

OPEN ACCESS

Full open access to this and thousands of other papers at <http://www.la-press.com>.

New Short Term Prediction Method for Chemical Carcinogenicity by Hepatic Transcript Profiling Following 28-Day Toxicity Tests in Rats

Hiroshi Matsumoto¹, Yoshikuni Yakabe¹, Fumiyo Saito¹, Koichi Saito², Kayo Sumida², Masaru Sekijima³, Koji Nakayama³, Hideki Miyaura¹, Masanori Otsuka¹ and Tomoyuki Shirai⁴

¹Chemicals Assessment and Research Center, Chemicals Evaluation and Research Institute, Japan, 1600 Shimotakano, Sugito-machi, Kitakatsushika-gun, Saitama 345-0043, Japan. ²Sumitomo Chemical Co., Ltd., 1-98, 3-Chome, Kasugade-Naka, Konohana-ku, Osaka 554-8558, Japan. ³Advanced Medical Science Research Center, Mitsubishi Chemical Medience Corporation, 14 Sunayama, Kamisu, Ibaragi, Japan. ⁴Department of Experimental Pathology and Tumor Biology, Nagoya City University Graduate School of Medical Sciences, 1 Kawasumi, Mizuho-cho, Nagoya, 467-8601, Japan. Corresponding author email: matsumoto-hiroshi@ceri.jp

Abstract: We have previously shown the hepatic gene expression profiles of carcinogens in 28-day toxicity tests were clustered into three major groups (Group-1 to 3). Here, we developed a new prediction method for Group-1 carcinogens which consist mainly of genotoxic rat hepatocarcinogens. The prediction formula was generated by a support vector machine using 5 selected genes as the predictive genes and predictive score was introduced to judge carcinogenicity. It correctly predicted the carcinogenicity of all 17 Group-1 chemicals and 22 of 24 non-carcinogens regardless of genotoxicity. In the dose-response study, the prediction score was altered from negative to positive as the dose increased, indicating that the characteristic gene expression profile emerged over a range of carcinogen-specific doses. We conclude that the prediction formula can quantitatively predict the carcinogenicity of Group-1 carcinogens. The same method may be applied to other groups of carcinogens to build a total system for prediction of carcinogenicity.

Keywords: toxicogenomics, carcinogenicity, hepatocarcinogen, microarray, prediction method

Cancer Informatics 2011:10 259–271

doi: [10.4137/CIN.S7789](https://doi.org/10.4137/CIN.S7789)

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.



Introduction

Carcinogenicity is one of the most important endpoints of chemical safety evaluations, not only for pharmaceutical compounds but also for industrial chemicals. The two-year rodent carcinogenicity studies are generally used to judge the carcinogenicity of chemicals, but they are very expensive and need long test periods. Hence the carcinogenic potential of many chemicals remains unknown.

A number of alternative methods have been developed to screen carcinogens more easily. McCann et al reported in 1975 that carcinogenicity could be predicted from strength and a pattern of mutation of *Salmonella*.¹ Then, Zeiger et al identified rodent carcinogens and non-carcinogens by multiple genetic tests, including a mutagenicity test, and concluded that the *Salmonella* mutagenicity test is effective for the identification of mutagens and potential carcinogens, but not chemicals classed as non-mutagenic in the *Salmonella* mutagenicity test.² Elcombe et al studied acute and subacute biochemical and tissue changes as biomarkers to predict non-genotoxic carcinogenicity in rodents.³ However, a further verification study was required because only nine chemicals were tested. Ito et al developed the 8 week medium-term liver bioassay system by quantifying glutathione S-transferase placental-form (GST-P) positive foci as markers in F344 rat livers, which was employed in the International Conference on Harmonization⁴; 59 out of 64 (92%) hepatocarcinogens gave positive results, irrespective of their mutagenicity.⁵ The cancerogenic peroxisomal proliferators that suppress GST-P expression showed false negative in this method.

Microarray technologies enable the comprehensive analysis of gene expression, and their development has led to the emergence of the promising new scientific field of toxicogenomics. Toxicogenomics has been applied to the elucidation of toxicity mechanisms, exploration of biomarkers, and prediction of toxicity.^{6,7} Mathijs et al reported discrimination of genotoxic carcinogens from non-genotoxic carcinogens by using GeneChip array data derived from primary mouse hepatocytes; the two classes of carcinogens were separated from each other by hierarchical clustering, and the genes responsive to genotoxic carcinogens were extracted.⁸ However, the prediction of carcinogenicity using the “characteristic” genes was not reported. Ziegelbauer et al classified 29 chemicals

into genotoxic carcinogens, non-genotoxic carcinogens and non-hepatocarcinogens by using GeneChip data obtained from short-term animal experiments, and they tried to build a formula to predict the type of carcinogen by support vector machine method, resulting in a concordance of 88% for the validation data.⁹

In a previous study,¹⁰ we performed a hierarchical cluster analysis of the gene expression data obtained from rat liver in a 28-day repeated-dose toxicity study of 73 chemicals, comprising 47 carcinogens and 26 non-carcinogens, which were selected on the basis of their chemical and toxicological diversity. These carcinogens were separated into three major groups without relying on the selected gene and the administration period in the cluster analysis. We identified three “characteristic” gene sets, each of which showed gene expression changes specific for one of the three groups of carcinogens, suggesting that prediction formulae should be built by using “characteristic” gene sets from each group.

Here, as the first step towards development of a prediction method for carcinogenicity, a small “characteristic” gene set for Group-1 carcinogens was selected, and a prediction formula was built by using the support vector machine method. The performance of the prediction formula was examined by using validation chemicals in addition to the carcinogens in the other two groups demonstrated in the previous study. The effective dose range of the prediction formula was identified by a dose-response study.

Materials and Methods

Test chemicals

Eighty-six chemicals with known carcinogenicity were selected on the basis of their chemical and toxicological diversity from the US National Toxicology Program (NTP) database (<http://ntp.mehs.nih.gov/>) and the Chemical Carcinogenesis Research Information System (CCRIS) database (<http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?CCRIS>). They include 17 Group-1 carcinogens (Table 3a), 23 Group-2 carcinogens (Table 3b), and 6 Group-3 carcinogens (Table 3c), as well as 24 non-carcinogens and 16 validation chemicals consisting of 11 carcinogens and 5 non-carcinogens, and these were based on our previous study.¹⁰ The Group-1 carcinogens consisted of hepatocarcinogens mainly, especially mutagenic hepatocarcinogens. The Group-2



carcinogens consisted of mutagenic carcinogens and non-mutagenic carcinogens. Furthermore, carcinogens with estrogenic activity were included in Group-3 mainly. Two non-carcinogens, 4'-(chloroacetyl)acetanilide and 3-chloro-p-toluidine, were used in the grouping of carcinogens included in the validation data, because they showed gene expression patterns very similar to those of Group-1 carcinogens in the previous study.¹⁰

A summary of toxicity test conditions, carcinogenic properties, and carcinogen group number, as determined in the previous study of the test chemicals,¹⁰ is presented in Table 1. In this study, 17 Group-1 carcinogens and 24 non-carcinogens were used as training chemicals for prediction formula building; 23 Group-2 and 6 Group-3 carcinogens, and 16 validation chemicals, were used to validate the prediction formula (Table 2).

Animals and treatment

The 28-day repeat-dose toxicity study was performed as previously described.¹⁰ Fischer 344 (F344) rats were randomly assigned to two groups (treatment and control) consisting of 4 rats per group, and each rat was given a test chemical dissolved in a suitable vehicle (gum Arabic, corn oil, distilled water, carboxymethylcellulose or methylcellulose) or vehicle alone, by oral gavage once a day for 28 days. The dosage of each chemical was set at approximately its minimum carcinogenic doses (for carcinogens) or its maximum tolerated doses (for non-carcinogens) on the basis of the information in NTP (<http://ntp.niehs.nih.gov/>), CCRIS (<http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?CCRIS>) database, and published literature.

Four carcinogens (diethylnitrosoamine [T01], N-nitrosomorpholine [T03], 1,4-dioxane [T14], and thioacetamide [T16]) and 2 non-carcinogens (lithocholic acid [N15] and alpha-tocopherol [N19]) were selected at random from 17 Group-1 carcinogens and 24 non-carcinogens for dose-response studies. Dose-response studies with four dosages were then conducted to examine the dose dependency of the expression of the genes selected as the predictive genes, and the dose dependency of the power of the prediction method. The maximum dose was set as the maximum tolerated dose, and the remaining three doses were set up with a common ratio of 5.

In this study, we obtained ethics approval for the use of animals in all our animal testing.

Microarray experiments

Gene expression in the liver was measured by using a custom microarray, NEDO-ToxArrayIII (NGK Insulators, Ltd. Nagoya, Japan), consisting of 6709 unique genes, and data processing was performed as described previously.¹⁰ The raw data are available for download from the Gene Expression Omnibus repository (<http://www.ncbi.nlm.nih.gov/geo/>) at the National Center for Biotechnology Information (Accession ID, GSE16394). In a previous paper, 1359 genes were selected based on cut-off criteria, and it was used to build a prediction formula for carcinogenicity.¹⁰

Selection of predictive genes

The Group-1 “characteristic” genes were defined as genes induced or repressed specifically by administration of Group-1 carcinogens (ie, not by administration of non-carcinogens). They were selected by two criteria, as follows: 1) an absolute t-value >5 obtained from the Welch's t-test comparing log₂ratio values between the training sets of Group-1 carcinogens and non-carcinogens, where the log₂ratio is the logarithm to the base 2 of the mean signal intensity ratio between the treated sample and the corresponding vehicle control; and 2) an absolute value of the log₂ratio >0.8 in more than 70% of the test Group-1 carcinogens and less than 30% of the test non-carcinogens. Five genes were selected by applying these criteria to the training data of 17 Group-1 carcinogens and 24 non-carcinogens (Table 2).

Prediction formula building

The prediction formula for carcinogenicity of Group-1 carcinogens was built from the gene expression data of the 5 predictive genes (Table 4) with 17 Group-1 carcinogens and 24 non-carcinogens (Table 1) as training chemicals. Support Vector Machine (SVM), which is widely employed in toxicogenomics for prognostic prediction and marker searches,^{11,12} was used for the prediction formula building.

Because the formula is best suited to the training data, a number of validation datasets are required to assess its general applicability. To overcome this problem, we built the prediction formula as follows



Table 1. Summary of the toxicity tests with 86 test chemicals, their carcinogenic properties, and carcinogen groups clustered in this study.

Test no. ^a	Name	Source ^b	Toxicity test		Vehicle ^d	Histo-path. ^e	Carcinogenic properties			Carcinogen group no. ⁱ
			Dose ^c				Car. ^f	Hepato-car. ^g	Muta. ^h	
T01	Diethylnitrosoamine	B	20 (30, 6, 1.2, 0.24)		DW	FICI, SCN	+	+	+	1
T02	<i>N</i> -Nitrosodimethylamine	A	0.2		DW	NO	+	+	+	1
T03	<i>N</i> -Nitrosomorpholine	B	10 (141, 28.2, 5.64, 1.13)		DW	FICI, SCN	+	+	+	1
T04	<i>N</i> -Nitrosopiperidine	G	10		DW	DHH, FICI	+	+	+	1
T05	2-Nitropropane	A	40		CO	GI	+	+	+	1
T06	3'-Methyl-4-dimethylamino-azobenzene	B	50		CO	AH, PHH	+	NA	+	1
T07	2-Acetylaminofluorene	A	6		CO	DHH, VH	+	+	+	1
T08	MeIQx	H	20		1%CMC	NO	+	+	+	1
T09	Furan	C	10		CO	AH, HHN, VH	+	+	+	1
T10	Quinoline	E	25		CO	HHN, IMF, SCN	+	+	+	1
T11	2,4-Diaminotoluene	D	10		DW	PHH	+	+	+	1
T12	Methapyrilene HCl	A	50		DW	DHH, FICI, HHN, IMF, SCN	+	NA	-	1
T13	Acetamide	A	1180		DW	NO	+	+	-	1
T14	1,4-Dioxane	C	1000 (1000, 200, 40, 8)		DW	PHH	+	LP	-	1
T15	Methyl carbamate	B	500		DW	AH, GI, IMF	+	+	-	1
T16	Thioacetamide	C	20 (25, 5, 1, 0.2)		DW	AH, HHN	+	+	-	1
T17	Urethane	A	80		DW	NO	+	+	-	1
C01	<i>N</i> -Ethyl- <i>N</i> -nitrosourea	A	3		DW	NO	+	-	+	2
C02	4-Nitroquinoline-1-oxide	A	2		CO	NO	+	-	+	2
C03	4-Dimethylaminoazobenzene	C	50		CO	NO	+	N	+	2
C04	2-Amino-1-methyl-6-phenylimidazo[4,5- <i>b</i>]pyridine(PhIP)	H	5		1%CMC	NO	+	-	+	2
C05	Safrole	CA	300		CO	ECH, SCN	+	+	+	2
C06	Benzo[<i>a</i>]pyrene	C	15		CO	NO	+	-	+	2
C07	7,12-Dimethylbenz[<i>a</i>]anthracene	A	1		CO	NO	+	-	+	2
C08	3-Methylcholanthrene	A	2		CO	NO	+	-	+	2
C09	Clofibrate	A	250		CO	PHH	+	+	-	2
C10	Di(2-ethylhexyl)adipate	C	1000		CO	NO	+	-	-	2
C11	Di(2-ethylhexyl)phthalate	A	300		CO	PHH	+	+	-	2
C12	Phenytol	B	160		DW	NO	+	E	-	2
C13	Butylated hydroxyanisole	C	750		CO	NO	+	-	-	2
C14	<i>d</i> -Limonene	A	1000		CO	NO	+	-	-	2
C15	Aldrin	I	0.3		CO	NO	+	-	-	2
C16	Chlorendic acid	B	100		DW	NO	+	-	-	2
C17	1,4-Dichlorobenzene	B	300		CO	PHH	+	+	-	2



C18	Hexachlorobenzene	A	5		CO	NO	+	+	+	+	-	2
C19	Alpha-Hexachloro- cyclohexane	C	20		CO	PHH	+	+	+	+	-	2
C20	Trichloroethylene	A	700		CO	NO	-		+		-	2
C21	Tetrachloroethylene	A	100		CO	NO	-		+		-	2
C22	Trichloroacetic acid	A	300		DW	NO	-	NA	+		-	2
C23	DL-Ethionine	C	30		CO	NO	-	+	+		-	2
C24	Benz[<i>a</i>]anthracene	B	50		CO	NO	-	NA	+		+	3
C25	Phenobarbital	C	100		DW	PHH	-	NA	+		-	3
C26	Diethylstilbestrol	B	10		CO	VH	-	E	+		-	3
C27	Ethinylestradiol	A	0.5		CO	ATH	-	+	+		-	3
C28	Chloroform	A	90		CO	NO	-	+	+		-	3
C29	Pentachloroethane	B	200		CO	NO	-	-	+		-	3
N01	2-Chloromethylpyridine HCl	B	150		DW	NO	-	-	+		+	-
N02	2-Chloro- <i>p</i> -phenylenediamine SO ₄	A	100		1%CMC	NO	-	-	+		+	-
N03	2,6-Diaminotoluene	A	10		DW	NO	-	-	+		+	-
N04	8-Hydroxyquinoline	B	25		CO	NO	-	-	+		+	-
N05	4-Nitroanthranilic acid	B	1000		CO	GI	-	-	+		+	-
N06	1-Nitronaphthalene	B	100		CO	IMF	-	-	+		+	-
N07	4-Nitro- <i>o</i> -phenylenediamine	A	250		1%CMC	NO	-	-	+		+	-
N08	<i>p</i> -Phenylenediamine · 2HCl	A	60		DW	NO	-	-	+		+	-
N09	2,5-Toluenediamine SO ₄	B	50		1%CMC	NO	-	-	+		+	-
N10	L-Ascorbic acid	C	1000		DW	NO	-	-	-		-	-
N11	Aspirin	A	27		CO	NO	-	-	-		-	-
N12	Caprolactam	A	375		DW	NO	-	-	-		-	-
N13	Indomethacin	C	5		CO	NO	-	-	-		-	-
N14	Lindane	A	10		CO	NO	-	-	-		-	-
N15	Lithocholic acid	C	1000		5%AG.S	NO	-	-	-		-	-
N16	D-Mannitol	C	(750, 150, 30, 6)		DW	NO	-	-	-		-	-
N17	Phthalamide	B	1000		CO	NO	-	-	-		-	-
N18	Sodium benzoate	C	1000		DW	GI	-	-	-		-	-
N19	Alpha-Tocopherol	C	1000		CO	NO	-	NA	-		-	-
N20	2-Chloroethanol	A	40		DW	NO	-	-	-	+	+	-
N21	Iodoform	C	200		CO	NO	-	-	-	NA	-	-
N22	DL-Menthol	D	1000		CO	PHH	-	-	-	-	-	-
N23	Benzoin	C	500		5%AGS	NO	-	-	-	NA	-	-
N24	1-Chloro-2-propanol	F	100		DW	NO	-	-	-	+	+	-
V01	4-Aminoazobenzene	C	50		CO	NO	-	-	+	+	-	-
V02	Carbon tetrachloride	C	50		CO	SCN	-	-	+	+	-	-
V03	Dichlorodiphenyl- trichloroethane	A	25		CO	DHH	-	-	+	+	-	-

(Continued)



Table 1. (Continued)

Test no. ^a	Name	Source ^b	Toxicity test		Vehicle ^d	Histo-path. ^e	Carcinogenic properties			Carcinogen group no. ⁱ
			Dose ^c				Car. ^f	Hepato-car. ^g	Muta. ^h	
V04	Acetaminophen	C	700		1%CMC + tw80	NO	+	+	-	
V05	2-Nitro-p-phenylenediamine	A	100		1%CMC	NO	+	-	+	
V06	1-Nitropyrene	A	5		CO	NO	+	-	+	
V07	Dieldrin	A	0.3		CO	NO	+	-	-	
V08	Methyl-N'-nitro-N-nitrosoguanidine (MNNG)	B	0.05		DW	NO	+	-	+	
V09	Methyleugenol	C	40		0.5%MC	NO	+	+	-	
V10	o-Nitrotoluene	B	300		CO	NO	+	+	-	
V11	Tris(2-chloroethyl)phosphate	C	88		CO	NO	+	-	-	
V12	Quercetin	C	200		0.1%CMC	NO	-	-	+	
V13	4-Acetylaminofluorene	J	40		CO	NO	-	-	+	
V14	Glutaraldehyde	C	50		DW	NO	-	-	+	
V15	4'-(Chloroacetyl)acetanilide	A	250		CO	NO	-	-	+	
V16	3-Chloro-p-toluidine	B	300		CO	AH, PHH	-	-	-	

Notes: ^aT, Group-1 carcinogen used for training set; C, other group carcinogen; N, non-carcinogen; V, validation set including carcinogen and non-carcinogen; ^bA, Sigma-Aldrich Co. (St. Louis, MO); B, Tokyo Chemical Co., Ltd. (Tokyo, Japan); C, Wako Pure Chemical Industries, Ltd. (Osaka, Japan); D, Junsei Chemical Co., Ltd (Tokyo, Japan); E, Kishida Chemical Co., Ltd. (Osaka, Japan); F, Fluka Chemical Co. (Buchs, Switzerland); G, Kanto Chemical Co., Inc. (Tokyo, Japan); H, Nard Institute, Ltd (Hyogo, Japan); I, AccuStandard Inc. (New Haven, CT); J, Lancaster Synthesis, Inc. (Windham, NH); ^cmg/kg/day; ^d5% AGS; 5.0 w/v% gum Arabic solution, CO; Corn oil, DW; Distilled water, 1%CMC; 1% carboxymethylcellulose sodium solution; 0.5%MC; 0.5% methylcellulose solution. ^eHistopathology: AH, Apoptosis of hepatocytes; ATH, Atrophy of hepatocytes; DHH, Diffuse hypertrophy of hepatocytes; ECH, Eosinophilic change of hepatocytes; FIC1, Focal inflammatory cell infiltrates in liver; GI, glycogen increment; IMF, increment of mitotic figure in hepatocytes; HHN, Hypertrophy of hepatocyte nuclei; NO, No histological abnormalities; PHH, Periportal hypertrophy of hepatocytes; SCN, Single cell necrosis of hepatocytes; VH, Vacuolization of hepatocytes; ^fCarcinogenicity; ^gHepato-carcinogenicity; E, Equivocal; NA, Data not available; LP, Limited positive; ^hMutagenicity; ⁱResults based on those of a previous study (Matsumoto et al., 2009).

**Table 2.** Number of chemicals in each subgroup.

Subgroup name	# of chemicals
Group-1 carcinogen	17
Group-2 carcinogen	23
Group-3 carcinogen	6
Non-carcinogen	24
Validation	16
Total	86

(Fig. 1); non-redundant random sampling was used to select 11 carcinogens and 11 non-carcinogens from the training dataset (17 training Group-1 carcinogens and 24 training non-carcinogens), and a working linear formula was built by using the classification mode of SVMlight (<http://svmlight.joachims.org/>). SVMlight generated the distance of each test sample data from the hyperplane in 5 dimensional space as a linear function of the gene expression changes (log2ratio values) of the 5 predictive genes. The linear function was used as the working formula, where the distance from hyperplane was defined as the prediction score, with positive values for carcinogens

and negative values for non-carcinogens. The random sampling and calculations were performed 3000 times to generate 3000 working linear formulas. The medians of the coefficients for each gene and the median intercept value were then calculated and used as the coefficients in the final prediction formula.

Validation of the performance of the prediction formula

To validate the prediction performance, the formula was applied to the data from the Group-1 training chemicals, the Group-2 and Group-3 carcinogens in the previous study,¹⁰ and the validation chemicals. Also, the prediction formula was applied to dose-response data to elucidate the dose response of the prediction performance.

Results

Toxicity studies

For 28 days, male F344 rats were treated daily with either minimum carcinogenic doses of carcinogens or maximum tolerated doses of non-carcinogens,

Table 3. Chemicals of each carcinogen group.

Test no. ^a	Name	Test no. ^a	Name	Test no. ^a	Name
(a) Group-1					
T01	Diethylnitrosoamine	T07	2-Acetylaminofluorene	T14	1,4-Dioxane
T02	<i>N</i> -Nitrosodimethylamine	T08	MelQx	T15	Methyl carbamate
T03	<i>N</i> -Nitrosomorpholine	T09	Furan	T16	Thioacetamide
T04	<i>N</i> -Nitrosopiperidine	T10	Quinoline	T17	Urethane
T05	2-Nitropropane	T11	2,4-Diaminotoluene		
T06	3'-Methyl-4-dimethylamino-azobenzene	T12	Methapyrilene HCl		
		T13	Acetamide		
(b) Group-2					
C01	<i>N</i> -Ethyl- <i>N</i> -nitrosoarea	C08	3-Methylcholanthrene	C17	1,4-Dichlorobenzene
C02	4-Nitroquinoline-1-oxide	C09	Clofibrate	C18	Hexachlorobenzene
C03	4-Dimethylaminoazobenzene	C10	Di(2-ethylhexyl)adipate	C19	Alpha-Hexachloro-cyclohexane
C04	2-Amino-1-methyl-6-phenyl-imidazo[4,5- <i>b</i>]pyridine(PhIP)	C11	Di(2-ethylhexyl)phthalate	C20	Trichloroethylene
C05	Safrole	C12	Phenytoin	C21	Tetrachloroethylene
C06	Benzo[<i>a</i>]pyrene	C13	Butylated hydroxyanisole	C22	Trichloroacetic acid
C07	7,12-Dimethylbenz[<i>a</i>]-anthracene	C14	<i>d</i> -Limonene	C23	DL-Ethionine
		C15	Aldrin		
		C16	Chlorendic acid		
(c) Group-3					
C24	Benz[<i>a</i>]anthracene				
C25	Phenobarbital				
C26	Diethylstilbestrol				
C27	Ethinylestradiol				
C28	Chloroform				
C29	Pentachloroethane				

Notes: ^aT, Group-1 carcinogen used for training set; C, other group carcinogen.

**Table 4.** Gene symbol, Gene name, Refseq ID and expression direction of the predictive genes.

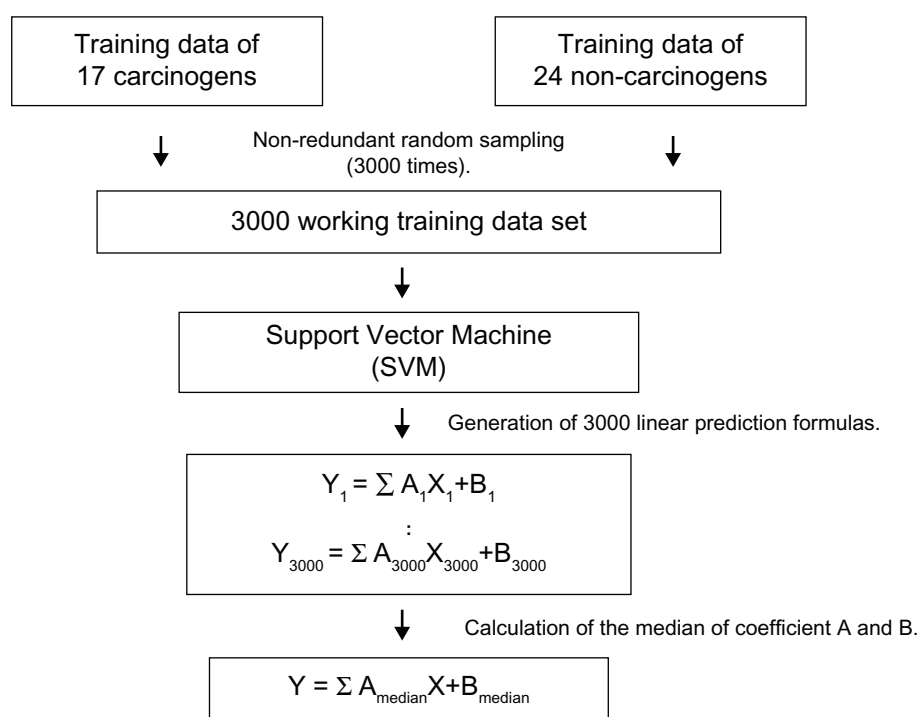
#	Gene symbol	Gene name	Refseq ID	Expression*
1	Ccng1	Cyclin G1	NM_012923	Up
2	Abcb1b	ATP-binding cassette, sub-family B (MDR/TAP), member 1B	NM_012623	Up
3	Mgmt	O-6-methylguanine-DNA methyltransferase	NM_012861	Up
4	Pbsn	Probasin	NM_019125	Up
5	Inmt	Indolethylamine N-methyltransferase	NM_001109022	Down

Note: *Up, Log2ratio of the gene was positive (ie, expression up-regulated by administration of Group 1 carcinogens); Down, log2ratio of the gene was negative (ie, expression down-regulated by administration of Group 1 carcinogens).

and then subjected to histopathological examination. For 15 of the 23 hepatocarcinogens, histological abnormalities such as modest hypertrophy of hepatocytes and enlargement of hepatocyte nuclei were observed 28 days post-treatment. Generally, the histological abnormalities observed were those expected after treatment with genotoxic carcinogens (Table 1). These findings support the assumption that the animals would have developed hepatic tumours if administration of carcinogens had continued.¹⁰

Predictive gene selection

Ideally, a single marker gene that undergoes expression change after administration of each carcinogen, but not any of the non-carcinogens, would be identified and used to predict the carcinogenicity of the chemicals. We considered it unlikely that such a gene would be found in this study, because factors other than carcinogenesis might affect the expression of the gene, and multiple carcinogenic mechanisms might be promoted by the various Group-1 carcinogens. Therefore in this



where Y (prediction score) is the distance of test data from the hyperplane, X is the gene expression change, A is the coefficient and B is the intercept

Figure 1. Flow diagram showing the generation of the prediction formula. From 17 carcinogens and 24 non-carcinogens, 3000 non-redundant sets of 11 carcinogens and non-carcinogens each were randomly selected, and the gene expression data of the predictive genes for those chemical was used as the training data sets to generate 3000 linear prediction formulas by SVM. The median value of the coefficients of 3000 linear formulae was set as the coefficient of the final prediction formula.



study, we used multiple genes as predictive genes in building the prediction formula for carcinogenicity.

The difference of carcinogen group and the non-carcinogen group which may have an unequal change is assessed in Welch's t-value that is the statistics of the Welch's t-test. We employed the t-value obtained as the first criterion to identify "characteristic" Group-1 carcinogen-responsive genes. However, if the variances of the gene expression changes were very small for both administration of carcinogen and administration of non-carcinogen, a gene could meet the first criterion without the expression change being of sufficient magnitude to be biologically significant. Consequently a second criterion was introduced, whereby a substantial gene expression change was observed after administration of more than a certain proportion of the carcinogens and less than a certain proportion of the non-carcinogens.

To determine the appropriate threshold values for the two criteria above, prediction formulae were built by using various "characteristic" gene sets obtained by varying the absolute value of the t-value threshold from 2 to 5 (criterion 1) and varying the proportions of carcinogens (>50% to >80%) and non-carcinogens (<30%) inducing substantial gene expression changes (criterion 2), where a substantial gene expression change was defined as an absolute $\log_2\text{ratio} > 0.8$. Using absolute t-values of >2, >3, >4 and >5 for the first criterion, 608, 250, 69, and 14 genes, respectively, were selected. These were reduced to 17, 14, 10 and 5 genes, respectively, when a second criterion of substantial gene expression change in >50% carcinogens and <30% non-carcinogens was added. These results indicate that the majority of the genes that showed substantial expression-change differences between the training carcinogens and non-carcinogens did not show substantial gene expression changes in response to a large proportion of the carcinogens, suggesting that both criteria are needed to select an efficient gene set for building the prediction formula for carcinogenicity. When the absolute t-value was set at >2 and the proportion of carcinogens was varied from >50% to >80%, the number of selected genes was from 17 to 2, and the performance of the prediction formula reached a maximum at >70% of the proportion. Similar results were obtained with other absolute t-values, although the predictive score increased slightly with increasing absolute

t-value. To maximize the prediction performance, we selected 5 Group-1 "characteristic" genes with absolute t-value >5, and substantially altered expression in response to >70% carcinogens as the final predictive genes (Table 4). Of these 5 genes, 4 were up-regulated and 1 was down-regulated by exposure to Group-1 carcinogens.

Development and validation of the prediction formula

Three thousand working linear prediction formulae were built by using the 5 predictive genes selected above and 3000 working training data sets of 11 carcinogens and 11 non-carcinogens, extracted at random from the total training data (17 carcinogens and 24 non-carcinogens). Within the working formulae, the coefficients for each gene, the intercept values and the concordance rates of 3000 training data sets tended towards a normal distribution as the number of formulae increased. In 3000 prediction formulae, the variance of the gene coefficients ranged from 8.3×10^{-4} to 4.2×10^{-2} , the variance of the intercept values was 5.6×10^{-3} , and the average concordance rate was 98.8% (standard deviation = $\pm 2.0\%$).

The final prediction formula was built by using the medians of the gene coefficients and intercept values of the 3000 working linear prediction formulae; the formula was then applied separately to the training and the validation data to assess its prediction performance (Fig. 2). The prediction scores of the 17 Group-1 carcinogens and 24 non-carcinogens used as training data were positive and negative, respectively; although in the case of two training carcinogens, quinoline (T10) and urethane (T17), the error bar (Standard deviation) extended to negative values. Thus, the concordance of the prediction outcome of the training data was 100%.

All Group-2 carcinogens were predicted as negative carcinogenicity, with the exception of safrole (C05), which had a positive score. All Group-3 carcinogens were predicted as negative carcinogenicity.

Three out of 5 non-carcinogens in validation chemicals were correctly predicted as negative carcinogenicity, but remaining two were judged as positive carcinogenicity, those being 4'-(chloroacetyl) acetanilide (V15), and 3-chloro-p-toluidine (V16). Three (carbon tetrachloride [V02], methyleugenol

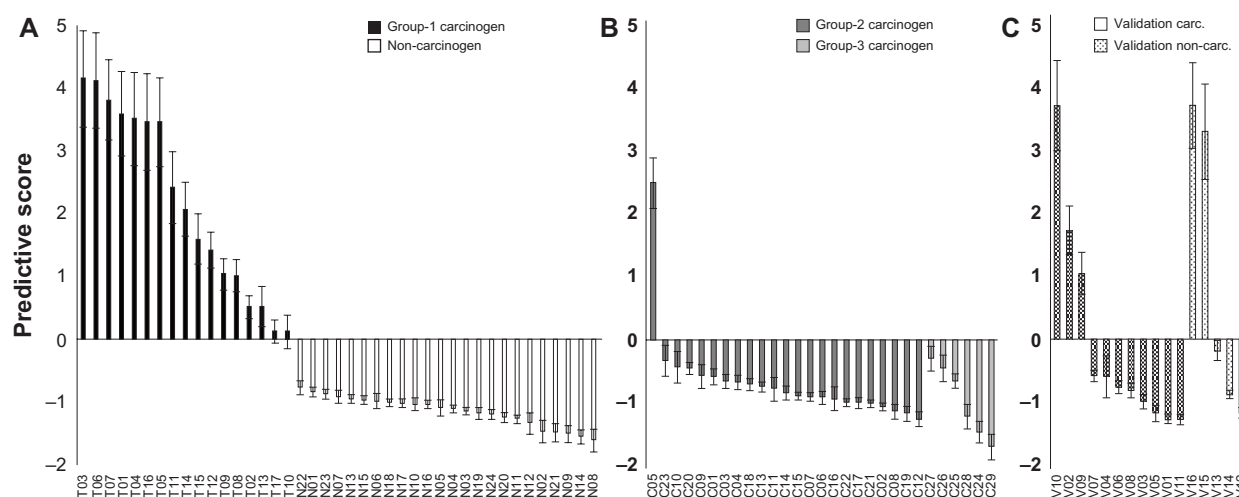


Figure 2. Comparison of scores of training and validating chemicals predicted by the prediction formula for Group-1 carcinogens. (A) Training Group-1 carcinogens and non-carcinogens, (B) Group-2 and -3 chemicals used in the previous clustering analysis (Matsumoto et al, 2009), and (C) validation carcinogens and non-carcinogens. Data represent the mean \pm standard deviation of prediction scores derived from 3000 working prediction formulae. The names of the compounds (T01 to N14, C05 to C29, and V10 to V12) are listed in Table 1.

[V09] and o-nitrotoluene [V10]) of the 11 carcinogens in carcinogens had positive scores, indicating that these chemicals are possible Group-1 carcinogens.

Dose response of the prediction score

Dose-response studies were conducted with four Group-1 carcinogens (diethylnitrosoamine [T01], N-nitrosomorpholine [T03], 1,4-dioxane [T14],

and thioacetamide [T16]) and two non-carcinogens (lithocholic acid [N15] and alpha-tocopherol [N19]) selected from the Group-1 carcinogens and non-carcinogens at random. The prediction formula was applied to the resultant data, and prediction scores were estimated (Fig. 3). The prediction score for the highest dose of diethylnitrosoamine (T01; 20 mg/kg/day) and N-nitrosomorpholine (T03; 10 mg/kg/day) could not be estimated because the animals died

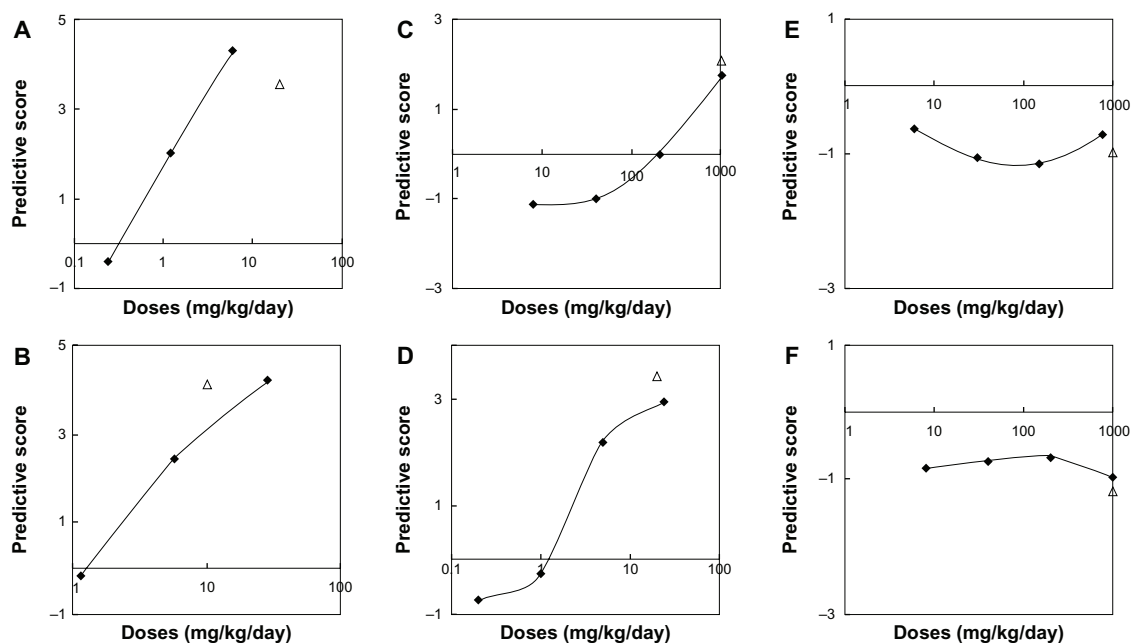


Figure 3. Dose-response relationship of prediction scores of four Group-1 carcinogens and two non-carcinogens. (A) Diethylnitrosamine (T01), (B) N-nitrosomorpholine (T03), (C) 1,4-dioxane (T14), (D) thioacetamide (T16), (E) lithocholic acid (N15) and (F) alpha-tocopherol (N19). The triangle marks show prediction score of training data of each chemical.



earlier. For all compounds tested, the prediction score was nearly coincident with the score obtained from the training data at the same dose, supporting the high reproducibility of the prediction scores. In the predictive gene set, the change in gene expression (log₂ ratio) increased for the 4 up-regulated genes and decreased for the 1 down-regulated genes with increasing doses of carcinogens, with the exception of Pbsn at the highest dose of thioacetamide. Consequently, the prediction score was negative at the lowest dose and increased to a positive value with increasing doses in the case of all four carcinogens. In contrast, no gene expression change depending on the administration was observed, and the prediction score remained a relatively constant negative value with increasing doses in the case of the two non-carcinogens. The intersection doses (DS₀; doses at which the predictive score curves crossed score = 0), were 0.44, 1.38, 192, and 1.28 mg/kg/day for diethylnitrosamine (T01), N-nitrosomorpholine (T03), 1,4-dioxane (T14), and thioacetamide (T16), respectively. We compared the DS₀ values with the tumorigenic doses (TD₅₀) values reported for the same chemicals (ie, 0.0265, 0.109, 267, and 11.5 mg/kg/day respectively) in the Carcinogenic Potency Database of Berkeley University (CPDB; <http://potency.berkeley.edu/>). The double logarithmic plot of TD₅₀ vs. DS₀ was approximated by the linear equation $TD_{50} = 1.3285 \times DS_0 - 0.485$ ($r = 0.85$).

Discussion

Prediction formula performance

Usually, a prediction formula is built by maximizing the prediction performance of a training data set, and therefore the prediction performance is strongly dependent on the training data set. If atypical data are mixed in the training data, this can lower the prediction performance of the resultant formula when it is applied to more typical data. Therefore, to minimize the influence of atypical data in the training data set, and to build a prediction formula that is widely applicable, we built 3000 prediction formulas from 3000 working training data sets extracted from the training data set at random. The medians of the coefficients and intercept values of the 3000 working linear prediction formulas were used to the final formula. This final formula is not over-fitting to the training data set. Rather, it was the

most frequently emerging formula in the training set and was therefore expected to give the most plausible prediction result.

The final prediction formula was able to correctly predict the carcinogenicity of all training carcinogens and non-carcinogens, though the predictive scores of two carcinogens, quinoline (T10) and urethane (T17), were smaller than the standard deviation. The dosage of quinoline used here (25 mg/kg/day) was similar to the dosage (22.3 mg/kg/day) that promoted tumors in rats in a previous study,¹³ and it is possible to take into account dose-response of the predictive score that the gene expression change specific to Group-1 carcinogen become faint. In contrast, the dosage of urethane used here (80 mg/kg/day), was much higher than the lowest dosage (1.1 mg/kg/day) that promoted tumors in a previous study,¹⁴ so the small predictive score cannot be explained by the dosage amount.

Importantly, six non-mutagenic carcinogens (methapyrilene HCl [T12], acetamide [T13], 1,4-dioxane [T14], methyl carbamate [T15], thioacetamide [T16] and urethane [T17]) were correctly predicted by the prediction formula, indicating that this method has advantages over other assessment such as Ames test in the prediction of carcinogenicity.

The prediction scores of Group-2 and Group-3 carcinogens had negative values, with the exception of safrole (C05) from Group-2. These results indicate that the prediction formula is specific to Group-1 carcinogens. The safrole predicted as Group-1 carcinogens were in the same cluster as Group-1 carcinogens under some conditions in hierarchical clustering analysis (data not shown), suggesting that these carcinogens share features of the gene expression profile of both groups and that multiple processes occur concurrently in carcinogenesis.

In the 11 validation carcinogens, the three non-mutagenic carcinogens carbon tetrachloride (V02), methyleugenol (V09) and o-nitrotoluene (V10) were predicted as Group-1 carcinogens. This result was confirmed by hierarchical cluster analysis, in which these chemicals resided in the Group-1 cluster.

The two non-carcinogens 3-chloro-p-toluidine (V16) and 4'-(chloroacetyl)acetanilide (V15) which were not included in the training data because they previously showed gene expression changes similar to Group-1 carcinogens,¹⁰ were predicted as carcinogens. The dosages of 3-chloro-p-toluidine



and 4'-(chloroacetyl)acetanilide used here (300 and 250 mg/kg/day, respectively), were at least 3 times the maximum dosages (100 and 67.6 mg/kg/day, respectively) used in previous carcinogenicity tests.^{15,16}

Dose-response relationship of the prediction score

For all Group-1 carcinogens examined, the prediction score decreased from positive to negative values with decreasing doses, and the DS0 values were highly correlated with TD50 values previously published (CPDB; <http://potency.berkeley.edu/line>). These results indicate that the gene expression changes specific to Group-1 carcinogens emerge above the specific dosage amount dependent on carcinogens and the specific dosage can be used as indicator of TD50 value of the test carcinogen.

Function of the predictive genes

For the predictive genes, we selected 5 genes with expression changes specific to Group-1 carcinogens.

Abcb1b. The gene encoding Abcb1b (ATP-binding cassette, sub-family B (MDRTAP), member 1) showed the largest t-value in the Group-1 carcinogen vs. non-carcinogen comparison, and it showed substantially altered expression following administration of each of the Group-1 carcinogens; hence, it contributed the most to the prediction score. The Abcb1b gene and its product, P-glycoprotein, which functions as a drug efflux transporter, show increased expression in rat liver after administration of carcinogens such as 2-acetylaminofluorene (T07; a Group-1 carcinogen), its metabolite N-hydroxy-acetylaminofluorene, and aflatoxine B1¹⁷ and it is also demonstrated that P-glycoprotein was associated with a more progressed phenotype of the liver malignancy.¹⁸ Although Abcb1b mRNA and P-glycoprotein are up-regulated by N-hydroxy-acetylaminofluorene and aflatoxine B1, they do not confer resistance to these chemicals in NIH 3T3-mdr1b cells, this is unlike the resistance to many cytotoxic drugs that is conferred by their transport by P-glycoprotein, and induction of the Abcb1b gene may result from an increase in transcription factors responsive to DNA damage induced by these carcinogens.¹⁹

Mgmt. The Mgmt gene encodes an enzyme involved in the DNA repair of O(6)-alkylguanine, which is the

major mutagenic and carcinogenic lesion in DNA. The Mgmt gene was substantially up-regulated by administration of most Group-1 carcinogens, including two non-mutagenic carcinogens, 1,4-dioxane and thioacetamide, both of which increased the expression of the Mgmt gene by 290%. The carcinogen 1,4-dioxane has been found to be non-mutagenic in 5 in vitro assays, including a Salmonella assay.²⁰ DNA damage has been observed in rat liver after a single oral administration of 1,4-dioxane (2550 mg/kg)²¹; however, neither DNA damage nor DNA repair was observed in the livers of F344 rats after a single oral administration of 1,4-dioxane (1000 mg/kg).²² Here, repeated doses test were performed over 28 days. No substantial change in Mgmt gene expression was observed for all non-carcinogens, nine of which were mutagenic. For example, two structural isomers (2,4-diaminotoluene [T11] and 2,6-diaminotoluene [N03]) were equally mutagenic in a Salmonella assay but differed in terms of carcinogenicity, and altered Mgmt gene expression was observed only after exposure to the carcinogenic isomer, 2,4-diaminotoluene. The difference in the Mgmt gene response to these structural isomers might be explained if only 2,4-diaminotoluene (T11) were mutagenic in the livers of rats, owing to differences in the metabolic pathways of 2,6-diaminotoluene (N03) being present in vivo compared with in vitro.²³

Ccng1. Ccng1 (cyclin G1) is one of the target genes of the transcription factor p53 and is induced in a p53-dependent manner in response to DNA damage. It plays roles in G2/M arrest, damage recovery, and growth promotion after cellular stress.²⁴

In summary, we developed a new gene-expression-based prediction method for carcinogenicity of Group-1 carcinogens, as a model case. Our final prediction formula used the data from 5 Group-1 carcinogen-responsive genes to correctly predict the carcinogenicity of Group-1 carcinogens regardless of mutagenicity. Advantages of this method are that the reliability of the prediction can be quantitatively evaluated by the prediction score value, and the TD50 value of chemicals might be estimated by the response of the prediction score. The prediction formula built here can be applied only to Group-1 carcinogens, which constitute only 37% of the carcinogens tested in our previous study,¹⁰ but the same method can be applied to other groups of carcinogens. Therefore, we are currently developing similar prediction



formulae for Group-2 and Group-3 carcinogens, so as to be able to predict all types of carcinogens.

Disclosures

Author(s) have provided signed confirmations to the publisher of their compliance with all applicable legal and ethical obligations in respect to declaration of conflicts of interest, funding, authorship and contribution, and compliance with ethical requirements in respect to treatment of human and animal test subjects. If this article contains identifiable human subject(s) author(s) were required to supply signed patient consent prior to publication. Author(s) have confirmed that the published article is unique and not under consideration nor published by any other publication and that they have consent to reproduce any copyrighted material. The peer reviewers declared no conflicts of interest.

References

1. McCann J, Choi E, Yamasaki E, Ames BN. Detection of carcinogens as mutagens in the Salmonella/microsome test: assay of 300 chemicals. *Proc Natl Acad Sci U S A*. 1975;72:5135–9.
2. Zeiger E. Identification of Rodent carcinogens and noncarcinogens using genetic toxicity tests: premises, promises, and performance. *Regul Toxicol Pharmacol*. 1998;28:85–95.
3. Elcombe CR, Odum J, Foster JR, et al. Prediction of rodent nongenotoxic carcinogenesis: evaluation of biochemical and tissue changes in rodents following exposure to nine nongenotoxic NTP carcinogens. *Environ Health Perspect*. 2002;110:363–75.
4. Ito N, Tsuda H, Tatematsu M, et al. Enhancing effect of various hepatocarcinogens on induction of preneoplastic glutathione S-transferase placental form positive foci in rats—an approach for a new medium-term bioassay system. *Carcinogenesis*. 1988;9:387–94.
5. Ito N, Tamano S, Shirai T. A medium-term rat liver bioassay for rapid in vivo detection of carcinogenic potential of chemicals. *Cancer Sci*. 2003;94:3–8.
6. Sawada H, Takami K, Asahi S. A toxicogenomic approach to drug-induced phospholipidosis: analysis of its induction mechanism and establishment of a novel in vitro screening system. *Toxicol Sci*. 2005;83:282–92.
7. Zidek N, Hellmann J, Kramer PJ, Hewitt PG. Acute hepatotoxicity: a predictive model based on focused illumina microarrays. *Toxicol Sci*. 2007;99:289–302.
8. Mathijs K, Brauers KJ, Jennen DG, et al. Discrimination for genotoxic and nongenotoxic carcinogens by gene expression profiling in primary mouse hepatocytes improves with exposure time. *Toxicol Sci*. 2009;112:374–84.
9. Ellinger-Ziegelbauer H, Gmuender H, Bandenburg A, Ahr HJ. Prediction of a carcinogenic potential of rat hepatocarcinogens using toxicogenomics analysis of short-term in vivo studies. *Mutat Res*. 2008;637:23–39.
10. Matsumoto H, Yakabe Y, Saito K, et al. Discrimination of carcinogens by hepatic transcript profiling in rats following 28-day administration. *Cancer Inform*. 2009;7:253–69.
11. Man TK, Chintagumpala M, Visvanathan J, et al. Expression profiles of osteosarcoma that can predict response to chemotherapy. *Cancer Res*. 2005;65:8142–50.
12. Wagner M, Naik DN, Pothan A, et al. Computational protein biomarker prediction: a case study for prostate cancer. *BMC Bioinformatics*. 2004;5:26.
13. Hasegawa R, Furukawa F, Toyoda K, Sato H, Imaida K, Takahashi M. Sequential analysis of quinoline-induced hepatic hemangioendothelioma development in rats. *Carcinogenesis*. 1989;10:711–6.
14. Schmähl D, Port R, Wahrendorf J. A dose-response study on urethane carcinogenesis in rats and mice. *Int J Cancer*. 1977;19:77–80.
15. National Cancer Institute (NCI). Bioassay of 3-chloro-p-toluidine for possible carcinogenicity, CAS No. 95-74-9. *Natl Cancer Inst Carcinog Tech Rep Ser*. 1978;145:1–100.
16. National Cancer Institute (NCI). Bioassay of 4'-(chloroacetyl)-acetanilide for possible carcinogenicity, CAS No. 140-49-8. *Natl Cancer Inst Carcinog Tech Rep Ser*. 1979;177:1–103.
17. Hill BA, Brown PC, Preissegger KH, Silverman JA. Regulation of *mdr1b* gene expression in Fischer, Wistar and Sprague-Dawley rats in vivo and in vitro. *Carcinogenesis*. 1996;17:451–7.
18. Bradley G, Sharma R, Rajalakshmi S, Ling V. P-glycoprotein expression during tumor progression in the rat liver. *Cancer Res*. 1992;52:5154–61.
19. Santoni-Rugiu E, Silverman JA. Functional characterization of the rat *mdr1b* encoded P-glycoprotein: not all inducing agents are substrates. *Carcinogenesis*. 1997;18:2255–63.
20. Morita T, Hayashi M. 1,4-Dioxane is not mutagenic in five in vitro assays and mouse peripheral blood micronucleus assay, but is in mouse liver micronucleus assay. *Environ Mol Mutagen*. 1998;32:269–80.
21. Kitchin KT, and Brown JL. Is 1,4-dioxane a genotoxic carcinogen? *Cancer Lett*. 1990;53:67–71.
22. Goldworthy TL, Monticello TM, Morgan KT, et al. Examination of potential mechanisms of carcinogenicity of 1,4-dioxane in rat nasal epithelial cells and hepatocytes. *Arch Toxicol*. 1991;65:1–9.
23. Toyoda-Hokaiwado N, Inoue T, Masumura K, et al. Integration of in vivo genotoxicity and short-term carcinogenicity assays using F344 gpt delta transgenic rats: in vivo mutagenicity of 2,4-diaminotoluene and 2,6-diaminotoluene structural isomers. *Toxicol Sci*. 2010;114:71–8.
24. Kimura SH, Ikawa M, Ito A, Okabe M, Nojima H. Cyclin G1 is involved in G2/M arrest in response to DNA damage and in growth control after damage recovery. *Oncogene*. 2001;20:3290–300.

Publish with Libertas Academica and every scientist working in your field can read your article

“I would like to say that this is the most author-friendly editing process I have experienced in over 150 publications. Thank you most sincerely.”

“The communication between your staff and me has been terrific. Whenever progress is made with the manuscript, I receive notice. Quite honestly, I've never had such complete communication with a journal.”

“LA is different, and hopefully represents a kind of scientific publication machinery that removes the hurdles from free flow of scientific thought.”

Your paper will be:

- Available to your entire community free of charge
- Fairly and quickly peer reviewed
- Yours! You retain copyright

<http://www.la-press.com>