 5.0. All the microarray data were scaled by global normalization where the mean signal intensity of each data was adjusted to 500. A heat map representing gene expression levels was created using Spotfire software (Spotfire, Inc., Somerville, MA).

Biomarker gene sets

Three lists of probe sets of RG U34A arrays are selected from the literature whose signals were reported to be increased by certain chemical treatments: i) carcinogenicity-related gene probe sets (Ellinger-Ziegelbauer *et al.*, 2004), ii) PPARα-regulated gene probe sets (Richert *et al.*, 2003), and iii) glutathione deficiency-correlated gene probe sets (Kiyosawa *et al.*, 2004a). The information for the selected probe sets is summarized in Table 2, and lists of the probe sets for each biomarker gene set are presented from Table 3 to Table 5. Since the selected probe sets were those of RG U34A array, the corresponding probe sets of RAE 230A array were determined by selecting "good match probe sets", where the probe sets whose corresponding probe sequences do not overlap between those of RG U34A and RAE 230A array were removed. The "good match probe set" information is provided in the NetAffx Website (Liu *et al.*, 2003).

Calculation of TGP1 score

The TGP1 score was calculated as shown in Fig. 1. Signal log ratio was calculated by dividing the mean signal value of the chemical-treated group by that of corresponding control. First, the sum of the signal log ratios for the used probe sets was calculated, and then divided by the number of probe sets used (Index 1). Next, the sum of squared signal log ratios for the used probe sets was calculated, and then divided by the number of probe sets used (Index 2). Finally, the TGP1 score was calculated by multiplying Index 1 with Index 2.

Individual gene expression analysis

The fold change in carcinogenesis- and glutathione deficiency-related genes is calculated by dividing signal data of the chemical-treated group by mean signal data (n=3 for rat liver and n=2 for rat hepatocytes) of corresponding control. The signal log ratio

Table 2. Summary of biomarker gene sets used in the present study.

Biomarker	Number of probe sets		Reference
	RG U34A	RAE 230A	
Carcinogenicity-related	55	26	Ellinger-Ziegelbauer *et al.*, 2004
PPARα-regulated	30	17	Richert *et al.*, 2003
Glutahione deficiency-related	69	45	Kiyosawa *et al.*, 2004

Three biomarker gene sets whose expression levels were reported to be increased by carcinogens, PPARα activators or glutathione depletors, were selected from the literature, and used for calculation of the TGP1 score. These genes were originally identified using a RG U34A GeneChip array, and the corresponding "good match" probe sets of the RAE 230A GeneChip array were re-selected according to the information provided by the NetAffx website.

Scoring the level of gene expression changes in microarray analysis.

Table 3. Carcinogenesis-related gene probe sets.

Affymetrix probe set ID		
RG U34A	RAE 230A	Gene name
X62952_at	1367574_at	Vimentin
L13039_s_at	1367584_at	Annexin A2
M58404_at	1367655_at	Thymosin, beta 10
X13044_at	1367679_at	CD74 antigen
rc_AI169327_at	1367712_at	Tissue inhibitor of metalloproteinase 1
D10729_s_at	1367786_at	Proteosome subunit, beta type 8
M32062_at	1367850_at	Fc receptor, IgG, low affinity III
M34253_at	1368073_at	Interferon regulatory factor 1
AF045464_s_at	1368121_at	Aldo-keto reductase family 7, member A3
M76704_s_at	1368311_at	*O*-6-Methylguanine-DNA methyltransferase
U62940_at	1368552_at	GrpE-like 1
U10894_s_at	1368558_s_at	Allograft inflammatory factor 1
AB010635_s_at	1368905_at	Carboxylesterase 2
U73174_g_at	1369061_at	Glutathione reductase
M31038_at	1369110_x_at	RT1 class Ib, locus Aw2
L12138_at	1369499_at	Thymidylate synthase
D42148_at	1369735_at	Growth arrest specific 6
U02322_s_at	1369783_a_at	Neuregulin 1
rc_AA900505_at	1369958_at	RhoB gene
X60351cds_s_at	1370026_at	Crystallin, alpha B
AJ011969mRNA_at	1370153_at	Growth differentiation factor 15
D10757_at	1370186_at	Proteosome subunit, beta type 9
E00717UTR#1_s_at	1370269_at	CYP1A1
U66470_at	1370361_at	Cell growth regulator with EF hand domain 1
M81855_at	1370583_s_at	ATP-binding cassette, sub-family B (MDR/TAP), member 1
J05181_at	1370688_at	Glutamate-cysteine ligase, catalytic subunit
AF084186_s_at	1370838_s_at	Alpha-spectrin 2
M15562_at	1370883_at	RT1 class II, locus Da
Z12298cds_s_at	1370956_at	Decorin
X56596_at	1371033_at	RT1 class II, locus Bb
S72506_s_at	1371089_at	Glutathione S-transferase Yc2 subunit
M63991_at	1371143_at	Serine (or cysteine) peptidase inhibitor, clade A, member 7
X75207_s_at	1371150_at	Cyclin D1
X51707cds_s_at	1371377_at	Ribosomal protein S19
rc_AI639488_at	1384427_at	Transformed mouse 3T3 cell double minute 2 (predicted)
AF083269_at	1386925_at	Actin-related protein 2/3 complex, subunit 1B
M60921_at	1386994_at	B-cell translocation gene 2, anti-proliferative
L03201_at	1387005_at	Cathepsin S
rc_AA875455_at	1387021_at	Wild-type p53-induced gene 1
AF001898_at	1387022_at	Aldehyde dehydrogenase family 1, member A1
U24174_at	1387391_at	Cyclin-dependent kinase inhibitor 1A
J02679_s_at	1387599_a_at	NAD(P)H dehydrogenase, quinone 1
M26125_at	1387669_a_at	Epoxide hydrolase 1
U17035_s_at	1387969_at	Chemokine (C-X-C motif) ligand 10
X57523_at	1388149_at	Transporter 1, ATP-binding cassette, sub-family B (MDR/TAP)
rc_AA893246_at	1388325_at	ATPase, H+ transporting, V1 subunit D
rc_AA944397_at	1388850_at	Heat shock protein 1, alpha

Table 3. Continued.

U49729_at	none	–
X70871_at	none	–
S56936_s_at	none	–
D38062exon_s_at	none	–
M17412_at	none	–
X59375mRNA_at	none	–
rc_AA894027_at	none	–
rc_AA945082_at	(1371089_at)	–

Probe sets whose signals are reported as increased by treatment of carcinogens in RG U34A GeneChip analysis were selected from the previous literature, and the corresponding probe sets of RAE 230A GeneChip were determined by referring to "good match probe sets" information provided by Affymetrix. Seven probe sets of RG U34A GeneChip did not show any corresponding probe sets in RAE 230A GeneChip, and were presented as "none" in the table. S72506_s_at and rc_AA945082_at are redundant probe sets of RG U34A GeneChip for 1371089_at probe set in RAE 230A GeneChip.

Table 4. PPARα-regulated gene probe sets.

Affymetrix probe set ID		
RG U34A	RAE 230A	Gene name
J02752_at	1367680_at	Acyl-Coenzyme A oxidase 1
AF072411_g_at	1367689_a_at	CD36 antigen
D43623_g_at	1367742_at	Carnitine palmitoyltransferase 1b
K03249_at	1368283_at	Enoyl-Coenzyme A, hydratase/3-hydroxyacyl Coenzyme A dehydrogenase
M57718mRNA_s_at	1368934_at	CYP4A10
V01235_at	1369111_at	Fatty acid binding protein 1, liver
X65296cds_s_at	1370363_at	Carboxylesterase 3
AF044574_g_at	1370818_at	2-4-Dienoyl-Coenzyme A reductase 2
rc_AA893239_at	1371012_at	2-Hydroxyphytanoyl-Coenzyme A lyase
L00320cds_f_at	1371076_at	CYP2B2
U08976_at	1386885_at	Enoyl coenzyme A hydratase 1
rc_AA946368_at	1386901_at	Similar to CD36 antigen
M21208mRNA_s_at	1387123_at	CYP17A1
J02749_g_at	1387783_a_at	Acetyl-Coenzyme A acyltransferase 1
Y09333_g_at	1388210_at	Acyl-CoA thioesterase 1
AB010428_s_at	1398250_at	Acyl-CoA thioesterase 1
AB005743_g_at	none	–
rc_AA924267_s_at	none	–
X07259cds_s_at	none	–
rc_AA925752_at	none	–
K01721mRNA_s_at	none	–
M14972_i_at	none	–
X72792cds_s_at	none	–
rc_AA799489_g_at	none	–
AF072411_at	(1367689_a_at)	–
L46791_at	(1370363_at)	–
J00728cds_f_at	(1371076_at)	–
M11251cds_f_at	(1371076_at)	–
J02749_at	(1387783_a_at)	–
Y09333_at	(1388210_at)	–

Probe sets whose signals are reported to be increased by treatment of PPARα activators in the RG U34A GeneChip analysis were selected from the literature, and the corresponding probe sets of RAE 230A GeneChip were determined by referring to "good match probe sets" information provided by Affymetrix. Eight probe sets of RG U34A GeneChip did not show any corresponding probe sets in RAE 230A GeneChip, and were presented as "none" in the table. Six probe sets of RG U34A GeneChip had redundant probe sets for the same genes, and the corresponding probe sets are presented in brackets in the table.

Table 5. Glutathione deficiency-related gene probe sets.

Affymetrix probe set ID		
RG U34A	RAE 230A	Gene name
rc_AI231807_at	1367559_at	Ferritin light chain
M29358_g_at	1367573_at	Ribosomal protein S6
J02752_at	1367680_at	Acyl-Coenzyme A oxidase 1
X78848cds_f_at	1367774_at	Glutathione *S*-transferase A5
M64733mRNA_s_at	1367784_a_at	Clusterin
AF026554_at	1367815_at	Solute carrier family 5, member 6
rc_AA892821_at	1367843_at	Aldo-keto reductase family 7, member A2
Y17295cds_s_at	1367969_at	Peroxiredoxin 6
U04733_s_at	1367988_at	CYP2C23
M57428_s_at	1368116_a_at	Ribosomal protein S6 kinase, polypeptide 1
AF045464_s_at	1368121_at	Aldo-keto reductase family 7, member A3
K00136mRNA_at	1368180_s_at	LOC494499 protein
AF025670_g_at	1368305_at	Caspase 6
D86745cds_s_at	1368376_at	Nuclear receptor subfamily 0, group B, member 2
U06274_s_at	1368397_at	UDP glycosyltransferase 2 family, polypeptide B4
AF087943_s_at	1368490_at	CD14 antigen
U73174_at	1369061_at	Glutathione reductase
AF068202_at	1369069_at	A kinase (PRKA) anchor protein 1
D13122_f_at	1369588_a_at	ATPase inhibitor
L11007_at	1369950_at	Cyclin-dependent kinase 4
rc_AI233261_i_at	1370030_at	Glutamate cysteine ligase, modifier subunit
rc_AI171506_at	1370067_at	Malic enzyme 1
D16478_at	1370164_at	Hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydratase (trifunctional protein), alpha subunit
U33500_g_at	1370566_at	retinol dehydrogenase 2
J05132_s_at	1370613_s_at	UDP glycosyltransferase 1 family, polypeptide A1
J05181_at	1370688_at	Glutamate-cysteine ligase, catalytic subunit
M13506_at	1370698_at	Liver UDP-glucuronosyltransferase, phenobarbital-inducible form
AJ001517cds_at	1370772_a_at	Hemochromatosis
S72506_s_at	1371089_at	Glutathione *S*-transferase Yc2 subunit
M20629_s_at	1371100_at	Esterase 2
M11794cds#2_f_at	1371237_a_at	Metallothionein
S70011_g_at	1372715_at	Sideroflexin 1 (predicted)
S55305_s_at	1386866_at	Tyrosine 3-monooxgenase/tryptophan 5-monooxgenase activation protein, gamma polypeptide
L07736_at	1386946_at	Carnitine palmitoyltransferase 1, liver
X04229cds_s_at	1386985_at	Glutathione *S*-transferase, mu 1
M33329_f_at	1387006_at	Rat senescence marker protein 2A gene, exons 1 and 2
E03424cds_s_at	1387221_at	GTP cyclohydrolase 1
M30282_at	1387323_at	Kallikrein B, plasma 1
M95762_at	1387372_at	Solute carrier family 6, member 13
J02679_s_at	1387599_a_at	NAD(P)H dehydrogenase, quinone 1
M26125_at	1387669_a_at	Epoxide hydrolase 1
X55286_g_at	1387848_at	3-Hydroxy-3-methylglutaryl-Coenzyme A reductase
rc_AI072634_at	1388155_at	Keratin complex 1, acidic, gene 18
D17309_at	1398310_at	Aldo-keto reductase family 1, member D1

Table 5. Continued.

rc_AI012807_at	1398765_at	Adaptor-related protein complex 2, mu 1 subunit
S56936_s_at	none	–
S82820mRNA_s_at	none	–
M33747_at	none	–
rc_AI638982_at	none	–
M33746mRNA#2_f_at	none	–
D00362_s_at	none	–
AF044574_at	none	–
S80431_s_at	none	–
rc_AI236284_s_at	none	–
M29249cds_at	none	–
D86745exon_s_at	none	–
X57133mRNA_at	none	–
X74565cds_g_at	none	–
U05784_s_at	none	–
AF087944mRNA_s_at	none	–
X77934cds_at	none	–
rc_AA859519_g_at	none	–
X57999cds_at	none	–
rc_AI231807_g_at	(1367559_at)	–
K01932_f_at	(1367774_at)	–
S72505_f_at	(1367774_at)	–
D00569_at	(1367777_at)	–
D86580_at	(1368376_at)	–
S65555_g_at	(1370030_at)	–

Probe sets whose signals were reported as increased by treatment of glutathione depletors in RG U34A GeneChip analysis were selected from the previous literature, and the corresponding probe sets of RAE 230A GeneChip were determined by referring to "good match probe sets" information provided by Affymetrix. Eighteen probe sets of RG U34A GeneChip did not show any corresponding probe sets in RAE 230A GeneChip, and were presented as "none" in the table. Six probe sets of RG U34A GeneChip had redundant probe sets for the same genes, and the corresponding probe sets are presented in brackets in the table.

was determined by logarithmic conversion of the fold change, where the base was set to 2. The heat map was created using Spotfire software (Spotfire, Inc.). The CYP1A1 mRNA level was determined by referring to signal data measured by 1370269_at gene probe set in GeneChip analysis.

RESULTS

TGP1 score

The TGP1 scores for carcinogenesis-related genes, PPARα-regulated genes and glutathione deficiency-related genes were calculated. For carcinogenesis-related genes, OPZ, CPZ, HCB, SS and WY showed a high TGP1 score among the compounds in the database (Fig. 2 A). For PPARα-regulated genes, WY, CFB, GFZ, BBr and ASA showed high TGP1 score (Fig. 2 B). For glutathione deficiency-related genes, OPZ, BBz, APAP and CMA showed high TGP1 score (Fig. 2 C).

In order to see the dose-dependency of the score, we selected several samples and showed them in Fig. 3A (carcinogenesis-related genes), 3B (PPARα-regulated genes), and 3C (glutathione deficiency-related genes) . It is obvious that the score showed good dose-dependency for all three of the marker gene sets. It is also interesting that some drugs show a sudden and large rise of the score suggesting a threshold of the toxicity at a certain dose level.

Comparison of TGP1 score between rat liver and rat hepatocytes

The TGP1 scores of rat liver and rat hepatocytes were calculated. For carcinogenesis-related genes, the

$$\text{Signal log ratio } (i) = \log_2\left\{\text{Average Signal } i \text{ (treated)} / \text{Average Signal } i \text{ (control)}\right\}$$

$$\text{Index 1} = \frac{\sum_{i=1}^{N} \text{Signal log ratio } (i)}{\text{Number of probe sets}}$$

$$\text{Index 2} = \frac{\sum_{i-1}^{N} \left\{\text{Signal log ratio } (i)\right\}^2}{\text{Number of probe sets}}$$

$$\text{TGP score 1} = (\text{Index 1}) \times (\text{Index 2})$$

Fig. 1. Calculation of TGP1 score.
The signal log ratio was calculated by dividing the mean signal value of the chemical-treated group by that of corresponding control. First, the sum of the signal log ratios for the used probe sets was calculated, and then divided by the number of probe sets used (Index 1). Next, the sum of squared signal log ratios for the used probe sets was calculated, and then divided by the number of probe sets used (Index 2). Finally, the TGP1 score was calculated by multiplying Index 1 with Index 2.

TGP1 scores in rat hepatocyte distributed from -10 to 4, while that of five chemicals, namely OPZ, CPZ, HCB and SS in rat livers showed quite high values exceeding 5 (Fig. 4 A). For PPARα-regulated genes, WY, CFB and BBr showed large TGP1 scores in both rat liver and rat hepatocyte (Fig. 4 B), whereas GFZ and ASA showed large TGP1 scores only in rat liver. For glutathione deficiency-related genes, OPZ, BBz, APAP and CMA showed large TGP1 scores only in rat liver (Fig. 4 C), whereas ANIT, INAH, ET and VPA showed small and negative TGP1 scores only in rat hepatocyte.

Gene expression profile in rat liver and rat hepatocytes

The expression profile of individual genes for carcinogenesis-related gene is shown in Fig. 5. It appears that variations among triplicates (liver *in vivo*) or duplicates (hepatocytes *in vitro*) are relatively small. Although the expression of each marker gene varied among the compounds, trends of gene expression as a whole are represented by the TGP1 scores. As it appeared that the contribution of induction of CYP1A1

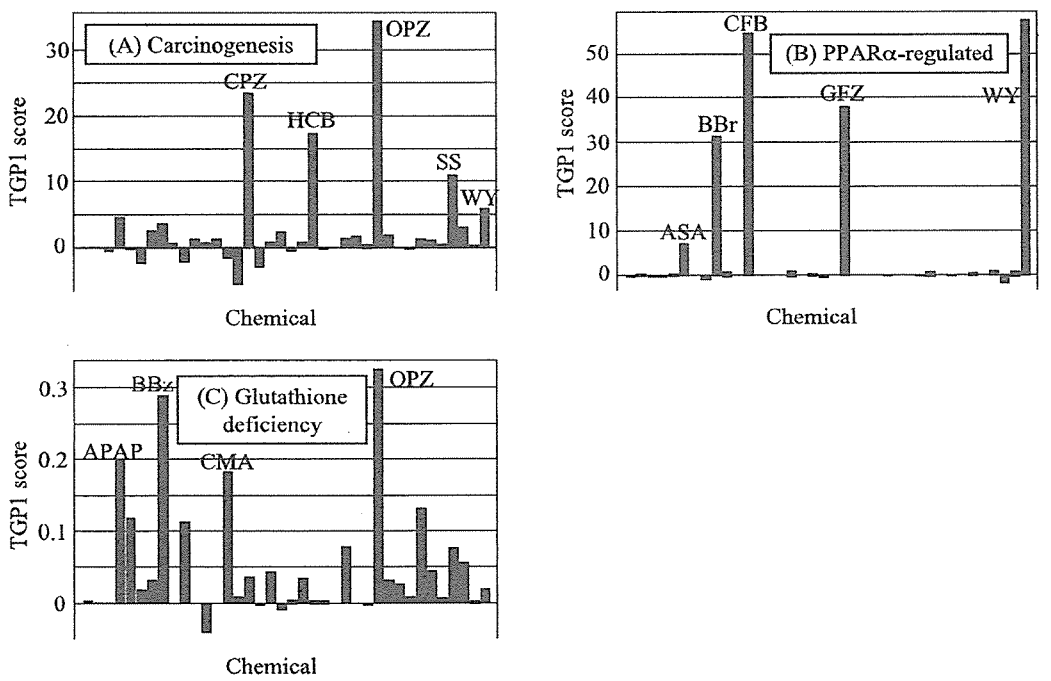


Fig. 2. TGP1 score.
TGP1 scores for (A) carcinogenesis-related genes, (B) PPARα-regulated genes and (C) glutathione deficiency-related genes were calculated for using the TGP database. The data of high dose were used for each chemical. The 38 chemicals are aligned on the abscissa in the order shown in Table 1 and the names are omitted for simplicity. As for the chemicals with high score, their names are indicated by abbreviations in the figure.

mRNA was the largest in the carcinogenesis-related genes, we focused on this gene (Fig. 6). Approximately a 16- to 500-fold increase of CYP1A1 mRNA was observed in OPZ, CPZ, HCB, SS-treated rat livers, whereas that in hepatocytes was only 2- to 8-fold. ANIT and DFNa showed small but consistent changes both *in vivo* and *in vitro*, and APAP and VPA showed no increase (or rather decrease) both *in vivo* and *in vitro*.

Fig. 7 shows the individual expression of glutathione deficiency-related genes *in vivo* and *in vitro*. Again, the variations among triplicates (liver *in vivo*) or duplicates (hepatocytes *in vitro*) appeared to be small, and the trends of gene expression as a whole were well represented by the TGP1 scores.

DISCUSSION

The objective of this study was to establish a one-dimensional score, which reflects the level of gene expression changes in certain biomarker gene sets. Such a score would be useful particularly when a large-scale toxicogenomics database and multiple biomarker information are available simultaneously. In the present study, we examined the usefulness of the score, named as TGP1, based on the signal log ratio values of the chemical-treated to the vehicle-treated rats.

As shown in Fig. 1, the TGP1 score comprised two elements, Index 1 and Index 2. Index 1 was set to capture the general tendency of gene expression changes; the absolute value would be high when the direction of the change is concomitant, whereas it would be low when they are random. In the present study, we selected three biomarker gene sets in which their expression is supposed to be uniformly up-regulated by treatment. The next element, Index 2, was set to capture the general size of gene expression changes, not considering their change in direction toward up- or

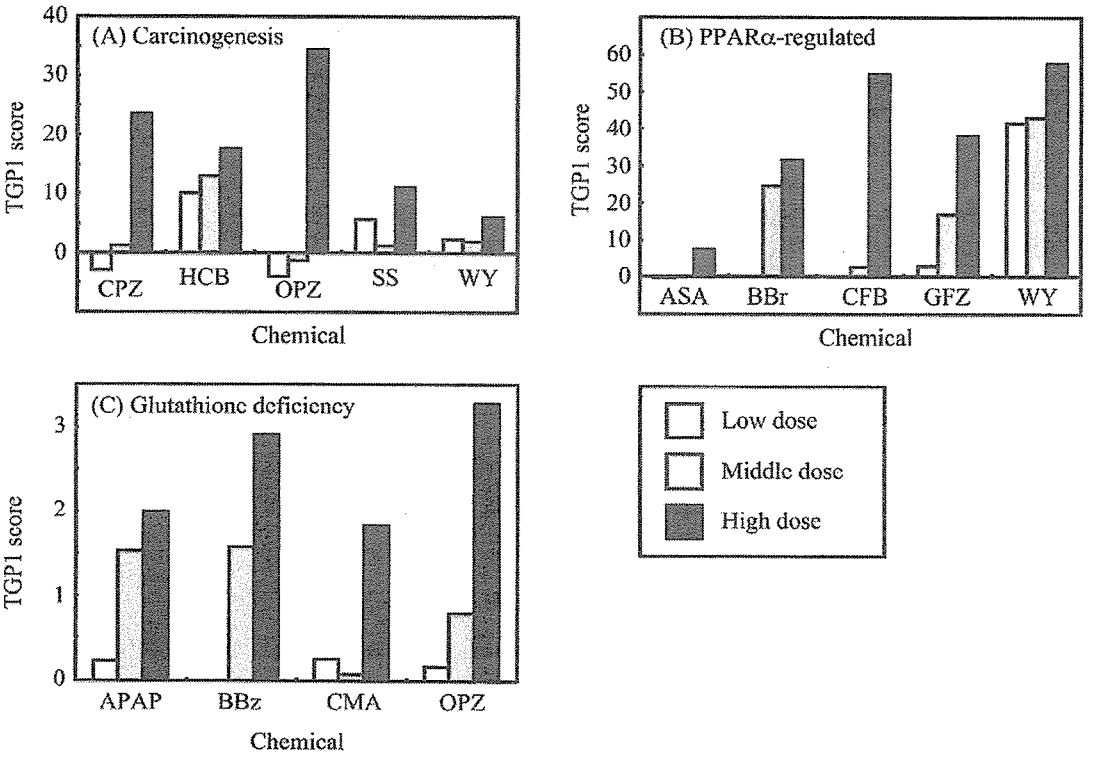


Fig. 3. Dose-dependency of TGP1 score.
The chemicals showing high TGP1 score (indicated in Figure 2) at high dose were selected and TGP1 scores for (A) carcinogenesis-related genes, (B) PPARα-regulated genes and (C) glutathione deficiency-related genes were calculated for low and middle doses in order to show dose-dependency. The abbreviations of the chemicals are shown in Table 1.

down-regulation by the chemical treatments. Since Index 2 includes the square of the signal log ratio, it tends to be higher when a biomarker gene set includes genes that show large changes in their mRNA levels. To examine the usefulness of the TGP1 score, three biomarker gene sets were selected from the previous literature whose expression levels are up-regulated by carcinogens (Table 3), PPARα activators (Table 4) and glutathione depletors (Table 5).

For carcinogenesis-related genes, the TGP1 score was high in rats treated with OPZ, CPZ, HCB, SS and WY (Fig. 2 A). OPZ is reported to have weak genotoxicity (Martelli *et al.*, 1998). CPZ and WY are reported to be non-genotoxic carcinogens in rodent (Gocke, 1996; Peters *et al.*, 1997). HCB and SS are reported to have carcinogenic potential in rodents (Smith *et al.*, 1985; Iatropoulos *et al.*, 1997). Thus, the TGP1 score appears to well express the carcinogenic risk of the chemicals. For PPARα-regulated genes, the TGP1 score was high in rats treated with CFB, WY, GFZ, BBr and ASA (Fig. 2 B). Of these, CFB, WY and GFZ are known to be potent PPARα activators (Gonzalez *et al.*, 1998). Besides these direct PPARα activators, it was reported that BBr has peroxisome proliferator activity (Bichet *et al.*, 1990). In addition, several non-steroidal anti-inflammatory drugs are reported to activate PPARα (Lehmann *et al.*, 1997), and therefore a high TGP1 score observed in ASA-treated rat liver would be reasonable. These results indicate that the TGP1 score works to express the PPARα-modulating property of the chemicals. For glutathione deficiency-related genes, the TGP1 score was high in rats treated with OPZ, BBz, APAP and CMA (Fig. 2 C). Of these, BBz, APAP and CMA are reported to deplete the hepatic glutathione content when overdosed (Lake *et al.*, 1989; Szymanska, 1996; James *et al.*, 2003). Although OPZ has not been reported to deplete hepatic glutathione in rats, it could be highly possible at extremely high dose, since OPZ attacks SH of gastric proton pump at low pH and this effect is prevented by

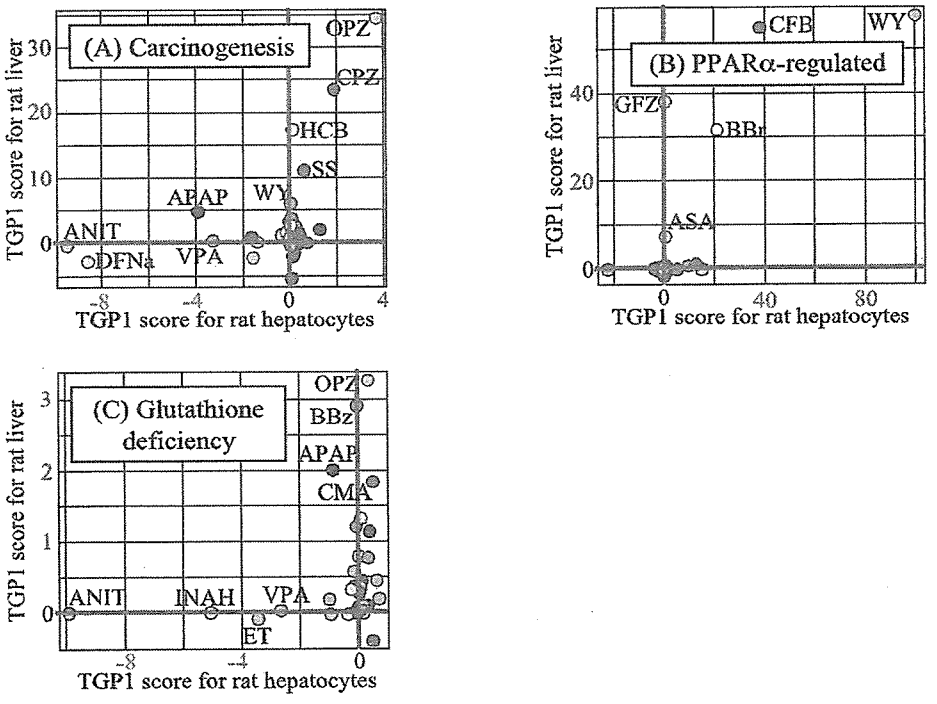


Fig. 4. Comparison of TGP1 scores between rat liver and rat hepatocytes.
Scatter plots of TGP1 scores for rat liver and rat hepatocytes are presented using (A) carcinogenesis-related genes, (B) PPARα-regulated genes and (C) glutathione deficiency-related genes.

glutathione (Morii *et al.*, 1989), and OPZ-glutathione conjugates are found in rat serum as metabolites (Weidolf *et al.*, 1992). Nonetheless, the results obtained from the three biomarker gene sets demonstrated that the TGP1 score successfully captured the characteristics of the compounds stored in our database.

In the next step, we compared the TGP1 scores between rat liver and hepatocytes. Since the TGP1 score is a one-dimensional value, we can make the comparison by plotting on X-Y coordinates. For carcinogenesis-related genes, TGP1 scores for some compounds including OPZ and CPZ showed positive correlation between rat liver and rat hepatocytes (Fig. 4 A). However, absolute values of TGP1 scores for hepatocytes were much smaller than those for liver, indicating that the genes included in the gene sets did not respond in the same way to the compounds between *in vivo* and *in vitro*. Similarly, the TGP1 score for glutathione deficiency-related genes did not show any correlation between liver and hepatocytes (Fig. 4 C), indicating that these gene sets can only be applied to data analysis of rat liver, and not to rat hepatocytes. ForPPARα-regulated genes, however, the scores for WY, CFB and BBr showed good positive correlation between rat liver and rat hepatocytes (Fig. 4 B). In contrast to the case of carcinogenesis-related genes, the absolute values of TGP1 scores were similar for *in vivo* and *in vitro*. It seems these gene sets could be used for assessing PPARα-modulating effects of chemicals in rat liver by using rat hepatocytes as an alternative experimental system. One of the premises of toxicogenomics research is to scale down toxicity studies from *in vivo* to an *in vitro* system, thereby reducing the number of experimental animals and the amount of test substances used (Boess *et al.*, 2003). It is expected that the TGP1 score enables us to suggest certain toxicological endpoints *in vivo*, which can be evaluated by *in vitro* systems using a large-scale toxicogenomics database and specific biomarker gene sets, such as PPARα-reg-

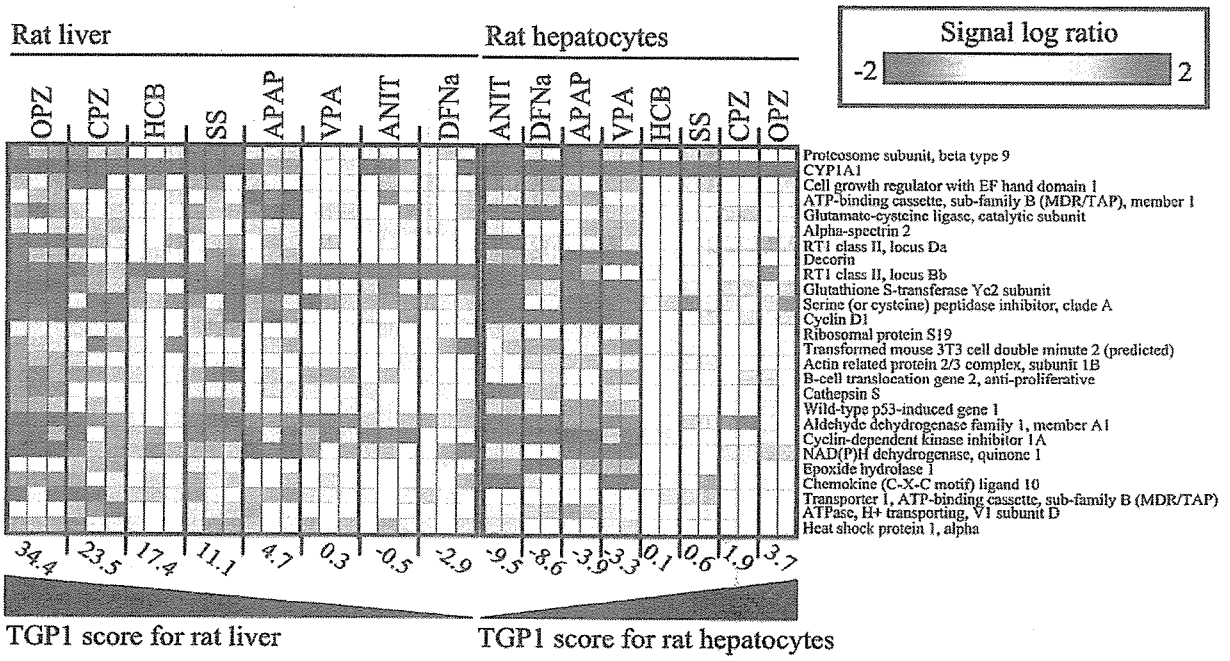


Fig. 5. Gene expression profile of carcinogenesis-related genes.
The fold change value for each of the carcinogenesis-related genes was calculated by dividing the signal value of chemical-treated rats or rat hepatocytes by the mean signal value of corresponding vehicle-treated rats (n=3) or rat hepatocytes (n=2), respectively, and the fold change values were converted to logarithm values where the base was set to 2. The heat map representing individual expression levels of carcinogenesis-related genes was created using the logarithm values of fold changes.

ulated genes.

The TGP1 score is useful for quickly surveying the gene expression changes in biomarker gene sets for multiple samples. However, the significance of changes in gene expression should be investigated in detail for the next step of data analysis. In the present study, the compounds whose TGP1 scores were relatively high or low in either rat liver or rat hepatocytes were selected, and expression profiles of carcinogenesis-related genes and glutathione deficiency-related genes are presented in Fig. 5 and Fig. 7, respectively. For carcinogenesis-related gene sets, CYP1A1 appeared to be the most remarkable gene that strongly affects the TGP1 score. The importance of the CYP1A1 gene in TGP1 score calculation for carcinogenesis-related gene sets is evident in Fig. 6 as well. CYP1A1 is under control of arylhydrocarbon receptor, a transcription factor that has been reported to play a key role in liver tumor promotion (Bock and Kohle, 2005). Therefore, the biological significance suggested by a high TGP1 score using carcinogenesis-related genes was thought to be appropriate for evaluating the carcinogenic risk of chemicals. On the other hand, genes that strongly affect the TGP1 score for glutathione deficiency-related genes were found to be aldo-keto reductase family 7, glutathione S-transferase Yc2 subunit, metallothionein or NAD(P)H dehydrogenase (Fig. 7). It is crucial to determine and investigate such key genes that strongly affect the TGP1 score to elucidate the molecular mechanism of toxicity.

One of the most promising goals of toxicogenomics application in drug development is an improved quality of safety assessment in human cases from experimental animal data, and the TGP database is constructed to achieve this goal by collecting both rat and human microarray data. When inter-species data analysis is challenged, however, a probe set conversion process from rat to human-like ortholog conversion is required. However, such converted-probe sets usually contain low-quality data for measuring gene expression signal in human samples. In such cases, straightforward data analysis would lead to the wrong conclusion. Therefore, the TGP1 score is thought to be inadequate to compare samples across the species. Additional scores that overcome this drawback should be needed for efficient data analysis of large-scale microarray database such as our TGP database.

In conclusion, we tested a simple one-dimensional score, named as TGP1, that reflects the level of gene expression changes in certain biomarker gene sets. The usefulness of this score was demonstrated using the TGP database. The score is useful for surveying the expression changes in multiple biomarker gene sets using the large-scale toxicogenomics database, which can reduce the labor- and time-consuming task of conventional multivariate statistical analysis. This scoring system is now incorporated in our analysis system where TGP1 scores are automatically calculated and displayed. Additional sophisticated one-dimensional scores should be invented to facilitate gene expression analysis for a large-scale toxicogenomics database with multiple biomarker gene sets.

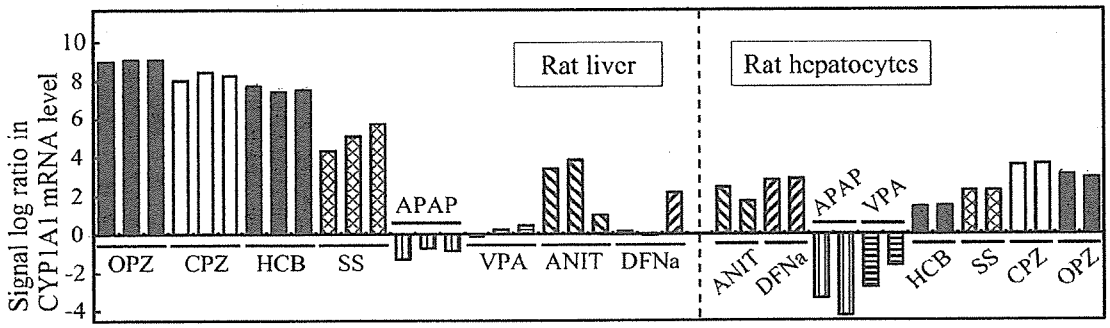


Fig. 6. Change in CYP1A1 mRNA level.
The fold change value for the CYP1A1 gene determined by 1370269_at probe set in GeneChip analysis was calculated by dividing the signal value of chemical-treated rats or rat hepatocytes by the mean signal value of corresponding vehicle-treated rats (n=3) or rat hepatocytes (n=2), respectively, and the fold change values were converted to logarithm values where the base was set to 2.

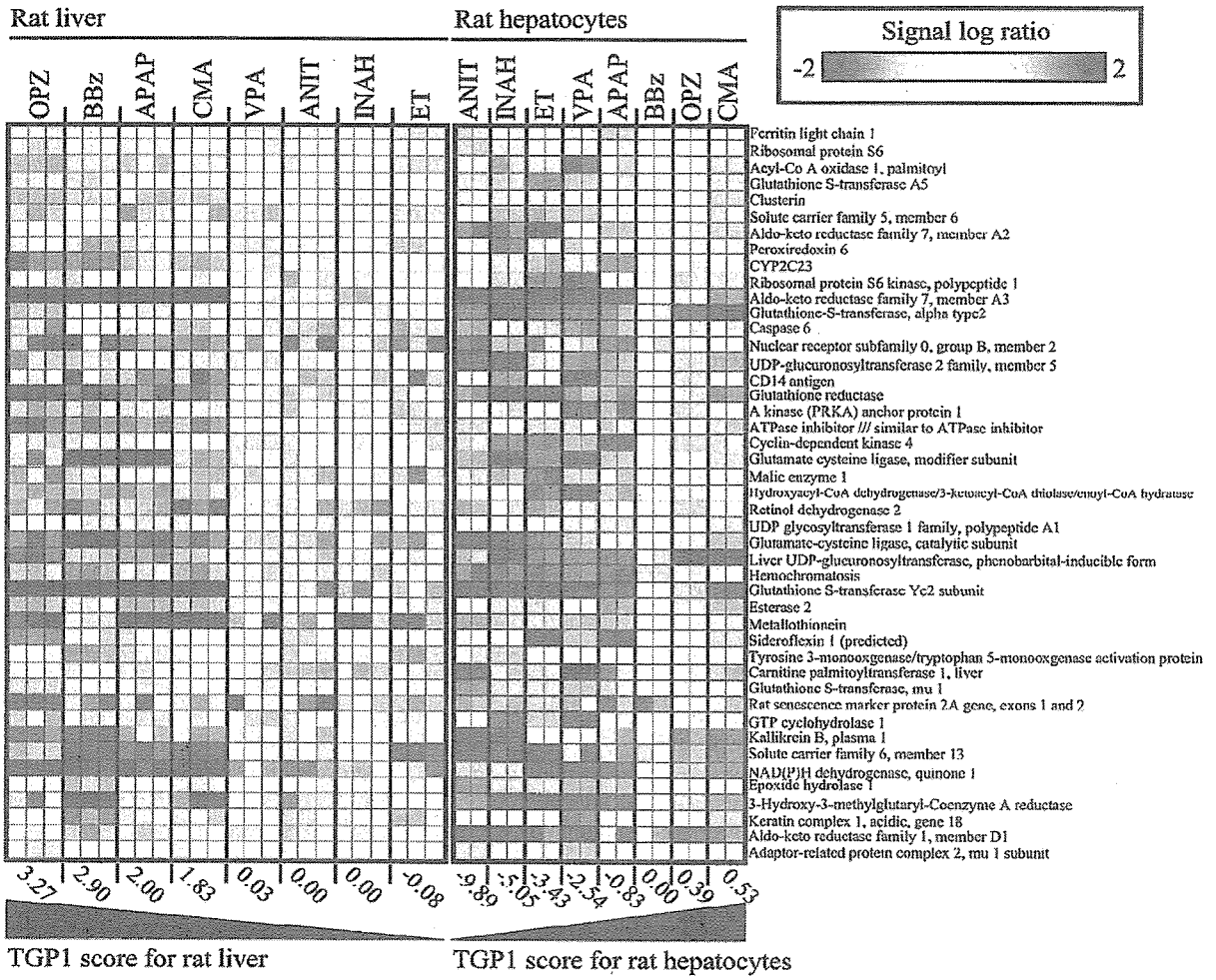


Fig. 7. Gene expression profile of glutathione deficiency-related genes.
The fold change value for each of the glutathione deficiency-related genes was calculated by dividing the signal value of chemical-treated rats or rat hepatocytes by the mean signal value of corresponding vehicle-treated rats (n=3) or rat hepatocytes (n=2), respectively, and the fold change values were converted to logarithm values where the base was set to 2. The heat map representing individual expression levels of glutathione deficiency-related genes was created using the logarithm values of fold changes.

ACKNOWLEDGMENT

This study was supported in part by a grant from the Ministry of Health, Labor and Welfare (H14-Toxico-001).

REFERENCES

Bichet, N., Cahard, D., Fabre, G., Remandet, B., Gouy, D. and Cano, J.P. (1990): Toxicological studies on a benzofuran derivative. III. Comparison of peroxisome proliferation in rat and human hepatocytes in primary culture. Toxicol. Appl. Pharmacol., **106**, 509-517.

Bock, K.W. and Kohle, C. (2005): Ah receptor- and TCDD-mediated liver tumor promotion: clonal selection and expansion of cells evading growth arrest and apoptosis. Biochem. Pharmacol., **69**, 1403-1408.

Boess, F., Kamber, M., Romer, S., Gasser, R., Muller, D., Albertini, S. and Suter, L. (2003): Gene expression in two hepatic cell lines, cultured primary hepatocytes, and liver slices compared to the *in vivo* liver gene expression in rats: Possible implications for toxicogenomics use of *in vitro* systems. Toxicol. Sci., **73**, 386-402.

Boverhof, D.R. and Zacharewski, T.R. (2006): Toxicogenomics in risk assessment: Applications and needs. Toxicol. Sci., **89**, 352-360.

Draghici, S. (2003): Analysis and visualization tools. In Data Analysis Tools for DNA Microarrays (Etheridge, A.M., Gross, L.J., Lenhart, S., Maini, P. K., Safer, H. M. and Voit, E. O., eds.), pp. 231-261. CRC Press, London.

Ellinger-Ziegelbauer, H., Stuart, B., Wahle, B., Bomann, W. and Ahr, H.J. (2004): Characteristic expression profiles induced by genotoxic carcinogens in rat liver. Toxicol. Sci., **77**, 19-34.

Gocke, E. (1996): Review of the genotoxic properties of chlorpromazine and related phenothiazines. Mutat. Res., **366**, 9-21.

Gonzalez, F.J., Peters, J.M. and Cattley, R.C. (1998): Mechanism of action of the nongenotoxic peroxisome proliferators: Role of the peroxisome proliferator-activator receptor alpha. J. Natl. Cancer Inst., **90**, 1702-1709.

Iatropoulos, M.J., Williams, G.M., Abdo, K.M., Kari, F.W. and Hart, R.W. (1997): Mechanistic studies on genotoxicity and carcinogenicity of salicylazosulfapyridine an anti-inflammatory medicine. Exp. Toxicol. Pathol., **49**, 15-28.

Ito, K., Kiyosawa, N., Kumagai, K., Manabe, S., Matsunuma, N. and Yamoto, T. (2006): Molecular mechanism investigation of cycloheximide-induced hepatocyte apoptosis in rat livers by morphological and microarray analysis. Toxicology, **219**, 175-186.

James, L.P., Mayeux, P.R. and Hinson, J.A. (2003): Acetaminophen-induced hepatotoxicity. Drug Metab. Dispos., **31**, 1499-1506.

Kaminski, N. and Friedman, N. (2002): Practical approaches to analyzing results of microarray experiments. Am. J. Respir. Cell Mol. Biol., **27**, 125-132.

Kiyosawa, N., Ito, K., Sakuma, K., Niino, N., Kanbori, M., Yamoto, T., Manabe, S. and Matsunuma, N. (2004a): Evaluation of glutathione deficiency in rat livers by microarray analysis. Biochem. Pharmacol., **68**, 1465-1475.

Kiyosawa, N., Tanaka, K., Hirao, J., Ito, K., Niino, N., Sakuma, K., Kanbori, M., Yamoto, T., Manabe, S. and Matsunuma, N. (2004b): Molecular mechanism investigation of phenobarbital-induced serum cholesterol elevation in rat livers by microarray analysis. Arch. Toxicol., **78**, 435-442.

Lake, B.G., Gray, T. J., Evans, J.G., Lewis, D.F., Beamand, J.A. and Hue, K.L. (1989): Studies on the mechanism of coumarin-induced toxicity in rat hepatocytes: Comparison with dihydrocoumarin and other coumarin metabolites. Toxicol. Appl. Pharmacol., **97**, 311-323.

Lehmann, J.M., Lenhard, J.M., Oliver, B.B., Ringold, G.M. and Kliewer, S.A. (1997): Peroxisome proliferator-activated receptors alpha and gamma are activated by indomethacin and other non-steroidal anti-inflammatory drugs. J. Biol. Chem., **272**, 3406-3410.

Liu, G., Loraine, A.E., Shigeta, R., Cline, M., Cheng, J., Valmeekam, V., Sun, S., Kulp, D. and Siani-Rose, M.A. (2003): NetAffx: Affymetrix probesets and annotations. Nucleic Acids Res., **31**, 82-86.

Martelli, A., Mattioli, F., Mereto, E., Brambilla Campart, G., Sini, D., Bergamaschi, G. and Brambilla, G. (1998): Evaluation of omeprazole genotoxicity in a battery of *in vitro* and *in vivo* assays. Toxicology, **130**, 29-41.

Morii, M., Takata, H. and Takeguchi, N. (1989): Acid activation of omeprazole in isolated gastric vesicles, oxyntic cells, and gastric glands. Gastroenterol., **96**, 1453-1461.

Peters, J.M., Cattley, R.C. and Gonzalez, F.J. (1997): Role of PPAR alpha in the mechanism of action of the nongenotoxic carcinogen and peroxisome proliferator Wy-14,643. Carcinogenesis, **18**, 2029-2033.

Richert, L., Lamboley, C., Viollon-Abadie, C., Grass, P., Hartmann, N., Laurent, S., Heyd, B., Mantion, G., Chibout, S.D. and Staedtler, F. (2003): Effects of clofibric acid on mRNA expression profiles in primary cultures of rat, mouse and human hepatocytes. Toxicol. Appl. Pharmacol., **191**, 130-146.

Sawada, H., Takami, K. and Asahi, S. (2005): A toxicogenomic approach to drug-Induced phospho-

lipidosis: Analysis of its induction mechanism and establishment of a novel *in vitro* screening system. Toxicol. Sci., **83**, 282-292.

Sehata, S., Kiyosawa, N., Makino, T., Atsumi, F., Ito, K., Yamoto, T., Teranishi, M., Baba, Y., Uetsuka, K., Nakayama, H. and Doi, K. (2004): Morphological and microarray analysis of T-2 toxin-induced rat fetal brain lesion. Food Chem. Toxicol., **42**, 1727-1736.

Smith, A.G., Francis, J.E., Dinsdale, D., Manson, M.M. and Cabral, J.R. (1985): Hepatocarcinogenicity of hexachlorobenzene in rats and the sex difference in hepatic iron status and development of porphyria. Carcinogenesis, **6**, 631-636.

Szymanska, J.A. (1996): Biochemical alterations as measures of acute and subacute hepatotoxicity of 1,3-dibromobenzene in rat. Arch. Toxicol., **71**, 99-106.

Takashima, K., Mizukawa, Y., Morishita, K., Okuyama, M., Kasahara, T., Toritsuka, N., Miyagishima, T., Nagao, T. and Urushidani, T. (2006): Effect of the difference in vehicles on gene expression in the rat liver-analysis of the control data in the Toxicogenomics Project Database. Life Sci., **78**, 2787-2796.

Urushidani, T. and Nagao, T. (2005): Toxicogenomics: The Japanese initiative. In Handbook of Toxicogenomics - Strategies and Applications (Borlak, J., ed.), pp. 623-631. Wiley - VCH.

Weidolf, L., Karlsson, K.E. and Nilsson, I. (1992): A metabolic route of omeprazole involving conjugation with glutathione identified in the rat. Drug Metab. Dispos., **20**, 262-267.

The Journal of Toxicological Sciences,
Vol.31, No.5, 449-469, 2006

COMPARISON OF GENE EXPRESSION PROFILES AMONG PAPILLA, MEDULLA AND CORTEX IN RAT KIDNEY

Kotaro TAMURA[1], Atsushi ONO[1], Toshikazu MIYAGISHIMA[1], Taku NAGAO[2] and Tetsuro URUSHIDANI[1,3]

[1]*Toxicogenomics Project, National Institute of Biomedical Innovation, 7-6-8 Saito-Asagi, Ibaraki, Osaka 567-0085, Japan*

[2]*National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-Ku, Tokyo 158-8501, Japan*

[3]*Department of Pathophysiology, Faculty of Pharmaceutical Sciences, Doshisha Women's College of Liberal Arts, Kodo, Kyotanabe, Kyoto 610-0395, Japan*

(Received August 18, 2006; Accepted September 11, 2006)

ABSTRACT — The aim of this study was to compare gene expression profiles in the different kidney regions as the basis for toxicogenomics. Rat kidney was separated into papilla, medulla and cortex, and total RNA was isolated from these and from the whole slice. Gene expression profiling was performed using Affymetrix Rat Genome 230 2.0 Array. When global normalization was applied, the expression of β-actin or GAPDH varied among the regions. It was considered that such a comparison could not be made, especially between papilla and other portions, since the production of total mRNA in the former was relatively low. In fact, ANOVA was performed on the gene expression values with global normalization in papilla, medulla, cortex, and whole slice, and the numbers of genes appeared to be the highest in papilla. It was also observed that many genes showed their maximum or minimum in the whole slice, which was theoretically impossible. To overcome the problems associated with global normalization, the "percelome" normalization (a way to obtain the values directly related to the copies of mRNA per cell) was employed to compare the regions. In applying this procedure, probe sets with regional difference in expression were efficiently extracted by ANOVA. When they were sorted by the fold difference to other regions, the higher rank was occupied by genes characteristic of the functions of kidney, i.e., channels, transporters and metabolic enzymes. Some of them were consistent with the literature and were related to pathophysiological phenomena. Comprehensive comparison of data of gene expression in the renal anatomical area will greatly enhance studies of the physiological function and mechanism of toxicity in kidney.

KEY WORDS: Toxicogenomics, Kidney, Gene expression, Regional difference, Rat

INTRODUCTION

The Toxicogenomics Project is a 5-year collaborative project by the National Institute of Health Sciences (NIHS) and 17 pharmaceutical companies in Japan which was started in 2002 (Urushidani and Nagao, 2005). In April 2005, some rearrangements were made and now the project is conducted by NIHS, the National Institute of Biomedical Innovation, and 15 pharmaceutical companies. Its aim is to construct a large-scale toxicology database of transcriptome for prediction of toxicity of new chemical entities in the early stage of drug development. About 150 chemicals, mainly medicinal compounds, have been selected, and the following are examined for each. The *in vivo* test using rat consists of a single administration test (3, 6, 9 and 24 hr with 4 dose levels including vehicle control) as well as a repeated administration test (3, 7, 14 and 28 days with 4 dose levels including vehicle control), and then data of body weight, general symptoms, histopathological examination of liver and kidney, and blood biochemistry are obtained from each animal. Gene

Correspondence: Tetsuro URUSHIDANI (E-mail: turushid@dwc.doshisha.ac.jp)

expression in liver (and kidney in some cases) is comprehensively analyzed by using Affymetrix GeneChip. An *in vitro* test using rat and human hepatocytes is also carried out to accomplish the bridging between the species. Although the main target of the project is liver, about 10 of the chemicals are typical nephrotoxicants and many others exhibited nephrotoxicity in addition to hepatotoxicity. In our project, we plan to perform gene expression analysis of kidney for up to 30 such chemicals.

Although the toxicogenomics technique is established as a powerful tool for prediction of nephrotoxicity of drugs (Thukra *et al.*, 2005), it is well known that kidney consists of a variety of cell types and that the physiological functions, including gene expression, differ between the anatomical portions, i.e., papilla, medulla, and cortex. Therefore, we expected that different gene expression profiles would be obtained either when kidney is analyzed as a whole or separated into each portion. For an exploratory test, we checked potential region-related differences in gene expression before starting analysis of drug effects on the kidney.

In employing global normalization, based on the assumption that the total amount of mRNA is constant, it can cause a bias in the comparison of the different portions, since the rate of transcription varies with the cell types. In our project, gene expression values can be converted to a value proportional to the copies of mRNA per cell (the values normalized by externally adding standard mRNA in an amount proportional to the DNA content in the homogenate) by employing a system, "percellome" (Kanno *et al.*, 2006). In the present study, quantification by this system was compared to that of global normalization.

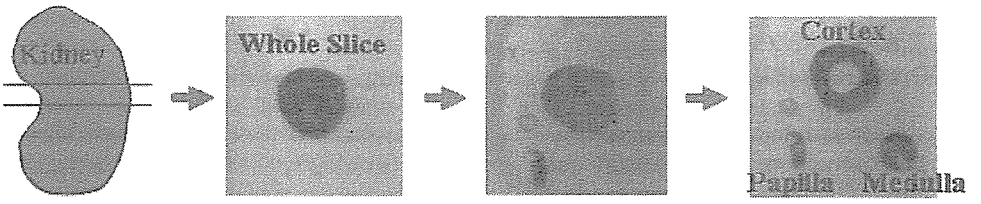
MATERIALS AND METHODS

Animals and Sampling

Male Sprague-Dawley rats were purchased from Charles River Japan Inc., (Kanagawa, Japan) at 5-weeks of age. After a 7-day quarantine and acclimatization period, 6 of the animals were euthanized by exsanguination from the abdominal veins and arteries under ether anesthesia. Kidneys were collected from each animal, and sliced horizontally at its middle portion with ca. 1 mm thickness by a pair of razor blades with a spacer in between. The slice was put into RNA later (Ambion, Austin, TX, USA) overnight for expression profiling. The fixed slices from three (No. 1 to 3) out of 6 animals were then separated into papilla, medulla, and cortex, as shown in Fig. 1. The remaining three (No. 4 to 6) were analyzed as a whole slice. The experimental protocols were reviewed and approved by the Ethics Review Committee for Animal Experimentation of the National Institute of Health Sciences.

Expression profiling

The kidney samples (whole slice, papilla, medulla and cortex) were homogenized using Mill Mixer (Qiagen) and zirconium beads. Total RNA was isolated from the kidney homogenate using RNeasy kit. Purity of the RNA was checked by gel electrophoresis, and the 260/280 nm ratio was between 2.0-2.2. Microarray analysis was conducted on 3 samples for each group by using GeneChip Rat Genome 230



Animal No. 1 ⇒ Whole slice No. 1
Animal No. 2 ⇒ Whole slice No. 2
Animal No. 3 ⇒ Whole slice No. 3
Animal No. 4 ⇒ Papilla No. 4, Medulla No. 4, Cortex No. 4
Animal No. 5 ⇒ Papilla No. 5, Medulla No. 5, Cortex No. 5
Animal No. 6 ⇒ Papilla No. 6, Medulla No. 6, Cortex No. 6

Fig. 1. Sampling and preparation of kidney for GeneChip analysis.

2.0 Arrays (Affymetrix, Santa Clara, CA, USA) containing 31,099 probe sets. The procedure was basically conducted according to the manufacturer's instructions using Superscript Choice System (Invitrogen, Carlsbad, CA, USA) and T7-(dT)24-oligonucleotide primer (Affymetrix) for cDNA synthesis, cDNA Cleanup Module (Affymetrix) for purification, and IVT Labeling Kit (Affymetrix) for synthesis of biotin-labeled cRNA. Twenty μg of the fragmented cRNA was hybridized to a Rat Genome 230 2.0 Array for 18 hr at 45°C at 60 rpm after which the array was washed and stained by streptavidin- phycoerythrin using Fluidics Station 450 (Affymetrix) and then scanned by Gene Array Scanner (Affymetrix).

Percellome normalization

We primarily use global mean normalization for data analysis in our project. However, to compare a sample from a different kidney region, in which the activity of transcription might be different, global normalization based on the total mRNA appeared not to provide the correct result. In the present study, we employed "percellome normalization" to normalize data as well.

The Percellome method reported by Kanno *et al.* (2006) was shown to be effective to compare the data from different tissues or different platform. In this method, the grade-dosed spike cocktail (GSC), which consists of five different Bacillus subtilis mRNA with different concentrations, was added to the tissue homogenate in proportion to its DNA contents before RNA extraction, assuming that all the cells contain a fixed amount of genomic DNA (g/cell) across the samples. The copy number of the RNA in one cell was calculated from the function of signal values of GSC and their copy number. Therefore, the effect of the difference in the total gene expression between samples can be avoided and a direct comparison of the expression data of samples from different sources is enabled.

The GSC used in this study was generously provided by Dr. Kanno, and the experimental procedure and calculations of the percellome method was conducted according to the original procedure. By separate experiments, we confirmed that GSC did not affect signals of other RNA and that the effect on global mean value was negligible.

Statistical Analysis

The data was analyzed by using GeneSpring version 6.1 (Silicon Genetics, Santa Clara, CA, USA), 2003 (Microsoft, Redmond, WA, USA) and MeV version 3.1 (The Institute for Genomic Research, Rockville, MD, USA). Expression data were customarily normalized using the mean value, multiplied by 500 (global mean normalization). Filtering of the data was performed either by flags (present, absent and marginal call) or ANOVA (Snedecor and Cochran, 1989).

RESULTS

First, analyses were conducted by our routine procedure, global normalization, in which each signal intensity was divided by the mean of each chip and multiplied by an arbitrary number to adjust the mean of each to the same value. In order to check the efficacy of the normalization procedure, the values of β-actin and GAPDH in each sample were divided by the mean value of the total 12 samples (Table 1). Expressed by this number, the values of β-actin varied between 0.91 - 0.99 in papilla, 0.98 - 1.12 in medulla, 0.97 - 1.14 in cortex, and 0.94 - 1.04 in the whole slice, resulting in an overall range of 0.91 - 1.14 (25% variation). The expression of GAPDH varied 0.76 - 0.85 in papilla, 1.01 - 1.05 in medulla, 1.11 - 1.19 in cortex, and 1.02 - 1.04 in the whole slice, resulting in an overall range of 0.76 - 1.19 (50% variation). Although they are relatively close to 1 (the mean value was 0.94, 1.02, and 1.03 for β-actin and 0.81, 1.03 and 1.16 for GAPDH in papilla, medulla, and cortex, respectively), the differences seemed to be too large to be ignored.

Using the data analyzed with global normalization, Pearson's correlation was calculated for each pair of the samples and is shown in Table 2A. The correlation coefficients were quite high among 3 samples from the same region, i.e., all were larger than 0.985. The correlation between papilla and others were lower than that between medulla and cortex, i.e., 0.821 - 0.860 between papilla and medulla, 0.741 - 0.789 between papilla and cortex, whereas it was 0.936 - 0.955 between medulla and cortex. The correlation coefficients between each of the three regions and the whole slice were in this order: cortex (0.981 - 0.987) > medulla (0.954 - 0971) > papilla (0.770 - 0.814), suggesting that gene expression in the whole slice preferentially represents that in cortex but it does not well represent that in papilla.

To try to identify genes with differential expression in the three regions, we first used a parameter of GeneChip data, detection call (present, marginal, and absent). Probe sets with present call in all samples in a region but all absent in other regions were selected as region-specific genes (Fig. 2). Genes with present call