

Scoring the level of gene expression changes in microarray analysis.

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

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

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

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

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 PPAR α -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

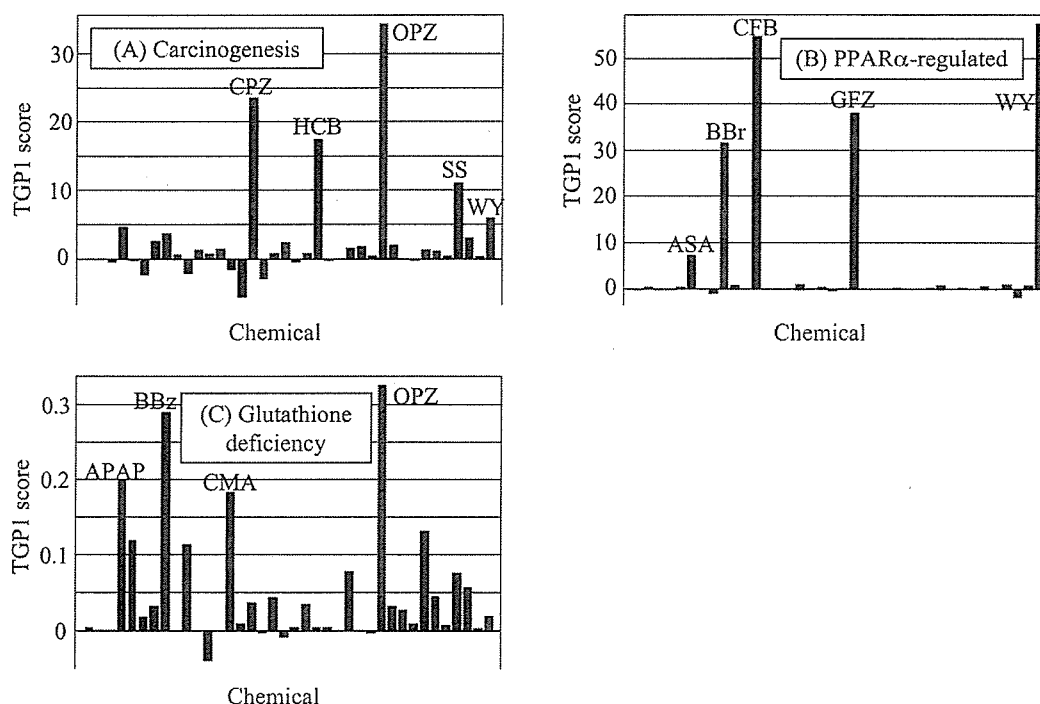


Fig. 2. TGP1 score.

TGP1 scores for (A) carcinogenesis-related genes, (B) PPAR α -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 TGPI 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 TGPI, based on the signal log ratio values of the chemical-treated to the vehicle-treated rats.

As shown in Fig. 1, the TGPI 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-

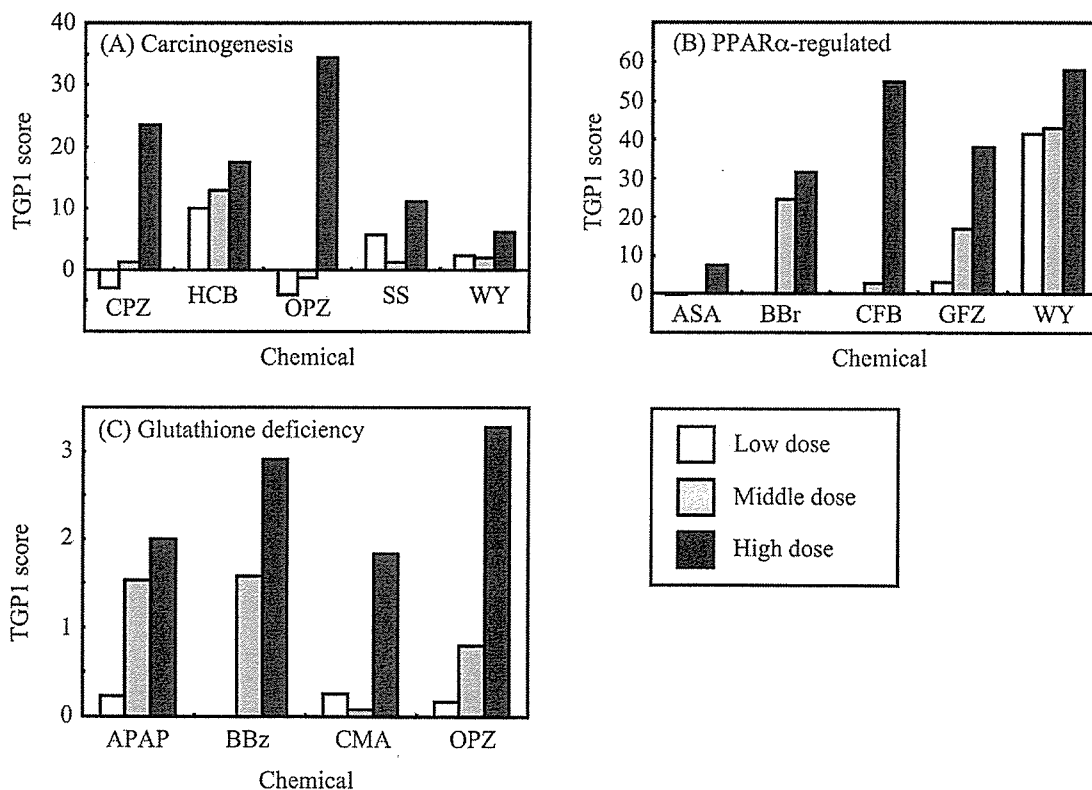


Fig. 3. Dose-dependency of TGPI score.

The chemicals showing high TGPI score (indicated in Figure 2) at high dose were selected and TGPI scores for (A) carcinogenesis-related genes, (B) PPAR α -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.

Scoring the level of gene expression changes in microarray analysis.

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), PPAR α 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 (Goetze, 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 PPAR α -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 PPAR α activators (Gonzalez *et al.*, 1998). Besides these direct PPAR α 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 PPAR α (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 PPAR α -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

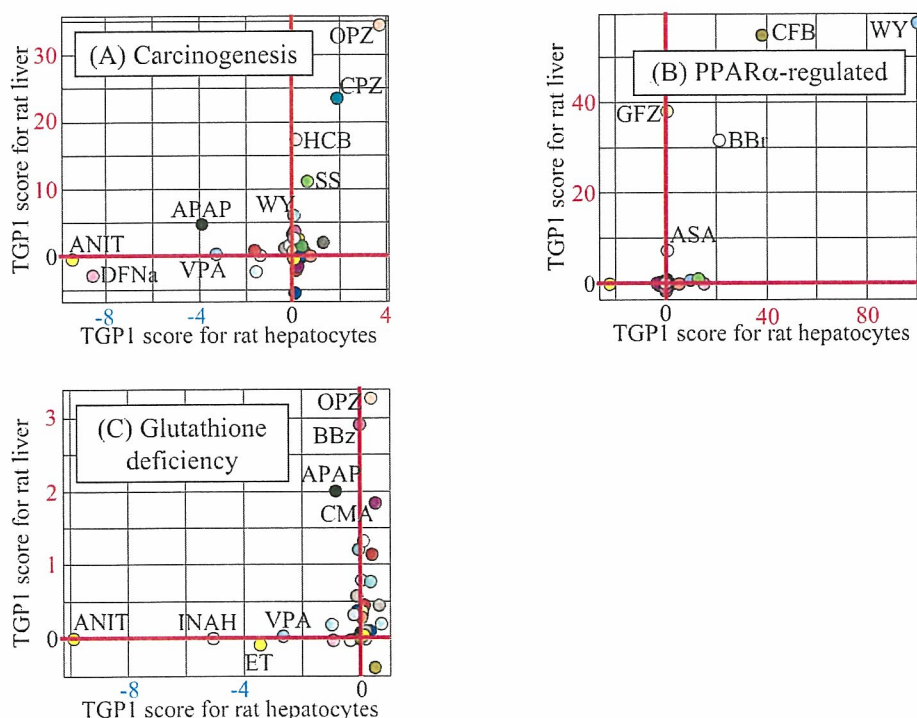


Fig. 4. Comparison of TGP1 scores between rat liver and rat hepatocytes. Scatter plots of TGP1 scores for rat liver and rat hepatocytes are presented using (A) carcinogenesis-related genes, (B) PPAR α -regulated genes and (C) glutathione deficiency-related genes.

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. For PPAR α -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 PPAR α -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 PPAR α -reg-

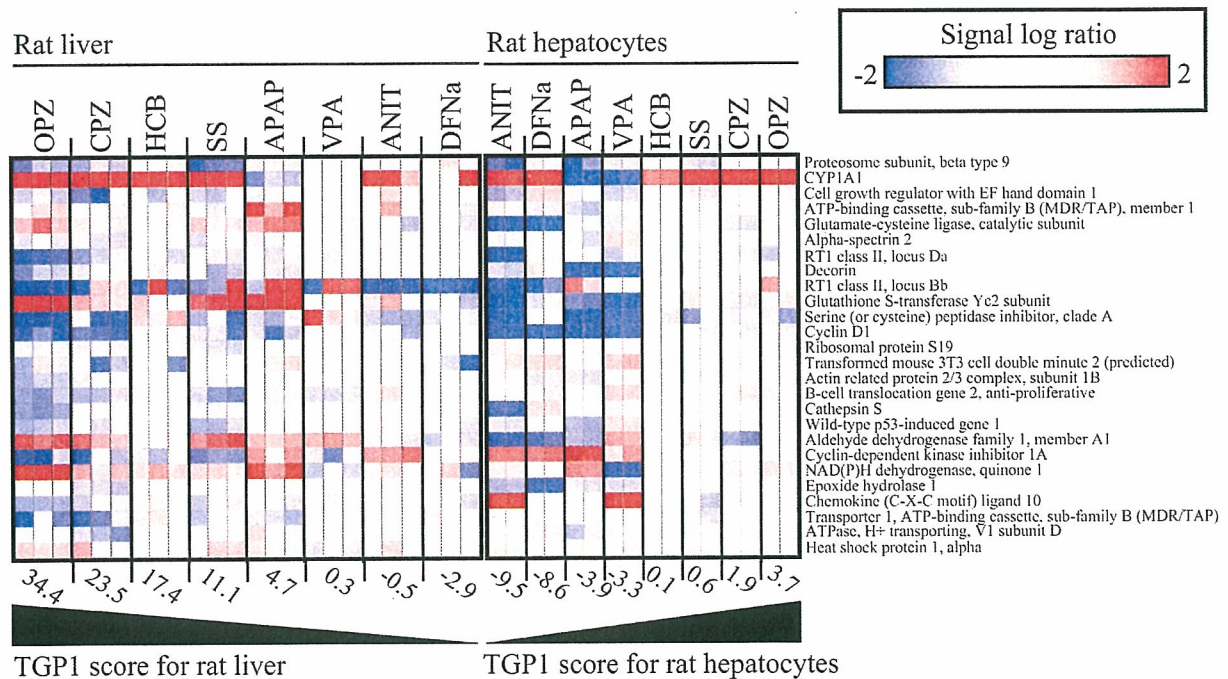


Fig. 5. Gene expression profile of carcinogenesis-related genes.

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.

Scoring the level of gene expression changes in microarray analysis.

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 toxicogenom-

ics 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.

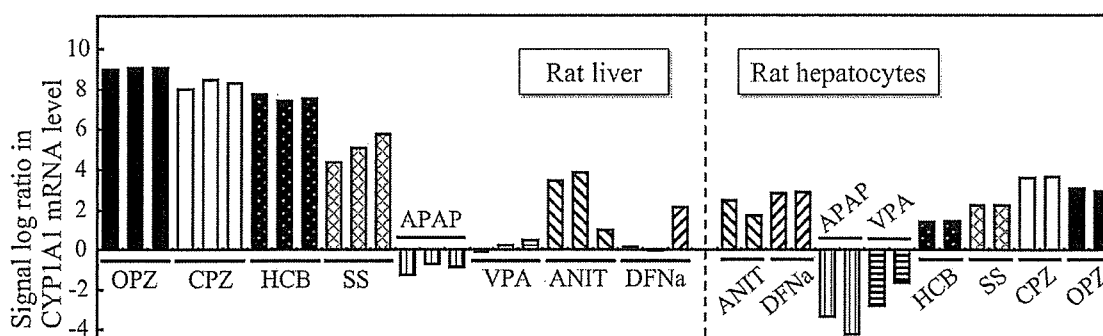


Fig. 6. Change in CYP1A1 mRNA level.

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.

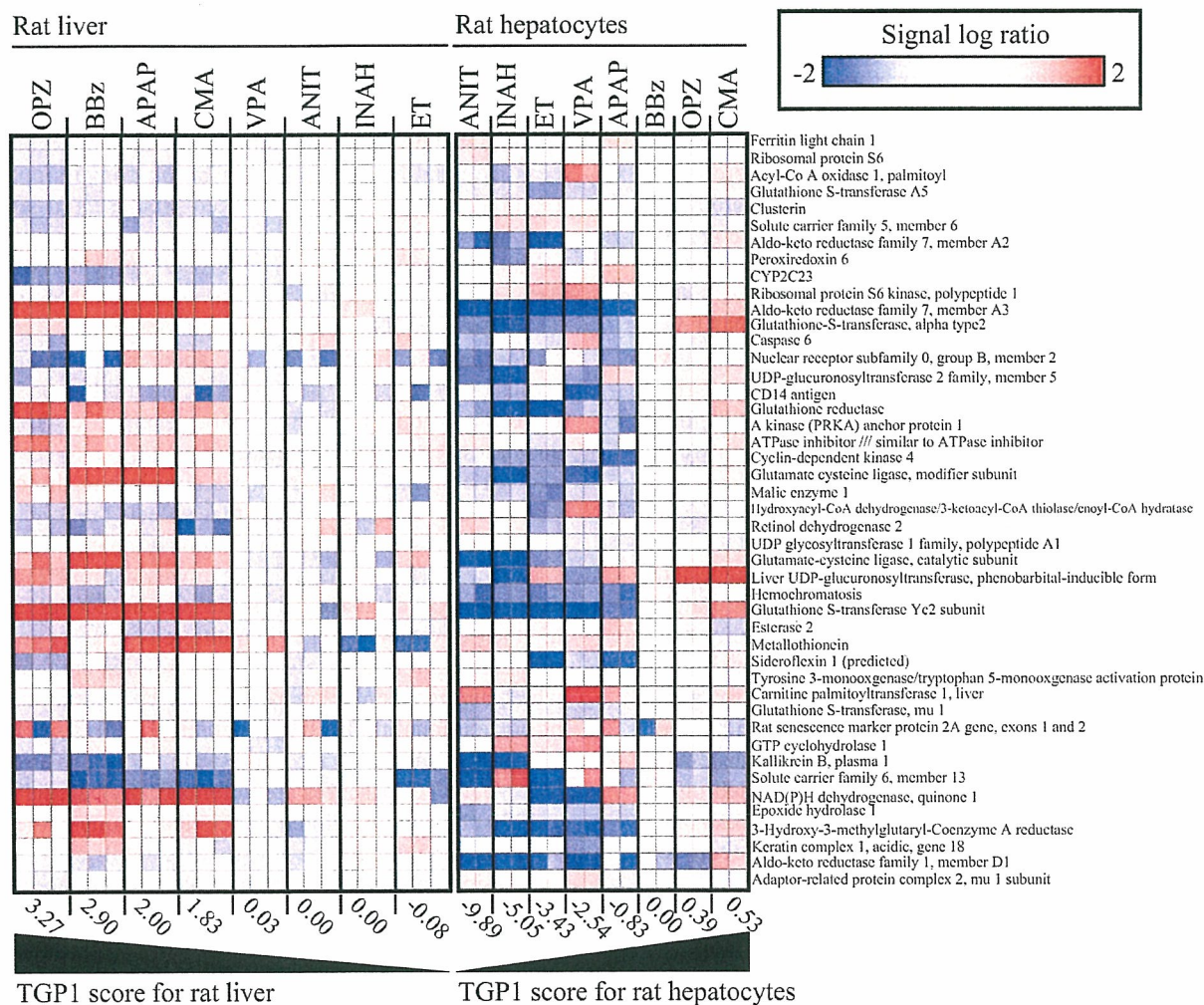


Fig. 7. Gene expression profile of glutathione deficiency-related genes.

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.

Scoring the level of gene expression changes in microarray analysis.

- 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., Beaman, 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., Lambole, C., Viollon-Abadie, C., Grass, P., Hartmann, N., Laurent, S., Heyd, B., Mantion, G., Chibout, S.D. and Staedtler, F. (2003): Effects of clofibrac 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.

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

Kotaro TAMURA¹, Atsushi ONO¹, Toshikazu MIYAGISHIMA¹,
Taku NAGAO² and Tetsuro URUSHIDANI^{1,3}

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

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

³Department 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 nephrotoxics 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 com-

pared 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

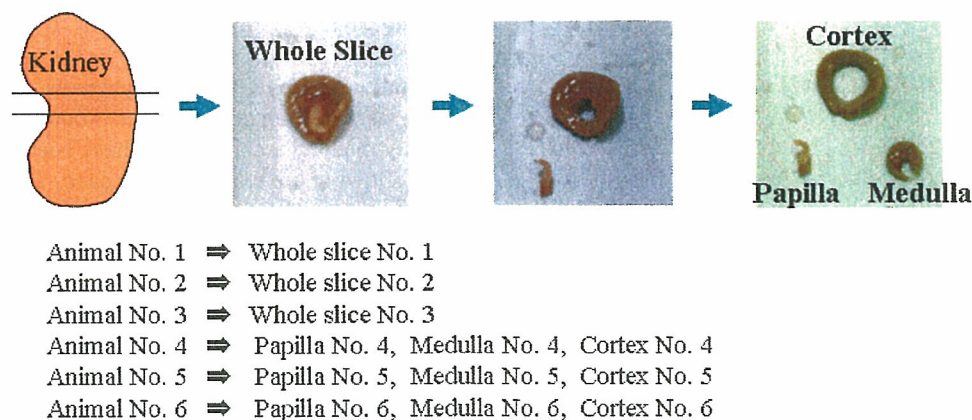


Fig. 1. Sampling and preparation of kidney for GeneChip analysis.

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 ver-

sion 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 - 0.971) > 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

Table 1. Signal intensity of β -actin and GAPDH.

| Value type | Gene Title | Papilla | | | Medulla | | | Cortex | | | Whole | | |
|--------------------------|----------------|---------|-------|-------|---------|-------|-------|--------|-------|-------|-------|-------|-------|
| | | No.1 | No.2 | No.3 | No.1 | No.2 | No.3 | No.1 | No.2 | No.3 | No.1 | No.2 | No.3 |
| Signal intensity | β -actin | 19619 | 20290 | 18700 | 20158 | 22911 | 20379 | 19881 | 23365 | 20759 | 19277 | 19278 | 21390 |
| | GAPDH | 17008 | 16029 | 15284 | 20169 | 21086 | 20291 | 22338 | 23834 | 23006 | 20765 | 20430 | 20458 |
| Mean of signal intensity | β -actin | 0.95 | 0.98 | 0.90 | 0.98 | 1.11 | 0.99 | 0.96 | 1.13 | 1.00 | 0.93 | 0.93 | 1.03 |
| | GAPDH | 0.85 | 0.81 | 0.77 | 1.01 | 1.06 | 1.02 | 1.12 | 1.20 | 1.16 | 1.04 | 1.03 | 1.03 |

Expression of β -actin and GAPDH in the three portions of kidney slice, papilla, medulla, and cortex, from three different rats (No. 1 - 3) as well as in the whole slice of three different rats (No. 4 - 6). Signal intensity of each gene was divided by the mean of all probes in the chip multiplied by 500 (global mean normalization, upper columns), after which each value was divided by the mean of these 12 values (per gene normalization).

Gene expression in rat kidney.

in all samples of papilla but absent in all of medulla and cortex were 448, and only 27 (6%) were all present in the whole slice. Genes with present call in all samples of medulla but absent in others were quite rare, i.e., 18 probe sets, of which 2 (11%) had present call in all of the whole slices. Cortex-specific probes were found to be 44, and 34 (77%) were present in all of the samples of the whole slice. The relatively small number in the latter two portions indicates that most of the genes are common between medulla and cortex, and the gene expression in papilla is unique. These results suggest again that the region specific genes (in other than cortex) are difficult to detect by analysis of the whole slice.

The above results clearly indicate that the population of genes expressed in each region is quite differ-

ent. Theoretically, absolute values of expression should be used when an accurate comparison is made between regions with different total mRNA contents. In order to further elucidate this point, the "percellome procedure" was employed to compare with global normalization.

The mean of copy numbers (or the values directly related to copy numbers) of β -actin was calculated to be 234, 291, 341, and 309 in papilla, medulla, cortex and the whole slice, respectively, i.e., the ratio in papilla, medulla, and cortex (whole slice = 1) was 0.76, 0.94 and 1.10, respectively. The mean of copy numbers of GAPDH was calculated to be 208, 298, 369 and 343, in papilla, medulla, cortex and the whole slice, respectively, i.e., the ratio in papilla, medulla, cortex (whole slice = 1) was 0.61, 0.87 and 1.08, respectively. These

Table 2. Pearson's correlation coefficient between samples.

| A | | Papilla | | | Medulla | | | Cortex | | | Whole | | | |
|---------|-----|---------|-------|-------|---------|-------|-------|--------|-------|-------|-------|-------|-------|--|
| | | No1 | No2 | No3 | No1 | No2 | No3 | No1 | No2 | No3 | No4 | No5 | No6 | |
| Papilla | No1 | 1 | 0.989 | 0.986 | 0.847 | 0.843 | 0.837 | 0.765 | 0.758 | 0.761 | 0.814 | 0.811 | 0.812 | 0.90 – 0.85 – 0.80 – 0.75 – – 0.75 |
| | No2 | | 1 | 0.985 | 0.853 | 0.853 | 0.846 | 0.778 | 0.773 | 0.775 | 0.827 | 0.825 | 0.825 | |
| | No3 | | | 1 | 0.815 | 0.812 | 0.811 | 0.738 | 0.73 | 0.735 | 0.789 | 0.785 | 0.786 | |
| Medulla | No1 | | | | 1 | 0.994 | 0.991 | 0.947 | 0.933 | 0.937 | 0.958 | 0.967 | 0.956 | |
| | No2 | | | | | 1 | 0.991 | 0.943 | 0.935 | 0.938 | 0.957 | 0.968 | 0.955 | |
| | No3 | | | | | | 1 | 0.953 | 0.943 | 0.953 | 0.964 | 0.972 | 0.962 | |
| Cortex | No1 | | | | | | 1 | 0.991 | 0.992 | 0.986 | 0.983 | 0.986 | | |
| | No2 | | | | | | | 1 | 0.991 | 0.983 | 0.979 | 0.984 | | |
| | No3 | | | | | | | | 1 | 0.984 | 0.98 | 0.985 | | |
| Whole | No4 | | | | | | | | | 1 | 0.996 | 0.994 | | |
| | No5 | | | | | | | | | | 1 | 0.992 | | |
| | No6 | | | | | | | | | | | 1 | | |

| B | | Papilla | | | Medulla | | | Cortex | | | Whole | | | |
|---------|-----|---------|-------|-------|---------|-------|-------|--------|-------|-------|-------|-------|-------|--|
| | | No1 | No2 | No3 | No1 | No2 | No3 | No1 | No2 | No3 | No4 | No5 | No6 | |
| Papilla | No1 | 1 | 0.989 | 0.985 | 0.837 | 0.833 | 0.826 | 0.755 | 0.755 | 0.752 | 0.801 | 0.797 | 0.801 | 0.90 – 0.85 – 0.80 – 0.75 – – 0.75 |
| | No2 | | 1 | 0.982 | 0.843 | 0.844 | 0.834 | 0.768 | 0.769 | 0.764 | 0.813 | 0.81 | 0.814 | |
| | No3 | | | 1 | 0.801 | 0.798 | 0.795 | 0.725 | 0.723 | 0.724 | 0.771 | 0.766 | 0.77 | |
| Medulla | No1 | | | | 1 | 0.994 | 0.991 | 0.946 | 0.934 | 0.937 | 0.957 | 0.967 | 0.955 | |
| | No2 | | | | | 1 | 0.991 | 0.942 | 0.936 | 0.937 | 0.956 | 0.967 | 0.954 | |
| | No3 | | | | | | 1 | 0.953 | 0.945 | 0.953 | 0.963 | 0.971 | 0.961 | |
| Cortex | No1 | | | | | | 1 | 0.992 | 0.992 | 0.987 | 0.984 | 0.986 | | |
| | No2 | | | | | | | 1 | 0.992 | 0.986 | 0.982 | 0.986 | | |
| | No3 | | | | | | | | 1 | 0.985 | 0.981 | 0.985 | | |
| Whole | No4 | | | | | | | | | 1 | 0.996 | 0.994 | | |
| | No5 | | | | | | | | | | 1 | 0.992 | | |
| | No6 | | | | | | | | | | | 1 | | |

A: Calculated on the data of global normalization.

B: Calculated on the data of percellome.

values clearly differ from those by global normalization, namely, the expression values in papilla were apparently overestimated by global normalization. The correlation coefficients calculated from the values normalized by percellome (Table 2B) were almost identical to that in Table 2A except between papilla and others, which showed a relatively large decrease in value. These results again indicated that global normalization was problematic, especially for papilla among the portions, most possibly because of the low amount of mRNA production in that region, compared with others.

After normalization either by global mean or percellome, the genes with absent call in all the samples were discarded and analyzed by ANOVA ($p < 0.01$) in order to extract probe sets showing different expression in any region(s). The numbers of extracted probe sets were 12,322 for percellome and 8,161 for global normalization. Fig. 3 shows the results of K-means clustering (Euclidean, 10 clusters) of the expression values converted into z-scores in order to see the trend of expression in each region.

By percellome normalization (Fig. 3A), clusters of probe sets with characteristic region-dependent

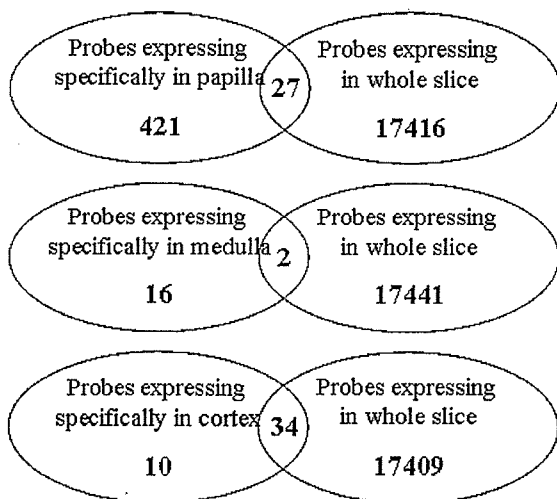


Fig. 2. Venn diagram of region-specific genes extracted by flags (present, absent and marginal call). The probes having "present call" in all samples of papilla and "absent call" in all samples of medulla and cortex were considered to be papilla-specific. The medulla- and cortex-specific probes were also extracted in the same manner and examined whether they were absent in the whole slice.

expression were efficiently extracted. It was also obvious from the figure that there was only one cluster containing only 293 probe sets that showed specifically high expression in papilla, and the other probe sets showed the lowest value in papilla compared to any other clusters. This indicates that most of the probe sets showed the lowest expression in papilla and the highest ones were exceptional. On the other hand, the clusters of global normalization (Fig. 3B) showed various, inconsistent patterns. Contrasting with the case of percellome normalization, there were many clusters in which gene expression was highest in papilla, or where papilla was equivalent to other region(s).

An obvious contradiction was noted between the two normalizations as described above, so further comparison was made. After the elimination of the probe sets that had absent call in all samples, the probe sets were classified into categories where the values showed maximum or minimum. Fig. 4 shows their counts in papilla, medulla, cortex, and whole slice for each normalization method. In the case of percellome normalization (A), the numbers of probe sets showing a maximal value were dominantly found in cortex, while the most of those showing a minimum were in papilla. On the other hand, in the case of global normalization (B), many of the probe sets showing their maximum were in papilla, while the numbers of probe sets with minimal expression were distributed evenly among the samples. It is noteworthy that very many (>3,000) probe sets showed minimal expression in the whole slice. However, this is theoretically impossible because the whole slice contains all the other 3 portions. Therefore, it should be concluded that the extraction of the genes with region-specific expression based on global normalization gives an error, and thus percellome normalization should be used in this case.

Based on the above results, region-specific genes in kidney were extracted as follows. After selection by ANOVA ($p < 0.01$) for the data of percellome normalization, genes were categorized by the position at which they showed the larger expression value and then aligned in the order of their ratio to the minimum, for papilla (Table 3), medulla (Table 4) and cortex (Table 5). As is obvious from Figs. 3 and 4, the production of mRNA per cell was considered to be in the order of cortex>medulla>papilla, and the numbers of region-specific probe sets were also in this order. For simplicity, probe sets without any annotation were eliminated, and the ones with the ratio of >3 for papilla, >10 for medulla, and >30 for cortex, are presented in the tables.

Gene expression in rat kidney.

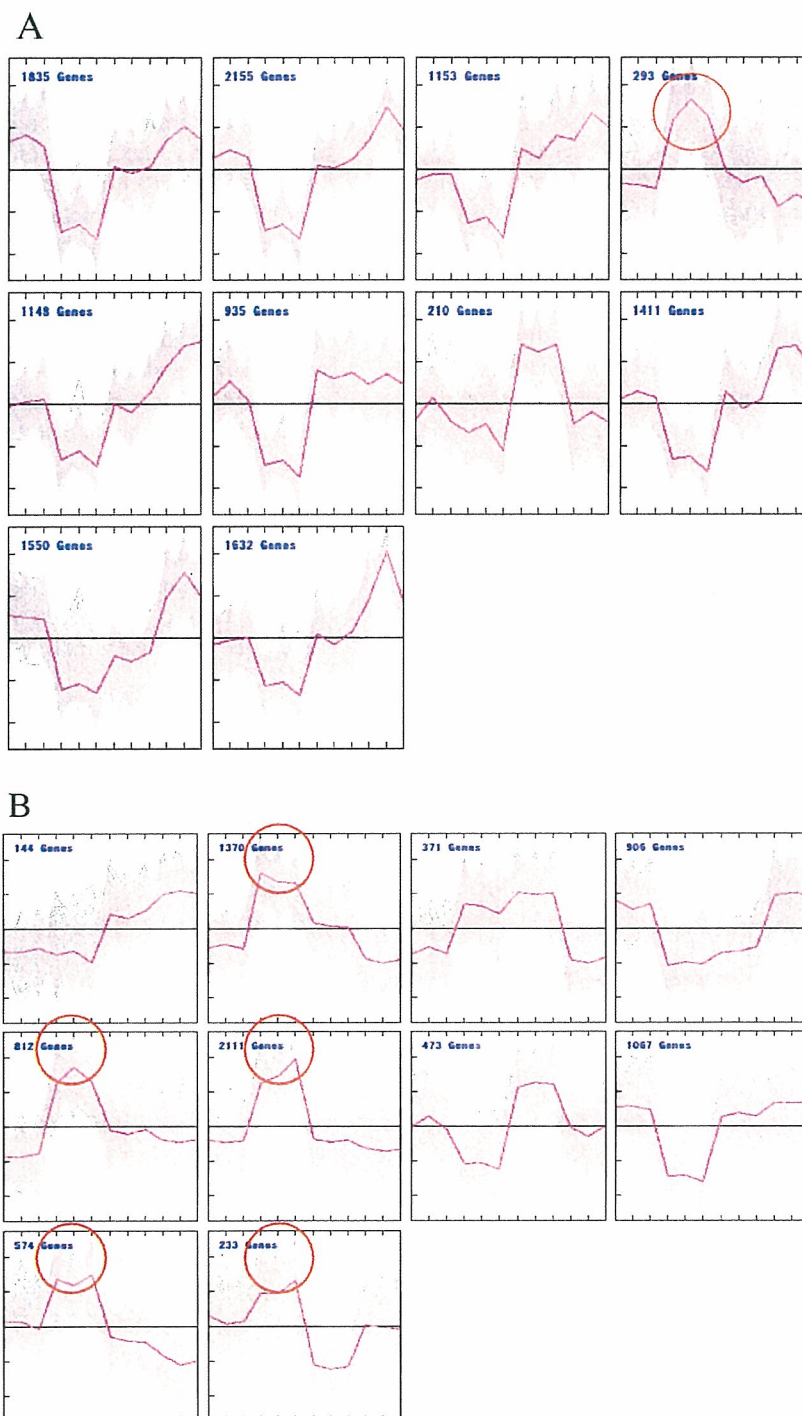


Fig. 3. K-means clustering of genes expressed in papilla, medulla, cortex, and whole slice of kidney. Data were processed by percellome normalization (A) or global mean normalization (B), and then converted into z-scores. K-means clustering (Euclidean, 10 clusters) was performed with MeV version 3.1 (The Institute for Genomic Research, Rockville, MD, USA). In each cluster, individual samples are aligned from left to right: whole slice (3), papilla (3), medulla (3), and cortex (3). Red lines indicate the mean of the probes within each cluster. Orange circles indicate where papilla showed specifically high expression values.

In all these three tables, the higher rank is generally occupied by the genes related to channels, transporters and metabolic enzymes, suggesting that the gene lists are meaningful for analysis of specific renal functions.

In most of the papilla-enriched genes (Table 3), the ratio of papilla/cortex is larger than that of papilla/medulla, since the composition of papilla is closer to medulla than to cortex. The exceptional genes (highest in papilla and lowest in medulla) are shaded in the table. The outstanding feature of this table is that heat shock proteins and cytoskeleton/extracellular matrix proteins are present, in addition to the channel/transporters and metabolic enzymes.

In most of the medulla-enriched genes (Table 4), the ratio of medulla/cortex is less than 2, suggesting that their expression is relatively similar between these two portions. The genes with ratio of >3 are shaded in the table, but they are only 3 sets, indicating that medulla-specific genes are rare. The higher rank of the list in Table 4 is also occupied by channel/transporters and metabolic enzymes, and cytoskeletal proteins are

scarce. A unique feature of medulla is the existence of 4 probe sets for prolactin receptor. This might occur simply because the quality of these multiple probes for one prolactin receptor is uniformly high.

Table 5 shows genes that showed the highest value in the cortex. It is noteworthy that many genes show more than 3 fold (shaded in the table) for the cortex/medulla ratio, indicating that there are many cortex-specific genes. The higher rank of this table is also occupied by channel/transporters and metabolic enzymes, but the numbers of metabolic enzymes are more prominent than in medulla.

Table 6 summarizes the genes categorized as channel/transporters, metabolic enzymes, cytoskeletons, and others for each portion.

DISCUSSION

The kidney is composed of various types of cells, and each portion (papilla, medulla, and cortex) has specific functions with wide variety, and the adverse effects of drugs vary with each portion. For example,

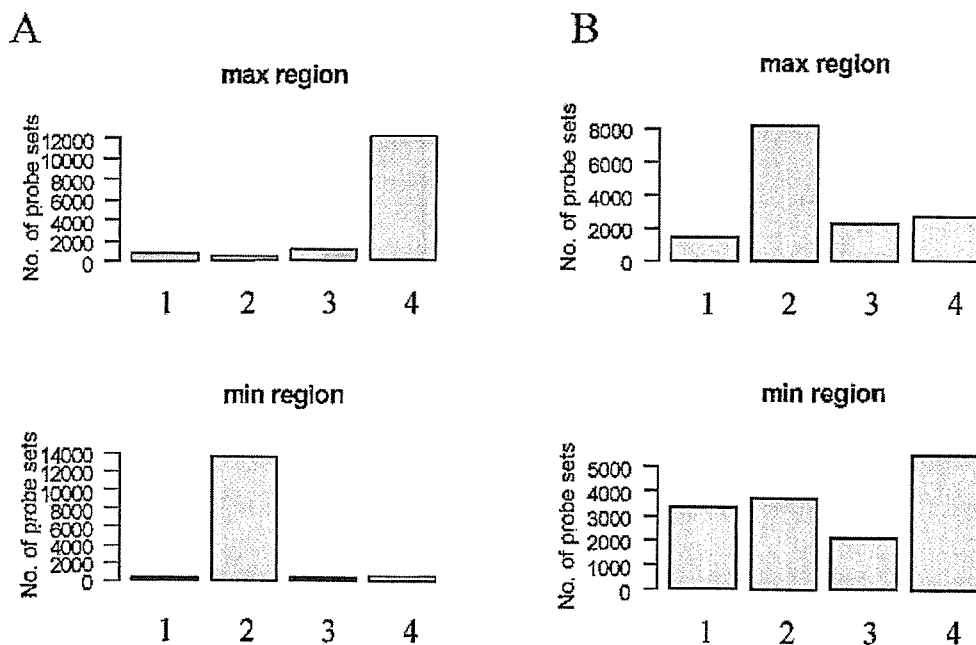


Fig. 4. The numbers of probe sets where their expression was maximal or minimal in whole slice, papilla, medulla, or cortex of kidney. After normalization by percellome (A) or global mean (B), probes with absent call in all samples were eliminated, and the numbers of probe sets with maximal (upper panels) or minimal (lower panels) were counted for each region, i.e., whole slice (1), papilla (2), medulla (3), and cortex (4).

Gene expression in rat kidney.

Table 3. A list of probe sets specifically expressed in papilla of kidney.

| Probe sets | papilla/ cortex | papilla/ medulla | medulla/ cortex | GENE_SYMBOL | GENE_NAME | channel transporter | Metabolic enzymes | cytoskeleton /extracellular matrix |
|------------|--------------------|---------------------|--------------------|-------------------|--|------------------------|----------------------|--|
| 1368259_at | 32.0 | 12.7 | 2.5 | Ptgs1 | prostaglandinendoperoxide synthase 1 | | | |
| 1389067_at | 29.5 | 10.7 | 2.8 | Slco4a1 | solute carrier organic anion transporter family, member 4a1 | | | |
| 1372190_at | 23.8 | 8.1 | 2.9 | Aqp4 | Aquaporin 4 | | | |
| 1394200_at | 22.9 | 4.9 | 4.7 | Hspa2 | heat shock protein 2 | | | |
| 1387092_at | 17.5 | 5.4 | 3.2 | Fxyd4 | FXFD domain-containing ion transport regulator 4 | | | |
| 1367847_at | 17.4 | 5.4 | 3.2 | Nupr1 | nuclear protein 1 | | | |
| 1388460_at | 16.8 | 5.6 | 3.0 | Capg_predicted | capping protein (actin filament), gelsolin-like (predicted) | | | |
| 1383469_at | 7.2 | 14.6 | 0.5 | Aldh1a3 | Aldehyde dehydrogenase family 1, subfamily A3 | | | |
| 1376711_at | 13.3 | 7.1 | 1.9 | Cldn11 | claudin 11 | | | |
| 1383319_at | 13.2 | 2.5 | 5.3 | Slc4a11_predicted | solute carrier family 4, sodium bicarbonate transporter-like, member 11 (predicted) | | | |
| 1367734_at | 13.1 | 7.9 | 1.7 | Akr1b4 | aldo-keto reductase family 1, member B4 (aldose reductase) | | | |
| 1368765_at | 10.7 | 2.5 | 4.2 | Clcnk1 | chloride channel K1 | | | |
| 1370229_at | 10.5 | 1.9 | 5.4 | Ndr4 | N-myc downstream regulated 4 | | | |
| 1369841_at | 10.3 | 4.2 | 2.4 | Hspa2 | heat shock protein 2 | | | |
| 1382303_at | 3.8 | 8.8 | 0.4 | RGD:1303187 | phosphatase and actin regulator 1 | | | |
| 1378690_at | 7.6 | 3.9 | 2.0 | Ly6a_predicted | lymphocyte antigen 6 complex, locus A (predicted) | | | |
| 1367661_at | 6.8 | 3.7 | 1.8 | S100a6 | S100 calcium binding protein A6 (calcyclin) | | | |
| 1374207_at | 6.6 | 2.5 | 2.6 | Agpt2 | angiopoietin 2 | | | |
| 1369113_at | 6.1 | 3.9 | 1.6 | Grem1 | gremlin 1 homolog, cysteine knot superfamily (<i>Xenopus laevis</i>) | | | |
| 1368858_at | 5.9 | 2.1 | 2.8 | Ugt8 | UDP-glucuronosyltransferase 8 | | | |
| 1368247_at | 5.4 | 3.7 | 1.5 | Hspa1a /// Hspa1b | heat shock 70kD protein 1A /// heat shock 70kD protein 1B | | | |
| 1370334_at | 3.2 | 5.4 | 0.6 | Plekhb1 | evectin-1 | | | |
| 1367650_at | 5.1 | 3.1 | 1.6 | Len7 | lipocalin 7 | | | |
| 1374861_at | 5.1 | 2.1 | 2.4 | Tle2_predicted | transducin-like enhancer of split 2, homolog of <i>Drosophila E(spl)</i> (predicted) | | | |
| 1387100_at | 4.1 | 5.1 | 0.8 | Aqp3 | aquaporin 3 | | | |
| 1369949_at | 5.0 | 3.6 | 1.4 | Lu | Lutheran blood group (Auberger b antigen included) | | | |
| 1369263_at | 5.0 | 2.4 | 2.0 | Wnt5a | wingless-type MMTV integration site 5A | | | |
| 1370312_at | 4.9 | 1.5 | 3.4 | Spon1 | spondin 1 | | | |
| 1388459_at | 4.7 | 2.9 | 1.6 | Col18a1 | collagen, type XVIII, alpha 1 | | | |
| 1388456_at | 4.3 | 3.0 | 1.4 | S100a1 | S100 calcium binding protein A1 | | | |
| 1393209_at | 3.6 | 4.3 | 0.8 | bsnd | Barter syndrome, infantile, with sensorineural deafness (Barttin) | | | |
| 1373733_at | 4.3 | 2.8 | 1.5 | Bok | Bcl-2-related ovarian killer protein | | | |
| 1388547_at | 4.2 | 2.5 | 1.7 | Cldn4_predicted | claudin 4 (predicted) | | | |

Table 3. Continued.

| Probe sets | papilla/ cortex | papilla/ medulla | medulla/ cortex | GENE_SYMBOL | GENE_NAME | channel transporter | Metabolic enzymes | cytoskeleton /extracellular matrix |
|--------------|--------------------|---------------------|--------------------|-------------------|--|------------------------|----------------------|--|
| 1396152_s_at | 4.1 | 2.4 | 1.7 | Igfbp5 | insulin-like growth factor binding protein 5 | | | |
| 1367812_at | 4.0 | 3.3 | 1.2 | Spnb3 | beta-spectrin 3 | | | |
| 1367577_at | 3.4 | 3.9 | 0.9 | Hspb1 | heat shock 27kDa protein 1 | | | |
| 1372755_at | 3.9 | 2.5 | 1.6 | Mal2 | mal, T-cell differentiation protein 2 | | | |
| 1370834_at | 3.8 | 1.8 | 2.2 | Hs3st1 | heparan sulfate (glucosamine) 3-O-sulfotransferase 1 | | | |
| 1388155_at | 3.8 | 3.3 | 1.2 | Krt1-18 | keratin complex 1, acidic, gene 18 | | | |
| 1372299_at | 3.7 | 2.3 | 1.6 | Cdkn1c | cyclin-dependent kinase inhibitor 1C (P57) | | | |
| 1387886_at | 3.6 | 2.7 | 1.3 | Prelp | proline arginine-rich end leucine-rich repeat protein | | | |
| 1368527_at | 1.7 | 3.6 | 0.5 | Ptgs2 | prostaglandin-endoperoxide synthase 2 | | | |
| 1388102_at | 1.9 | 3.5 | 0.5 | Ltb4dh | leukotriene B4 12-hydroxydehydrogenase | | | |
| 1370048_at | 3.4 | 2.8 | 1.2 | Edg2 | endothelial differentiation, lysophosphatidic acid G-protein-coupled receptor, 2 | | | |
| 1371004_at | 3.4 | 1.7 | 2.0 | Sort1 | sortilin 1 | | | |
| 1397830_at | 3.4 | 2.3 | 1.5 | Igfbp5 | Insulin-like growth factor binding protein 5 | | | |
| 1369953_a_at | 3.3 | 1.5 | 2.2 | Cd24 | CD24 antigen | | | |
| 1367912_at | 3.2 | 2.3 | 1.4 | Ltbp1 | latent transforming growth factor beta binding protein 1 | | | |
| 1387566_at | 3.2 | 1.6 | 2.0 | Pla2g4a | phospholipase A2, group IVA (cytosolic, calcium-dependent) | | | |
| 1370912_at | 3.2 | 2.3 | 1.4 | Hspa1b | heat shock 70kD protein 1B | | | |
| 1398318_at | 3.2 | 2.1 | 1.5 | Muc1 | mucin 1, transmembrane | | | |
| 1371625_at | 3.1 | 2.8 | 1.1 | Pygb | brain glycogen phosphorylase | | | |
| 1370026_at | 3.1 | 1.1 | 2.8 | Cryab | crystallin, alpha B | | | |
| 1393048_at | 3.1 | 1.1 | 2.8 | Adra2a | Adrenergic receptor, alpha 2a | | | |
| 1388143_at | 3.1 | 2.4 | 1.3 | Col18a1 | collagen, type XVIII, alpha 1 | | | |
| 1393958_at | 3.1 | 2.1 | 1.5 | Arhgap4 | Rho GTPase activating protein 4 | | | |
| 1369084_a_at | 3.1 | 2.8 | 1.1 | Bok | Bcl-2-related ovarian killer protein | | | |
| 1371499_at | 3.0 | 1.8 | 1.7 | Cd9 | CD9 antigen | | | |
| 1391830_at | 3.0 | 1.9 | 1.6 | Cpne8_predicted | copine VIII (predicted) | | | |
| 1368342_at | 3.0 | 1.9 | 1.6 | Ampd3 | adenosine monophosphate deaminase 3 | | | |
| 1384192_at | 3.0 | 2.1 | 1.4 | Chst1_predicted | carbohydrate (keratan sulfate Gal-6) sulfotransferase 1 (predicted) | | | |
| 1398431_at | 1.2 | 3.0 | 0.4 | Car8_predicted | carbonic anhydrase 8 (predicted) | | | |
| 1375170_at | 3.0 | 1.6 | 1.9 | S100a11_predicted | S100 calcium binding protein A11 (calizzarin) (predicted) | | | |
| 1367759_at | 3.0 | 1.2 | 2.5 | H1f0 | H1 histone family, member 0 | | | |
| 1387040_at | 3.0 | 1.6 | 1.8 | Mal | myelin and lymphocyte protein | | | |

After selection by ANOVA ($p < 0.01$) for the data of percellome normalization, genes maximally expressed in papilla were selected. The genes were aligned in the order of the ratio to the lower expression value, either in medulla or in cortex. As the genes listed here are expressed higher in medulla than cortex, in general, exceptional cases (ratio < 0.6) are shaded in the medulla/cortex column. The genes categorized to "channel/transporters", "metabolic enzymes", or "cytoskeleton/extracellular matrix" are also shaded. Proteases or enzymes involving signal transduction are not included in the category of "metabolic enzymes". For simplicity, genes with less than 3-fold specificity are omitted.

Gene expression in rat kidney.

Table 4. A list of probe sets specifically expressed in medulla of kidney.

| Probe sets | medulla/ papilla | cortex/ papilla | medulla/ cortex | GENE_SYMBOL | GENE_NAME | channel transporter | Metabolic enzymes | cytoskeleton /extracellular matrix |
|--------------|---------------------|--------------------|--------------------|---------------------------------|--|------------------------|----------------------|--|
| 1370377_at | 204.8 | 124.5 | 1.6 | Cyp2d9 /// Cyp2d10 | cytochrome P450, family 2, subfamily d, polypeptide 9 /// cytochrome P450, family 2, subfamily d, polypeptide 10 | | | |
| 1387567_at | 184.9 | 119.7 | 1.5 | Slc21a1 /// LOC497799 | solute carrier family 21, member 1 /// hypothetical gene supported by NM_017111 | | | |
| 1369401_at | 153.0 | 69.6 | 2.2 | Slc21a13 | solute carrier family 21, member 13 | | | |
| 1387328_at | 149.6 | 100.5 | 1.5 | Cyp2c | Cytochrome P450, subfamily IIC (mephenytoin 4-hydroxylase) | | | |
| 1368288_at | 147.9 | 90.7 | 1.6 | Gc | group specific component | | | |
| 1368498_a_at | 147.3 | 102.7 | 1.4 | RGD:621387 | kidney specific organic anion transporter | | | |
| 1386454_at | 133.0 | 46.7 | 2.8 | Slc23a3_predicted | solute carrier family 23 (nucleobase transporters), member 3 (predicted) | | | |
| 1370789_a_at | 132.6 | 91.8 | 1.4 | Prlr | prolactin receptor | | | |
| 1387987_at | 120.1 | 43.9 | 2.7 | Slc22a19 | solute carrier family 22 (organic anion transporter), member 19 | | | |
| 1390569_at | 117.4 | 74.7 | 1.6 | RGD:1359493 | similar to carnosinase 1 | | | |
| 1369450_at | 114.9 | 92.3 | 1.2 | UST5r | integral membrane transport protein UST5r | | | |
| 1369493_at | 89.1 | 62.1 | 1.4 | Prlr | prolactin receptor | | | |
| 1368575_at | 88.9 | 70.0 | 1.3 | Slc6a18 | solute carrier family 6 (neurotransmitter transporter), member 18 | | | |
| 1370824_at | 88.2 | 60.3 | 1.5 | Slc38a3 | solute carrier family 38, member 3 | | | |
| 1387382_at | 88.0 | 42.0 | 2.1 | Hnmt | histamine N-methyltransferase | | | |
| 1387303_at | 83.1 | 59.2 | 1.4 | Slc22a2 | solute carrier family 22 (organic cation transporter), member 2 | | | |
| 1378247_at | 81.5 | 41.3 | 2.0 | Eaf2 | ELL associated factor 2 | | | |
| 1373990_at | 78.5 | 24.0 | 3.3 | Slc7a12_predicted /// LOC361914 | solute carrier family 7 (cationic amino acid transporter, y+ system), member 12 (predicted) /// similar to solute carrier family 7 (cationic amino acid transporter, y+ system), member 12 | | | |
| 1389756_at | 78.0 | 57.9 | 1.3 | Melk_predicted | maternal embryonic leucine zipper kinase (predicted) | | | |
| 1384775_s_at | 74.6 | 51.7 | 1.4 | Tmprss8 | transmembrane protease, serine 8 (intestinal) | | | |
| 1370384_a_at | 73.9 | 60.3 | 1.2 | Prlr | prolactin receptor | | | |
| 1368208_at | 72.9 | 60.3 | 1.2 | Cml1 | camello-like 1 | | | |
| 1376944_at | 72.5 | 68.6 | 1.1 | Prlr | Prolactin receptor | | | |
| 1385132_at | 69.0 | 32.5 | 2.1 | Mybl1_predicted | myeloblastosis oncogene-like 1 (predicted) | | | |
| 1368651_at | 66.4 | 49.0 | 1.4 | Pklr | pyruvate kinase, liver and RBC | | | |
| 1368304_at | 65.8 | 57.1 | 1.2 | Fmo3 | Flavin containing monooxygenase 3 | | | |
| 1397205_at | 54.9 | 42.8 | 1.3 | Dhrs7_predicted /// LOC500672 | dehydrogenase/reductase (SDR family) member 7 (predicted) /// similar to Down-regulated in nephrectomized rat kidney #3 | | | |

Table 4. Continued.

| Probe sets | medulla/ papilla | cortex/ papilla | medulla/ cortex | GENE_SYMBOL | GENE_NAME | channel transporter | Metabolic enzymes | cytoskeleton /extracellular matrix |
|--------------|---------------------|--------------------|--------------------|--|--|------------------------|----------------------|--|
| 1398612_at | 51.8 | 37.6 | 1.4 | Akr1c12_predicted | aldo-keto reductase family 1, member C12 (predicted) | | | |
| 1384639_at | 51.6 | 49.0 | 1.1 | Dp111_predicted | deleted in polyposis 1-like 1 (predicted) | | | |
| 1368627_at | 49.7 | 35.7 | 1.4 | Rgn | regucalcin | | | |
| 1368366_at | 47.1 | 42.1 | 1.1 | Cml2 | Camello-like 2 | | | |
| 1387234_at | 44.2 | 42.4 | 1.0 | Azgp1 | alpha-2-glycoprotein 1, zinc | | | |
| 1368163_at | 43.3 | 43.2 | 1.0 | Dpp4 | dipeptidylpeptidase 4 | | | |
| 1372841_at | 41.7 | 32.3 | 1.3 | Dp111_predicted | deleted in polyposis 1-like 1 (predicted) | | | |
| 1398255_at | 38.7 | 20.9 | 1.8 | Slc15a2 | solute carrier family 15 (H+/peptide transporter), member 2 | | | |
| 1367905_at | 38.2 | 29.7 | 1.3 | Enpp3 | ectonucleotide pyrophosphatase/phosphodiesterase 3 | | | |
| 1370688_at | 33.0 | 27.8 | 1.2 | Gclc | glutamate-cysteine ligase, catalytic subunit | | | |
| 1368374_a_at | 31.9 | 29.2 | 1.1 | Ggt1 | gamma-glutamyltransferase 1 | | | |
| 1387218_at | 30.2 | 23.1 | 1.3 | Tff3 | trefoil factor 3 | | | |
| 1370714_a_at | 30.0 | 17.1 | 1.8 | Siat1 | sialyltransferase 1 | | | |
| 1373773_at | 29.4 | 25.3 | 1.2 | Gpm6a | glycoprotein m6a | | | |
| 1387357_at | 28.1 | 24.8 | 1.1 | Tmlhe | trimethyllysine hydroxylase, epsilon | | | |
| 1380962_at | 27.6 | 23.3 | 1.2 | Ace2 | Angiotensin I converting enzyme (peptidyl-dipeptidase A) 2 | | | |
| 1370144_at | 26.9 | 18.1 | 1.5 | Gtpbp4 /// LOC364763 /// LOC498786 | G protein-binding protein CRFG /// similar to GTP-binding protein NGB /// similar to GTP-binding protein NGB | | | |
| 1387209_at | 24.6 | 22.3 | 1.1 | Rgpr | regucalcin gene promotor region related protein | | | |
| 1367838_at | 24.5 | 17.9 | 1.4 | Cth | CTL target antigen | | | |
| 1390855_at | 23.9 | 12.7 | 1.9 | Prep | Prolyl endopeptidase | | | |
| 1371913_at | 23.4 | 11.8 | 2.0 | Tgfb1 | transforming growth factor, beta induced | | | |
| 1368234_at | 23.4 | 12.1 | 1.9 | Prep | prolyl endopeptidase | | | |
| 1388145_at | 22.5 | 22.4 | 1.0 | Tnxa | tenascin XA | | | |
| 1370365_at | 22.3 | 14.3 | 1.6 | Gss | glutathione synthetase | | | |
| 1381350_at | 20.3 | 17.4 | 1.2 | Idb4 | inhibitor of DNA binding 4 | | | |
| 1394022_at | 18.0 | 11.8 | 1.5 | Idb4 | inhibitor of DNA binding 4 | | | |
| 1368164_at | 17.9 | 14.7 | 1.2 | Blvra | biliverdin reductase A | | | |
| 1379300_at | 17.5 | 17.0 | 1.0 | Chst2_predicted | carbohydrate sulfotransferase 2 (predicted) | | | |
| 1387296_at | 17.1 | 15.7 | 1.1 | Cyp2j4 | cytochrome P450, family 2, subfamily J, polypeptide 4 | | | |
| 1377408_at | 16.9 | 16.2 | 1.0 | Pla2g6 | phospholipase A2, group VI | | | |
| 1369407_at | 16.8 | 10.4 | 1.6 | Tnfrsf11b | tumor necrosis factor receptor superfamily, member 11b (osteoprotegerin) | | | |
| 1382868_at | 16.5 | 9.9 | 1.7 | Sema6a_predicted | sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6A (predicted) | | | |
| 1387819_at | 16.3 | 16.3 | 1.0 | Ela1 | elastase 1, pancreatic | | | |
| 1372523_at | 16.3 | 14.9 | 1.1 | Gclc | glutamate-cysteine ligase, catalytic subunit | | | |