

RESEARCH

Open Access



Four functional genotoxic marker genes (*Bax*, *Btg2*, *Ccng1*, and *Cdkn1a*) discriminate genotoxic hepatocarcinogens from non-genotoxic hepatocarcinogens and non-genotoxic non-hepatocarcinogens in rat public toxicogenomics data, Open TG-GATEs

Chie Furihata^{1,2*}  and Takayoshi Suzuki³

Abstract

Background Previously, Japanese Environmental Mutagen and Genome Society/Mammalian Mutagenicity Study Group/Toxicogenomics Study Group (JEMS/MMS toxicogenomic study group) proposed 12 genotoxic marker genes (*Aen*, *Bax*, *Btg2*, *Ccnf*, *Ccng1*, *Cdkn1a*, *Gdf15*, *Lrp1*, *Mbd1*, *Phlda3*, *Plk2*, and *Tubb4b*) to discriminate genotoxic hepatocarcinogens (GTHCs) from non-genotoxic hepatocarcinogens (NGTHCs) and non-genotoxic non-hepatocarcinogens (NGTNHCs) in mouse and rat liver using qPCR and RNA-Seq and confirmed in public rat toxicogenomics data, Open TG-GATEs, by principal component analysis (PCA). On the other hand, the U.S. Environmental Protection Agency (US EPA) suggested seven genotoxic marker genes (*Bax*, *Btg2*, *Ccng1*, *Cgrf1*, *Cdkn1a*, *Mgmt*, and *Tmem47*) with Open TG-GATEs data. Four genes (*Bax*, *Btg2*, *Ccng1*, and *Cdkn1a*) were common in these two studies. In the present study, we examined the performance of these four genes in Open TG-GATEs data using PCA.

Results The study's findings are of paramount significance, as these four genes proved to be highly effective in distinguishing five typical GTHCs (2-acetylaminofluorene, aflatoxin B1, 2-nitrofluorene, *N*-nitrosodiethylamine and *N*-nitrosomorpholine) from seven typical NGTHCs (clofibrate, ethanol, fenofibrate, gemfibrozil, hexachlorobenzene, phenobarbital, and WY-14643) and 11 NGTNHCs (allyl alcohol, aspirin, caffeine, chlorpheniramine, chlorpropamide, dexamethasone, diazepam, indomethacin, phenylbutazone, theophylline, and tolbutamide) by PCA at 24 h after a single administration with 100% accuracy. These four genes also effectively distinguished two typical GTHCs (2-acetylaminofluorene and *N*-nitrosodiethylamine) from seven NGTHCs and ten NGTNHCs by PCA on 29 days after 28 days-repeated administrations, with a similar or even better performance compared to the previous 12 genes. Furthermore, the study's analysis revealed that the three intermediate GTHC/NGTHCs (methapyrilene, monocrotaline,

*Correspondence:

Chie Furihata

chiefurihata@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

and thioacetamide, which were negative in the *Salmonella* test but positive in the in vivo rat liver test) were located in the intermediate region between typical GTHCs and typical NGTHCs by PCA.

Conclusions The present results unequivocally demonstrate the availability of four genotoxic marker genes (*Bax*, *Btg2*, *Ccng1*, and *Cdkn1a*) and PCA in discriminating GTHCs from NGTHCs and NGTNHCs in Open TG-GATEs. These findings strongly support our recommendation that future rat liver in vivo toxicogenomics tests prioritize these four genotoxic marker genes, as they have proven to be highly effective in discriminating between different types of hepatocarcinogens.

Keywords Toxicogenomics, Genotoxic marker genes, Open TG-GATEs, Genotoxic carcinogen, Non-genotoxic carcinogen, Non-carcinogen

Introduction

Carcinogenicity testing is important to identify carcinogens in environmental chemicals [1], chemicals in daily life [2], and pharmaceutical drug and pesticide development [3, 4]. However, evaluating carcinogenicity using conventional 2-year rodent-based animal studies is time-consuming and labor-intensive [5]. Thus, there is an increased need to develop novel alternative approaches to these rodent bioassays for assessing the carcinogenicity of substances [6].

Carcinogens have traditionally been classified into two categories according to their presumed mode of action: genotoxic (GTC) and non-genotoxic (NGTC). An OECD expert group defined GTCs as having the potential to induce cancer by interacting directly with DNA and the cellular apparatus involved in preserving the integrity of the genome. In contrast, NGTCs have the potential to induce cancer without interacting directly with either DNA or the apparatus mentioned above [7]. Genotoxic carcinogens are usually identified based on positive results in different in vitro and in vivo test systems. These include detecting DNA strand breaks (comet assay) [OECD TG489 (in vivo)], unscheduled DNA synthesis [OECD TG482 (in vitro), OECD TG486 (in vivo)], sister chromatid exchange [OECD TG479 (in vitro)], chromosomal aberrations [OECD TG473 (in vitro)], DNA adduct formation [8], mitotic recombination [9], and gene mutation [OECD TG476 (in vitro)]. Standard tests of mutagenicity include the Ames test [OECD TG471], in vitro metaphase chromosome aberration assay [OECD TG473], in vitro micronucleus assay [OECD TG487], L5178Y/tk± mouse lymphoma (thymidine kinase) gene mutation assay [OECD TG490], in vivo micronucleus assay in rodents [OECD TG474], and transgenic rodent mutation assay [OECD TG488]. NGTCs show a threshold for exerting hazardous effects, and the various authoritative bodies set guidelines regarding appropriate exposure levels like other hazardous substances. They recommended clear differences between threshold and non-threshold carcinogens dealing with carcinogen classification and risk assessment [10].

Toxicogenomics, the application of transcription profiling to toxicology, has been widely used for elucidating the molecular and cellular actions of chemicals and other environmental stressors on biological systems, predicting toxicity before any functional damages, and classification of known or new toxicants based on signatures of gene expression. The success of a toxicogenomics study depends upon close collaboration among experts in different fields, including a toxicologist or biologist, a bioinformatician, a statistician, a physician, and, sometimes, a mathematician [11].

Previously, the JEMS/MMS toxicogenomics study group has been conducting studies using DNA microarray [12] and qPCR [13–17] to discriminate genotoxic hepatocarcinogens from non-genotoxic hepatocarcinogens in mice [13, 14, 16] and in rats [15] for use in an in vivo short-term toxicogenomics screening test for genotoxic carcinogens. We proposed 12 genotoxic marker genes (*Aen*, *Bax*, *Btg2*, *Ccnf*, *Ccng1*, *Cdkn1a*, *Gdf15*, *Lrp1*, *Mbd1*, *Phlda3*, *Plk2* and *Tubb4b*) to discriminate GTHCs from NGTHCs in mouse study using eight mouse GTHCs from five different functional groups [aromatic amines: 2-acetamidofluorene and 2,4-diaminotoluene; azobenzene: 4-dimethylaminoazobenzene; ester of carbamic acid: urethane; heterocyclic aromatic compound: quinoline; and nitrosamines: diisopropanolnitrosamine, 4-(methylnitrosamino)-1-(3-pyridyl) 1-butanone, and *N*-nitrosomorpholine] and four mouse non-genotoxic hepatocarcinogens [chlorinated aromatic hydrocarbons: 1,4-dichlorobenzene and dichlorodiphenyltrichloroethane; phthalate: di(2-ethylhexyl) phthalate; and heterocyclic organic compound: furan] [14]. These seven mouse genotoxic hepatocarcinogens [18–24] except urethane [25] also induced hepatocellular carcinoma in rats. The results suggested that these carcinogens had similar modes of action in both rat and mouse liver. As described previously [14], nine (*Aen*, *Bax*, *Btg2*, *Ccng1*, *Cdkn1a*, *Gdf15*, *Mbd1*, *Phlda3*, and *Plk2*) of 12 marker genes are members of gene families related to the intrinsic apoptotic signaling pathway in response to DNA damage by the p53 class mediator. *Ccnf* may be related to

DNA repair and DNA damage [26]. *Lrp1* may be related to tumor growth and metastasis, particularly by modulating three extracellular tumor environments [27]. *Tubb4b* may be related to metastasis in colon cancer [28]. Table 2 shows the symbol, gene name, and gene ID of the 12 corresponding rat genes. We confirmed the 12 genes in public rat toxicogenomics data, Open TG-GATEs [29]. We studied differentially expressed protein-coding genes with targeted RNA-Seq on freshly frozen rat liver tissues [30] and on formalin-fixed paraffin-embedded (FFPE) rat liver tissues [31, 32] after 28 days of treatment with chemicals using PCA.

The present study examined data from rat [male Crl:CD Sprague–Dawley (SD) rat, 6-week-old] Open TG-GATEs with newly selected genes. The Open Japanese Toxicogenomics Project-Genomics Assisted Toxicity Evaluation System (Open TG-GATEs) was developed by the Japanese Toxicogenomics Project (TGP) consortium and opened to the public in 2015 (<http://toxico.nibiohn.go.jp/english/>) [33]. The data include 170 chemicals. The data contain concurrent vehicle controls, three rats per group, three doses, and various time points (3, 6, 9, and 24 h after a single administration and 4, 8, 15, and 29 days after repeated administrations).

The registered data include five typical GTHCs [2-acetamidofluorene (AAF), aflatoxin B1 (AFL), 2-nitrofluorene (2NF), *N*-nitrosodiethylamine (DEN) and *N*-nitrosomorpholine (NNM)] and seven typical NGTHCs [clofibrate (CLO), ethanol (ETH), fenofibrate (FEN), gemfibrozil (GEM), hexachlorobenzene (HEX), phenobarbital (PHE) and WY-14643 (WY)] for comparison. Among them, three genotoxic hepatocarcinogens (AFL: mycotoxin; 2NF: nitrated polycyclic aromatic hydrocarbon; and NNM: nitrosamine) and all seven NGTHCs are different from those in our previous study [14]. Seven NGTHCs were fibric acid derivatives: CLO and GEM; simple alcohol: ETH; synthetic phenoxo-isobutyric acid derivative: FEN; chlorinated aromatic hydrocarbon: HEX; barbituric acid derivative: PHE; and thioacetic acid derivative: WY, a peroxisome proliferator. Six NGTHCs, except HEX, have functional groups different from those of our four previous NGTHCs.

The IWGT-Toxicogenomics meeting was held at ICEM in Toronto in 2022, where 12 genotoxicity marker genes from JEMS/MMS and seven marker genes from EPA [34] were discussed. Four genes (*Bax*, *Btg2*, *Ccng1*, and *Cdkn1a*) were common in these two studies. Then, we analyzed the four genes in the rat public toxicogenomics data “Open TG-GATEs”.

We analyzed 23 previous chemicals (five GTHCs, seven NGTHCs, and eleven NGTNHCs) [29] and three intermediate GTHCs/NGTHCs. We designated that typical GTHCs are mutagenic in the *Salmonella* test and

genotoxic in some in vivo liver tests, such as the micronucleus test, the transgenic rodent mutation assay, and the UDS test and carcinogenic in rat liver. Typical NGTHCs are not mutagenic in the *Salmonella* test, do not show genotoxicity in in vivo rat liver tests, and are carcinogenic in rat liver. NGTNHCs are not mutagenic in the *Salmonella* test and are not carcinogenic in rat liver. Intermediate GTHC/NGTHCs are negative in the *Salmonella* test and show contradictory results with in vivo rat liver tests. Table 1 summarizes this study’s classification of GTHC, NGTHC, NGTNHC, and intermediate GTHC/NGTHC [35–77].

In the present study, we propose that the four genes (*Bax*, *Btg2*, *Ccng1*, and *Cdkn1a*) instead of the previous twelve genes (*Aen*, *Bax*, *Btg2*, *Ccnf*, *Ccng1*, *Cdkn1a*, *Gdf15*, *Lrp1*, *Mbd1*, *Phlda3*, *Plk2*, and *Tubb4b*) are functional to discriminate genotoxic hepatocarcinogens from non-genotoxic hepatocarcinogens and non-genotoxic non-hepatocarcinogens in rat liver. The present study strongly supports our recommendation that future rat liver in vivo toxicogenomics tests prioritize these four genotoxic marker genes, as they have proven highly effective in discriminating between different types of hepatocarcinogens.

Methods

Chemicals

We analyzed data of the 26 chemicals: five typical GTHCs, seven typical NGTHCs, 11 NGTNHCs [29], and three intermediate GTHC/NGTHCs from Open TG-GATEs. Table 1 summarizes this study’s classification of GTHC, NGTHC, NGTNHC, and intermediate GTHC/NGTHC [35–77]. Open TG-GATEs presented five typical GTHCs: 2-acetamidofluorene (CAS 53–96-3, AAF), aflatoxin B1 (CAS 1402–68-2, AFL, IARC Group 1), 2-nitrofluorene (CAS 607–57–8, 2NF, IARC Group 2B), *N*-nitrosodiethylamine (CAS 55–18-5, DEN, IARC Group 2 A), and *N*-nitrosomorpholine (CAS 59–89-2, NNM, IARC Group 2B) at 24 h after a single administration and two typical GTHCs: AAF and DEN at 29 days after repeated administration. Open TG-GATEs presented seven typical NGTHCs after at 24 h after a single administration and at 29 days after repeated administration: four PPAR α agonists [clofibrate (CAS 637–07-0, CLO, IARC Group 3), fenofibrate (CAS 49562–28-9, FEN), gemfibrozil (CAS 25812–30–0, GEM, IARC Group 3), and WY-14643 (CAS 50892–23-4, WY)], two enzyme inducers [hexachlorobenzene (CAS 118–74-1, HEX, IARC Group 2B) and phenobarbital (CAS 50–06-6, PHE, IARC Group 2B)] and ethanol (CAS 64–17-5, ETH, IARC Group 1), which induced oxidative stress. Open TG-GATEs presented the data of many NGTNHCs, but we selected 11 familiar chemicals in this study.

Table 1 Examined GTHCs, NGTHCs, NGTNHCs and intermediate GTHCs/NGTHCs in the present study

No	Classification	Chemical	Abbreviations	<i>Salmonella</i> test	Male rat in vivo liver test	Carcinogenicity in male rat liver	Carcinogenicity in male rat in other organs	IARC group ³⁵	References
1	GTHCs	2-acetylaminofluorene	AAF	+	MN +, UDS +	+	mgl, ski	NL	36
2		aflatoxin B1	AFB	+	UDS +	+	-	1	36, 37, 38
3		2-nitrofluorene	2NF	+	UDS +	+	kid, sto	2B	36, 39
4		<i>N</i> -nitrosodiethylamine	DEN	+	UDS +, DNA single strand scission +	+	eso, kid, vsc	2A	36, 40, 41
5		<i>N</i> -nitrosomorpholine	NNM	+	UDS +	+	kid, adr, thy, pit	2B	36, 41–43
6	NGTHCs	clofibrate	CLO	-	Pig-a -, UDS -, DSB -, MN +	+	pan, smi, dermatofibrosarcoma, tes	3	36, 44–47
7		ethanol	ETH	-	comet -, DNA strand break +	+	adr, pan, pit	1	36, 48, 49
8		fenofibrate	FEN	-	comet + (weak), ND (UDS, MN)	+	pan, tes	NL	50, 51
9		gemfibrozil	GEM	-	ND	+	adr, pan, tes	3	36, 52, 53
10		hexachlorobenzene	HEX	-	(Dominant lethal test -)	+	bile duct, kid	2B	36, 54
11	NGTNHCs	phenobarbital	PHE	-	MN -, comet -, Pig-a -	+	adr, nas	2B	36, 45, 55, 56
12		WY-14643	WY	-	UDS -	+	Leydig cell, pancreatic acinar cell	NL	57, 58
13		allyl alcohol	AA	-	ND	-	-	NL	36, 59
14		aspirin	ASP	-	ND	-	-	NL	36, 60
15		caffeine	CAF	-	ND	-	-	3	36, 61
16		chlorpheniramine	CPA	-	ND	-	-	NL	36, 62
17		chlorpropamide	CPP	-	ND	-	-	NL	36, 63
18		dexamethasone	DEX	-	ND	-	-	NL	36, 64
19		diazepam	DIA	-	ND	-	-	3	36, 65
20		indomethacin	IND	-	ND	-	-	NL	36, 66
21		phenylbutazone	PBZ	-	ND	-	-	3	36, 67
22	intermediate (GTHCs/NGTHCs)	theophylline	THE	-	ND	-	-	3	36, 68, 69
23		tolbutamide	TOL	-	ND	-	-	NL	36, 70
24		methapyrilene	MP	-	MN +, comet -, UDS -	+	-	NL	36, 55, 71, 72
25		monocrotaline	MCT	-	MN +, DNA-DNA interstrand crosslinks +,	+	-	2B	36, 73, 74
26		thioacetamide	TAA	-	comet +, γ -H2AX +, MN -	+	bile duct	2B	36, 75–77

DSB DNA strand breaks, ND Not determined by UDS, MN and comet assays, NL Not listed

adr adrenal gland, eso esophagus, kid kidney, mgl mammary gland, nas nasal cavity, pan: pancreas, pit pituitary gland, ski skin, smi small, sto stomach, tes testes, thy thyroid gland, vsc vascular system

Table 2 Twelve genes analyzed in the present study

No	Symbol	Gene name	Gene ID
1	Aen	apoptosis enhancing nuclease	361,594
2	Bax	BCL2 associated X, apoptosis regulator	24,887
3	Btg2	BTG anti-proliferation factor 2	29,619
4	Ccnf	cyclin F	117,524
5	Ccng1	cyclin G1	25,405
6	Cdkn1a	cyclin-dependent kinase inhibitor 1A	114,851
7	Gdf15	growth differentiation factor 15	29,455
8	Lrp1	LDL receptor related protein 1	299,858
9	Mbd1	methyl-CpG binding domain protein 1	291,439
10	Phlda3	pleckstrin homology-like domain, family A, member 3	363,989
11	Plk2	polo-like kinase 2	83,722
12	Tubb4b	tubulin, beta 4B class IVb	296,554

NGTNHCs do not include carcinogens to other organs than liver [36]. They are allyl alcohol (CAS 107–18-6, AA), aspirin (CAS 50–78-2, ASP), caffeine (CAS 58–08-2, CAF, IARC Group3), chlorpheniramine (CAS 113–92-8, CPA), chlorpropamide (CAS 94–20-2, CPP), dexamethasone (CAS 50–02-2, DEX), diazepam (CAS 439–14-5, DIA, IARC Group 3), indomethacin (CAS 53–86-1, IND), phenylbutazone (CAS 50–33-9, PBZ, IARC Group 3), theophylline (CAS 58–55-9, THE, IARC Group 3), and tolbutamide (CAS 64–77-7, TOL) at 24 h after a single administration. Open TG-GATEs examined 10 of the 11 NGTNHCs except DEX 29 days after repeated administration. Open TG-GATEs presented three intermediate GTHC/NGTHCs: methapyrilene (CAS 91–80-5, MP), monocrotaline (CAS 315–22-0, MCT, IARC Group 2B), and thioacetamide (CAS 62–55-5, TAA, IARC Group 2B) at 24 h after a single administration and at 29 days after repeated administration. They are negative in the *Salmonella* test [36] but positive in the in vivo rat liver test [71–77]. AAF was a metabolite of 2NF [78]. The IARC classification does not contain FEN, WY, AA, ASP, CPA, CPP, DEX, IND, MP, and TOL.

Analyzed genes

Table 2 shows the symbol, gene name, and NCBI gene ID of the 12 analyzed rat genes.

Data analysis

We obtained relative fluorescence intensities of DNA microarray results of three individual rats in each dose- and time-group on the four genes from Open TG-GATEs (<http://toxico.nibiohn.go.jp/english/>). Data from Open TG-GATEs were transferred into log2 data

using the “R Project for Statistical Computing” (<https://www.r-project.org/>) to stabilize the variance. Ratio (exp/cont) log2 was calculated against the mean from the control group (Appendix A.1, A.2, B1, B2, C1, and C2). Discrimination of GTHCs vs. NGTHCs plus NGTNHCs was achieved by statistical analysis using the PCA program from the “R Project for Statistical Computing” (Appendix A.3, A.4, B3, B4, C3, and C4) as described previously [29]. Data for each dose in supplements and each point in figures presented in this paper are averages for three rats.

Results

Four genotoxic marker genes discriminate typical GTHCs from typical NGTHCs and NGTNHCs at 24 h after a single administration and 29 days after repeated administrations
At 24 h after a single administration, as Open TG-GATEs presented five typical GTHCs (AAF, AFL, DEN, 2NF, and NNM) and seven typical NGTHCs (CLO, ETH, FEN, GEM, HEX, PHE, and WY), we analyzed five typical GTHCs, seven typical NGTHCs, and optional 11 NGTNHCs (AA, ASP, CAF, CPA, CPP, DEX, DIA, IND, PBZ, THE, and TOL), with three doses, 23 chemicals, 69 data points. Appendix A.1 shows the gene expression profile (log2), Appendix A.2 shows the first principal component (PC1) and the second principal component (PC2) analyzed by PCA, and Fig. 1(A) shows the result of PCA of five typical GTHCs, seven typical NGTHCs, and 11 NGTNHCs 24 h after a single administration in a two-dimensional figure. Five typical GTHCs are separated from seven typical NGTHCs and 11 NGTNHCs by PCA in a two-dimensional graph, with (PC1), where GTHCs exhibit PC1 below −0.719 (DEN-low), and NGTHCs exhibit PC1 above −0.292 (DIA-high) (Appendix A.2). PC1 of DEN was −0.719, −2.611, −4.885 for DEN-low, DEN-middle and DEN-high respectively, and reflected dose-dependent changes in gene expressions. The distinction between GTHCs and NGTHCs by the four genes is not inferior to, but somewhat superior to, the distinction by the 12 genes presented previously [reference 29, Fig. 4(C)].

At 29 days after repeated administration, as Open TG-GATEs presented two typical GTHCs and seven typical NGTHCs, we analyzed two typical GTHCs (AAF-low, AAF-middle, AAF-high, DEN-low, and DEN-middle), seven typical NGTHCs with three doses, and optional choice of the 10 NGTNHCs (AA, ASP, CAF, CPA, CPP, DIA, PBZ, THE, TOL with three doses, and IND-low and IND-middle), 19 chemicals, 55 data points. Appendix A.3 shows the gene expression profile (log2), Appendix A.4 shows the first principal component (PC1) and the second principal component (PC2) analyzed by PCA,

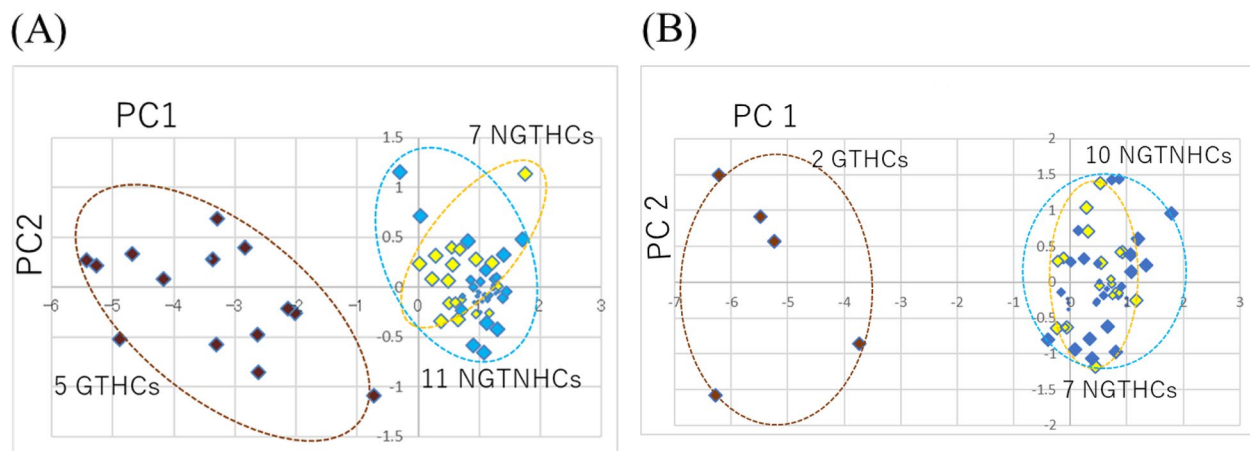


Fig. 1 **A** and **B** Analysis of rat liver toxicogenomics public data (OPEN TG-GATES, DNA microarray) by PCA with the four genes. Discrimination of GTHCs from NGTHCs and NGTNHCs at 24 h after a single administration [Fig. 1 (A)] and 29 days after repeated administrations [Fig. 2 (B)] by PCA with the four genotoxic marker genes (*Bax*, *Btg2*, *Ccng1*, and *Cdkn1*). The gene expression profile (log2) was calculated: the mean of each control group was calculated as 0 (log2), and the ratio (exp/cont) (log2) was calculated (Appendix A.1 for 24 h and Appendix A.3 for 29 days). These numerical values were analyzed by PCA (Appendix A.2 for 24 h and Appendix A.4 for 29 days). At 24 h, Fig. 1(A) shows the results of 23 chemicals with three doses, a total of 69 points: five typical GTHCs (brown-colored, AAF, AFL, DEN, NNM, and 2NF) were discriminated from seven typical NGTHCs (yellow-colored, CLO, ETH, FEN, GEM, HEX, PHE, and WY) and 11 NGTNHCs (blue-colored, AA, ASP, CAF, CPA, CPP, DEX, DIA, IND, PBZ, THE, and TOL). At 29 days, Fig. 1(B) shows the results of 19 chemical, total of 55 points: two typical GTHCs (brown-colored, AAF-low, -middle, and -high; DEN-low and -) were discriminated from seven typical NGTHCs (yellow-colored, CLO, ETH, FEN, GEM, HEX, PHE, and WY with three doses) and 10 NGTNHCs (blue-colored, AA, ASP, CAF, CPA, CPP, DIA, PBZ, THE, and TOL with three doses and IND-low and -middle). Each group is enclosed with an optional dashed ellipse. Five GTHCs [AAF: 2- acetamidofluorene, AFL: aflatoxin B1, 2NF: 2-nitrofluorene, DEN: *N*-nitrosodiethylamine and NNM: *N*-nitrosomorpholine], seven NGTHCs [CLO: clofibrate, ETH: ethanol, FEN: fenofibrate, GEM: gemfibrozil, HEX: hexa-chlorobenzene, PHE: phenobarbital, and WY: WY-14643], and 11 NGTNHCs (mostly pharmaceutical drugs) [AA: allyl alcohol, ASP: aspirin, CAF: caffeine, CPA: chlorpheniramine, CPP: chlorpropamide, DEX: dexamethasone, DIA: diazepam, IND: indomethacin, PBZ: phenylbutazone, THE: theophylline, and TOL: tolbutamide]

and Fig. 1(B) presents the result of PCA on two typical GTHCs, DEN and AAF, seven typical NGTHCs, and 10 NGTNHCs. Figure 1(B) demonstrates the discrimination of two typical GTHCs from seven typical NGTHCs and 10 NGTNHCs with PC1 by PCA in a two-dimensional graph, where GTHCs exhibit PC1 below -3.724 (DEN-low), and NGTHCs and NGTNHCs exhibit PC1 above -0.388 (CPA-high) (Appendix A.4). The distinction of GTHCs from NGTHCs plus NGTNHCs by four genes is no different from the distinction by 12 genes presented previously [reference 29, Fig. 4(F)].

Our research has uncovered a novel aspect of genotoxicity. We have identified four genotoxic marker genes that discriminate GTHCs from NGTHCs and NGTNHCs. This distinction is observed 24 h after a single administration and 29 days after repeated administration, significantly advancing our understanding of genetic mechanisms.

Three intermediate GTHC/NGTHCs are located in the intermediate region between typical GTHCs and typical NGTHCs by PCA with the four genes

In the current study, three additional intermediate GTHC/NGTHCs, methapyrilene (MP), monocrotaline

(MCT), and thioacetamide (TAA), are analyzed. They are negative in the *Salmonella* test but positive in the in vivo rat liver test [32–34].

At 24 h after a single administration of Open TG-GATES data, we analyzed five typical GTHCs (AAF, AFL, DEN, 2NF, and NNM), seven typical NGTHCs, optional choice of the 11 NGTNHCs (AA, ASP, CAF, CPA, CPP, DEX, DIA, IND, PBZ, THE, and TOL), and three intermediate GTHC/NGTHCs (MCT, MP, and TAA), with three doses, 26 chemicals, 78 data points. Appendix B.1 shows the gene expression profile (log2) of the four genes, Appendix B.2 shows the first principal component (PC1) and the second principal component (PC2) analyzed by PCA, and Fig. 2 (A) presents the result of PCA on five typical GTHCs, seven typical NGTHCs, 11 NGTNHCs, and three intermediate GTHC/NGTHCs 24 h after a single administration. The PCA results separate five typical GTHCs from seven typical NGTHCs and 11 NGTNHCs by PCA with PC1 in a two-dimensional graph, where GTHCs exhibit PC1 below -0.789 (DEN-low; PC1: -0.789 , PC2: -1.035). NGTHCs and NGTNHCs exhibit PC1 above -0.334 (IND-high; PC1: -0.334 , PC2: 1.355) (Appendix B.2). Speaking of the three intermediates GTHC/NGTHCs, TAA-high (PC1: -1.74 , PC2: -0.251)

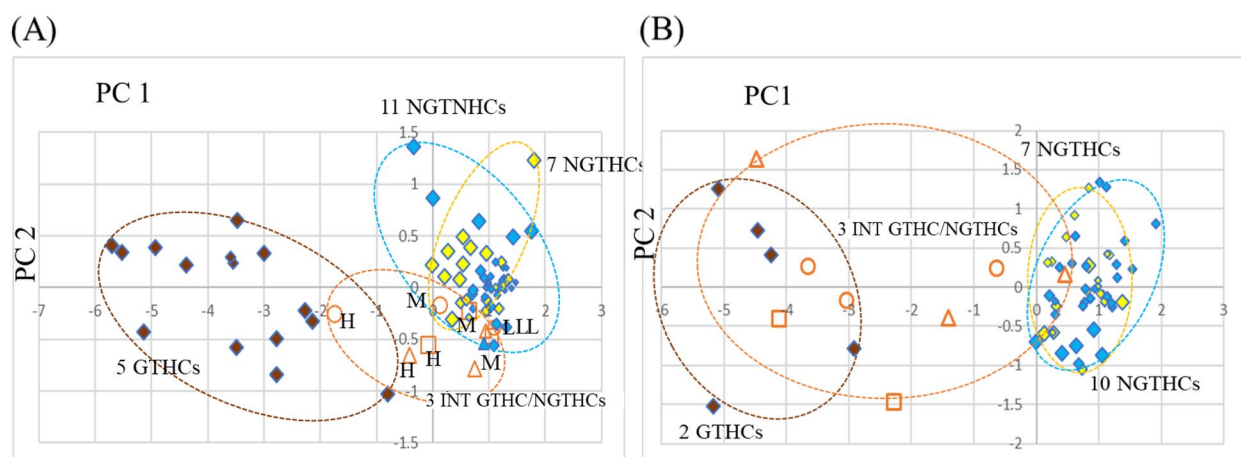


Fig. 2 **A** and **B** Analysis of three intermediate GTHC/NGTHCs in rat liver toxicogenomics public data (OPEN TG-GATES) by PCA with the four genes. Three additional intermediate GTHC/NGTHCs, methapyrilene (MP), monocrotaline (MCT), and thioacetamide (TAA), were analyzed with typical GTHCs, typical NGTHCs and NGTNHCs at 24 h after a single administration [Fig. 2(A)] and at 29 days after repeated administration [Fig. 2(B)] by PCA with four genotoxic marker genes (*Bax*, *Btg2*, *Ccng1*, and *Cdkn1*). The gene expression profile (log2) was calculated: the mean of each control group was calculated as 0 (log2), and ratio (exp/cont) (log2) was calculated (Appendix B.1 for 24 h and Appendix B.3 for 29 days). These numerical values were analyzed by PCA (Appendix B.2 for 24 h and Appendix B.4 for 29 days). At 24 h, Fig. 2(A) shows the results of 26 chemicals with three doses (low, middle and high), total of 78 points: five typical GTHCs (brown-colored, AAF, AFL, DEN, NNM, and 2NF) were discriminated from seven typical NGTHCs (yellow-colored, CLO, ETH, FEN, GEM, HEX, PHE, and WY) and 11 NGTNHCs (blue-colored, AA, ASP, CAF, CPA, CPP, DEX, DIA, IND, PBZ, THE, and TOL). GTHCs exhibit PC1 below -0.789 (DEN-low; PC1: -0.789 , PC2: -1.035). NGTHCs and NGTNHCs exhibit PC1 above -0.334 (IND-high; PC1: -0.334 , PC2: 1.355) (Appendix B.2). Speaking of the three intermediates GTHC/NGTHCs (MCT: \square , MP: Δ , and TAA: \circ), TAA-high (PC1: -1.742 , PC2: -0.251) is in the GTHCs area (below PC1: -0.789). MP-high (PC1: -0.408 , PC2: -0.659), MCT-high (PC1: -0.075 , PC2: -0.557) and MP-middle (PC1: 0.747 , PC2: -0.79) are in the intermediate area (orange dashed circle). TAA-middle (PCA: 0.135 , PC2: -0.173), TAA-low (PC1: 1.101 , PC2: -0.385), MP-low (PC1: 0.952 , PC2: -0.424), MCT-middle (PC1: 0.851 , PC2: -0.231), MCT-high (PC1: 1.043 , PC2: -0.47) and MCT-middle (PC1: 0.851 , PC2: -0.231) are in the NGTHCs and NGTNHCs area. At 29 days, Fig. 2(B) shows the results of 22 chemical, total of 63 points: two typical GTHCs (brown-colored, AAF-low, -middle, and -high; DEN-low and -high) were discriminated from seven typical NGTHCs (yellow-colored, CLO, ETH, FEN, GEM, HEX, PHE, and WY with three doses) and 10 NGTNHCs (blue-colored, AA, ASP, CAF, CPA, CPP, DIA, PBZ, THE, and TOL with three doses and IND-low and -middle). GTHCs exhibit PC1 below -2.906 (DEN-low). NGTHCs and NGTNHCs exhibit PC1 above 0.001 (CPA-high) (Appendix B.4). Three intermediate GTHC/NGTHCs, MCT-middle (PC1: -4.125 , PC2: -0.409), MP-high (PC1: -4.486 , PC2: 1.642), TAA-middle (PC1: -3.043 , PC2: -0.158), and TAA-high (PC1: -3.667 , PC2: 0.274), are in the GTHCs area. In contrast, MCT-low (PC1: -2.279 , PC2: -1.467), MP-middle (PC1: -1.403 , PC2: -0.398) and TAA-low (PC1: -0.651 , PC2: 0.248) are in the intermediate area between GTHCs and NGTHC and NGTNHCs. Only MP-low (PC1: 0.454 , PC2: 0.163) is in the NGTHCs and NGTNHCs area. Each group is enclosed with an optional dashed ellipse. Five GTHCs [AAF: 2- acetamidofluorene, AFL: aflatoxin B1, 2NF: 2-nitrofluorene, DEN: *N*-nitrosodiethylamine and NNM: *N*-nitrosomorpholine], seven NGTHCs (mostly pharmaceutical drugs) [AA: allyl alcohol, ASP: aspirin, CAF: caffeine, CPA: chlorpheniramine, CPP: chlorpropamide, DEX: dexamethasone, DIA: diazepam, IND: indomethacin, PBZ: phenylbutazone, THE: theophylline, and TOL: tolbutamide] and three intermediate GTHC/NGTHCs [MCT: monocrotaline, MP: methapyrilene, and TAA: thioacetamide]

is in the GTHCs area, and MCT-high (PC1: -0.075 , PC2: -0.557), MP-high (PC1: -0.408 , PC2: -0.659), and MP-middle (PC1: 0.747 , PC2: -0.79) are in the intermediate area. TAA-low (PC1: 1.10 , PC2: -0.385), TAA-middle (PC1: 0.135 , PC2: -0.173), MP-low (PC1: 0.952 , PC2: -0.424), MCT-middle (PC1: 0.651 , PC2: -0.231), and MCT-low (PC1: 1.043 , PC2: -0.47) are in the NGTHCs and NGTNHCs area. The results show that the three intermediate GTHC/NGTHCs (MCT, MP, and TAA) span the GTHCs, intermediate, and NGTHCs and NGTNHCs areas at 24 h after a single administration.

At 29 days, Open TG-GATES presented only two GTHCs. We analyzed two typical GTHCs (AAF with three doses and DEN-low and DEN-middle), seven

typical NGTHCs with three doses, optional 10 NGTNHCs (AA, ASP, CAF, CPA, CPP, DIA, PBZ, THE, TOL with three doses, and IND-low and IND-middle), and three intermediate GTHC/NGTHCs (MCT-low, MCT-middle, MP-low, MP-middle, MPP-high, TAA-low, TAA-middle, and TAA-high), 22 chemicals, 63 points. Appendix B.3 presents the gene expression profile (log2), Appendix B.4 presents a first principal component (PC1) and a second principal component (PC2) analyzed by PCA, and Fig. 2(B) presents the result of PCA on two typical GTHCs, seven typical NGTHCs, 10 NGTNHCs, and three intermediate GTHC/NGTHCs 29 days after repeated administration. Figure 2(B) shows the discrimination of two GTHCs from

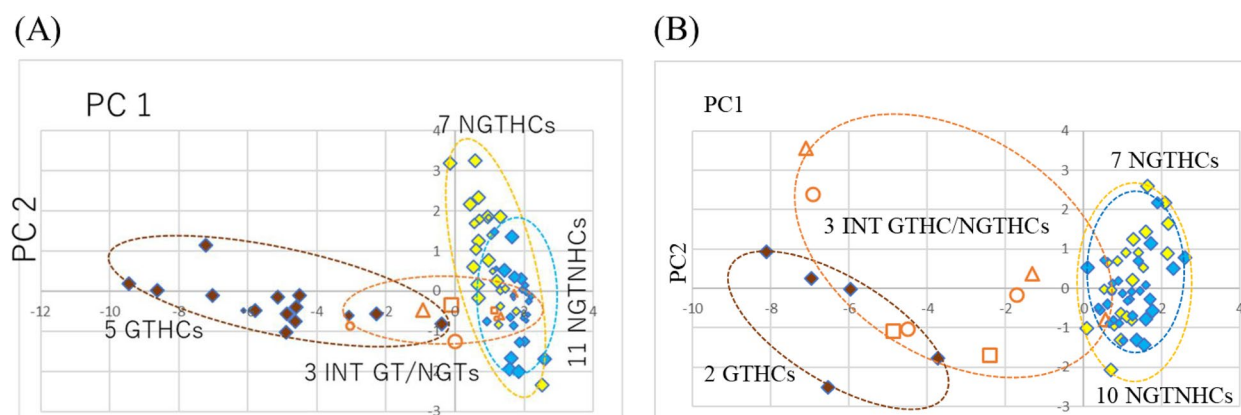


Fig. 3 **A** and **B** Comparison of three intermediate GTHC/NGTHCs analyzed by PCA with the 12 previously presented genes. We analyzed 23 chemicals plus three intermediate GTHCs/NGTHCs, a total of 26 chemicals, 78 points, analyzed by PCA with 12 previously presented genes [29] at 24 h after a single administration and a total of 22 chemicals, 63 points, at 29 days after repeated administrations. The gene expression profile (log2) was calculated: the mean of each control group was calculated as 0 (log2), and ratio (exp/cont) (log2) was calculated (Appendix C.1 for 24 h and Appendix C.3 for 29 days). These numerical values were analyzed by PCA (Appendix C.2 for 24 h and Appendix C.4 for 29 days). At 24 h after a single administration, The PCA result shows a clear separation of five typical GTHCs from seven typical NGTHCs and 11 NGTNHCs by PCA with PC1, where GTHCs exhibit PC1 below -0.381 (DEN-low, PC1: -0.381 , PC2: -0.822), and NGTHCs and NGTNHCs exhibit PC1 above -0.130 (NGTHC, FEN-high, PC1: -0.13 , PC2: 3.184) (Appendix C.2). Speaking of the three intermediate GTHC/NGTHCs (MCT: \square , MP: Δ , and TAA: \circ), MP-high (PC1: -0.941 , PC2: -0.472) and TAA-high (PC1: -3.04 , PC2: -0.869) are in the GTHCs area. MCT-high (PC1: -0.109 , PC2: -0.353) and TAA-middle (PC1: -0.024 , PC2: -1.236) are in the intermediate area. MCT-low (PC1: 1.688 , PC2: -0.149), MCT-middle (PC1: 1.149 , PC2: -0.489), MP-low (PC1: 1.702 , PC2: -0.059), MP-middle (PC1: 1.324 , PC2: -0.618), and TAA-low (PC1: 1.205 , PC2: -0.671) are in the NGTHCs and NGTNHCs area. The analysis with 12 genotoxic marker genes are similar to that with four genotoxic marker genes [Fig. 2(A)]. At 29 days after repeated administration, Fig. 3(B) demonstrates the discrimination of two GTHCs from seven NGTHCs and 10 NGTNHCs with PC1 by PCA, where GTHCs exhibit PC1 below -3.729 (DEN-low, PC1: -3.729 , PC2: -1.776), and NGTHCs and NGTNHCs exhibit PC1 above -0.084 (PHE-low, PC1: -0.084 , PC2: -1.015) (Appendix C.4). Three intermediate GTHC/NGTHCs (MCT: \square , MP: Δ , and TAA: \circ), MCT-middle (PC1: -4.873 , PC2: -1.08), MP-high (PC1: -7.093 , PC2: 3.554), TAA-middle (PC1: -4.502 , PC2: -1.021), and TAA-high (PC1: -6.931 , PC2: 2.398) are in the GTHCs area. MCT-low (PC1: -2.402 , PC2: -1.693), MP-middle (PC1: -1.322 , PC2: 0.371), and TAA-low (PC1: -1.718 , PC2: -0.155) are in the intermediate area between GTHCs and NGTHCs plus NGTNHCs. Only MP-low (PC1: 0.565 , PC2: -0.792) is in the NGTHCs and NGTNHCs area. The results of an analysis with 12 genotoxic marker genes [Fig. 3(B)] are similar to those with four genotoxic marker genes [Fig. 2(B)]

seven NGTHCs and 10 NGTNHCs with PC1 by PCA in a two-dimensional graph, where GTHCs exhibit PC1 below -2.906 (DEN-low). NGTHCs and NGTNHCs exhibit PC1 above 0.001 (CPA-high) (Appendix B.4). Speaking of the three intermediate GTHC/NGTHCs, MCT-middle (PC1: -4.125 , PC2: -0.409), MP-high (PC1: -4.486 , PC2: 1.642), TAA-middle (PC1: -3.043 , PC2: -0.158), and TAA-high (PC1: -3.667 , PC2: 0.274) are in the GTHCs area. In contrast, MCT-low (PC1: -2.279 , PC2: -1.467), MP-middle (PC1: -1.403 , PC2: -0.398) and TAA-low (PC1: -0.651 , PC2: 0.248) are in the intermediate area between GTHCs and NGTHC and NGTNHCs. Only MP-low (PC1: 0.454 , PC2: 0.163) is in the NGTHCs and NGTNHCs area. The results show that the three intermediate GTHC/NGTHCs (MCT, MP, and TAA), span the GTHCs and intermediate areas except MP-low after repeated administrations for 28 days. The 29-day results show that the three intermediate GTHC/NGTHCs are moving in the GTHCs and intermediate regions compared to the 24-h results.

Comparison of analysis of three intermediate GTHC/NGTHCs analyzed by PCA with the 12 previously presented genes

In the previous paper [29], we presented an analysis of a total of 23 chemicals, 69 data points, five typical GTHCs, seven typical NGTHCs, and 11 optional NGTNHCs at 24 h after a single administration, and a total of 19 chemicals, 55 data points, except for three intermediate GTHCs/NGTHCs. Therefore, we only present the result of 23 chemicals plus three intermediate GTHCs/NGTHCs, a total of 26 chemicals, 78 data points, analyzed with 12 previously presented genes at 24 h after a single administration, and a total of 22 chemicals, 63 data points, 29 days after repeated administrations, in the present paper.

At 24 h after a single administration, Appendix C.1 presents the gene expression profile (log2), Appendix C.2 presents the first principal component (PC1) and the second principal component (PC2) analyzed by PCA, and Fig. 3(A) presents the result of PCA of five typical GTHCs, seven typical NGTHCs, 11 NGTNHCs, and three intermediate GTHC/NGTHCs, analyzed with 12

genotoxic marker genes in a two-dimensional graph. The results show a clear separation of five typical GTHCs from seven typical NGTHCs and 11 NGTNHCs by PCA with PC1, where GTHCs exhibit PC1 below -0.381 (DEN-low, PC1: -0.381 , PC2: -0.822), and NGTHCs and NGTNHCs exhibit PC1 above -0.13 (NGTHC, FEN-high, PC1: -0.13 , PC2: 3.184) (Appendix C.2). Speaking of the three intermediate GTHC/NGTHCs, MP-high (PC1: -0.941 , PC2: -0.472) and TAA-high (PC1: -3.04 , PC2: -0.869) are in the GTHCs area. MCT-high (PC1: -0.109 , PC2: -0.353) and TAA-middle (PC1: -0.024 , PC2: -1.236) are in the intermediate area. MCT-low (PCA: 1.688 , PC2: -0.149), MCT-middle (PC1: 1.149 , PC2: -0.489), MP-low (PC1: 1.702 , PC2: -0.059), MP-middle (PC1: 1.324 , PC2: -0.618), and TAA-low (PC1: 1.205 , PC2: -0.671) are in the NGTHCs and NGTNHCs area. The analysis with 12 genotoxic marker genes are similar to that with four genotoxic marker genes [Fig. 2(A)].

At 29-day after repeated administrations in the previous paper [23], we presented an analysis of 19 chemicals, two typical GTHCs, seven typical NGTHCs, and 10 optional NGTNHCs except for three intermediate GTHC/NGTHCs. Therefore, we only present the result of 19 chemicals plus three intermediate GTHC/NGTHCs, 63 data points, analyzed with 12 previously presented genes in the present paper.

At 29 days, Appendix C.3 presents the gene expression profile (log2), and Appendix C.4 presents the first principal component (PC1) and the second principal component (PC2) analyzed by PCA. Figure 3(B) presents the results of PCA of two typical GTHCs (AAF-low, AAF-middle, AAF-high, DEN-low, and DEN-middle), seven typical NGTHCs with three doses, optional 10 NGTNHCs (AA, ASP, CAF, CPA, CPP, DIA, PBZ, THE, TOL with three doses, and IND-low and IND-middle), and three intermediate GTHC/NGTHCs (MCT-low, MCT-middle, MP-low, MP-middle, MP-high, TAA-low, TAA-middle, and TAA-high), 22 chemicals, 63 points, in a two-dimensional graph. Figure 3(B) demonstrates the discrimination of two GTHCs from seven NGTHCs and 10 NGTNHCs with PC1 by PCA, where GTHCs exhibit PC1 below -3.73 (DEN-low, PC1: -3.729 , PC2: -1.776), and NGTHCs and NGTNHCs exhibit PC1 above -0.084 (PHE-low, PC1: -0.084 , PC2: -1.015) (Appendix C.4). The distinction between GTHCs and NGTHCs by 12 genes [Fig. 3(B)] and by four genes [Fig. 2(B)] are similar. Speaking of the three intermediate GTHC/NGTHCs, MCT-middle (PC1: -4.873 , PC2: -1.08), MP-high (PC1: -7.09 , PC2: 3.55), TAA-middle (PC1: -4.50 , PC2: -1.02), and TAA-high (PC1: -6.93 , PC2: 2.40) are in the GTHCs area. MCT-low (PC1: -2.40 , PC2: -1.693), MP-middle (PC1: -1.322 , PC2: 0.371), and TAA-low (PC1: -1.72 ,

PC2: -0.155) are in the intermediate area between GTHCs and NGTHCs plus NGTNHCs. Only MP-low (PC1: 0.565 , PC2: -0.792) is in the NGTHCs and NGTNHCs area. The results of an analysis with 12 genotoxic marker genes [Fig. 3(B)] are similar to those with four genotoxic marker genes [Fig. 2(B)].

These findings strongly support our recommendation that future rat liver in vivo toxicogenomics tests should prioritize these four genotoxic marker genes, as they have proven to be highly effective in discriminating between different types of hepatocarcinogens.

Discussion

At the beginning of the twenty-first century, we expected that toxicogenomics approaches would clarify toxic modes of action of chemical compounds, including carcinogens, in a biological system. Although hundreds of studies were published [33, 79–81], practical in vivo short-term screening test methods for carcinogens using toxicogenomics (gene expression profiles) have yet to be established. About 90% of human carcinogens are genotoxic carcinogens [82]; genotoxic carcinogens are still a significant threat to human health. We have been working to develop an in vivo short-term genotoxic carcinogen screening method using gene expression profiles based on the toxic modes of action of chemical compounds and using data analysis by PCA. It would be desirable to distinguish genotoxic carcinogens from non-genotoxic carcinogens, and non-genotoxic noncarcinogens by gene expression profiles and PCA. For this purpose, it is necessary to identify marker genes that distinguish between genotoxic carcinogens, non-genotoxic carcinogens, and non-genotoxic noncarcinogens.

We used PCA for our statistical analysis. PCA is an unsupervised learning algorithm. PCA is not widely used in toxicogenomics but is commonly used in other biological fields, such as cancer analysis [83, 84]. Ringnér wrote that “PCA is often incorporated into genome-wide expression studies.” He explained that “samples can then be plotted, making it possible to visually assess similarities and differences between samples and determine whether samples can be grouped” [85]. PCA is advantageous because the results are clearly understood visually in a two-dimensional or three-dimensional figure with numerical values without bioinformatics knowledge. Previously, we successfully applied PCA to human lung cancer cell lines to discriminate four histopathological subtypes (adenocarcinoma, squamous cell carcinoma, large-cell carcinoma, and small-cell carcinoma) [86, 87]. We also successfully applied PCA to toxicogenomics to discriminate GTHCs from NGTHCs in mice [14], rat models [15, 30–32], and Open TG-GATEs [29].

The present gene data set (Appendix A.1, A.2, A.3, A.4; B.1, B.2, B.3, and B.4) may be functional in predicting the genotoxicity of hepato-carcinogenicity of new chemicals; we have added our data to for this purpose. Users can add their data [gene expression profile (log2)] in addition to one of our existing data (Appendix A.1, A.3, B.1, and B.3) one by one or in small numbers and conduct PCA (<https://www.r-project.org/>). For example, comparing Appendix A.2 and Appendix B.2, even with the addition of the 9 data points, for MCT24hL, MCT24hM, MCT24hH, MP24hL, MP24hM, MP24hH and TAA24hL, TAA24hM, TAA24hH for intermediate GTHC/NGTHCs, the PC1 for AAF24hL in GTHCs does not change significantly from −3.283 (Appendix A.2) to −3.462 (Appendix B.2).

Users can also attempt to calculate PC1 for their test compound using the following formula. The first principal component (Y1) is given by the linear combination of the variable X1, X2, ..., Xp.

$Y1 = a1X1 + a2X2 + \dots + a1pXp$ where $a1p$ is the eigenvector, which can be calculated with the PCA program in R, and Xp is the canonicalized logarithmic (log2)-transformed gene ratio (exp/cont), $[(x-\mu)/\sigma]$.

x is the logarithmic log2 of exp/cont, μ is the mean and σ is the standard deviation (<https://strata.uga.edu/software/pdf/pcaTutorial.pdf>).

When we calculate all data of 24 h and 29 days by R, $a11$, ..., $a1p$ of PC1 are $a(Bax)$: −0.501, $a(Btg2)$: −0.500, $a(Ccng1)$: −0.516, $a(Cdkn1a)$: −0.483. Users can calculate their PC1 (Y1) by introducing their xp into the following equation:

$$Y1 = (-0.501) \times [(xBax-0.320)/0.0790] + (-0.500) \times [(xBtg2-0.262)/0.946] + (-0.516) \times [(xCcng1-0.562)/1.1166] + (-0.483) \times [(xCdkn1a-0.402)/1.352]$$

In this equation, users can judge a possibility of GTHC by the cut-off value of −0.507 (median value between the maximum value of GTHC, DEN24hL, −1.168 and the minimum value of NGTHC and NGTNHC, IND24hH, −0.778) obtained in our experiment set, assuming that the difference of μ and σ in $[(x-\mu)/\sigma]$ will be small when total number of samples is greater. We also present the table to calculate the more accurate PC1. Users can add their data (Gene expression profiles (exp/cont) to excel data, then they will get their PC1 value together with a re-calculated cut-off value for GTHC, by integrating new data.

Similarly, the PC2 (Y2) value can be calculated by the following equation:

Recently, we published a paper to evaluate rat GTHC and NGTHC via selected gene expression patterns in the liver, as determined by NGS-targeted mRNA sequencing (RNA-Seq) and PCA [30]. We analyzed two typical GTHC (DEN and 3,3'-dimethylbenzidine·2HCl), a typical NGTHC [di-(2-ethylhexyl)phthalate], and 1,4-dioxan (DO), which has long been unclear whether it is a GTHC. The results suggested that PCA discriminated between two GTHCs and NGTHC and that DO resulted in an intermediate gene expression profile different from typical GTHC and NGTHC. In the "Discussion" of the paper, we showed that existing data from TG-GATES helped evaluate new RNA-Seq data by PCA.

We also performed FFPE RNA-Seq to compare a typical GTHC, 2-acetylaminofluorene (AAF), to genotoxicity equivocal p-cresidine (CRE). CRE is used as a synthetic chemical intermediate, and this compound is classified as an IARC 2B carcinogen and is mutagenic in the *Salmonella* test, which is non-genotoxic to rat livers as assessed by single-strand DNA damage analysis. PCA resulted in CRE as an NGTHC in our experiment. Our results suggest that FFPE RNA-Seq and PCA are useful for evaluating rat GTHCs and NGTHCs [31].

As described previously [29], in connection with restrictions on animal use, "OECD Guidelines for the Testing of Chemicals, [Repeated Dose 28-Day Oral Toxicity Study on Rodents (OECD TG, 2008, 407)] [Test No. 407: Repeated Dose 28-Day Oral Toxicity Study in Rodents | READ online (oecd-ilibrary.org)] is still valid for testing chemical toxicity. This assay determines the

general toxicity of chemicals in rodents after 28 days of oral dosing (e.g., effects on the liver, kidney, heart, and lungs). Despite restrictions on animal testing, this test will continue to be applied. We can use the animal organs from the test collaboratively and the samples, reducing the number of experimental animals used. Using FFPE samples is also available for RNA-Seq and spatial transcriptomic [88] and helps to reduce the number of experimental animals used.

Recently, Gi et al. reported ten genotoxic marker genes (*Aen*, *Cdln1a*, *Phlda3*, *Nudt5*, *Mybl1*, *Glx3*, *Atp6v1f*, *Mok*, *Cyria*, *Sugct*) from the Open TG-GATES (with five genotoxic hepatocarcinogens) [89]. *Aen*, *Cdln1a*, and *Phlda3* are common with our previous 12 genotoxic

$$Y2 = (0.516 \times [(xBax-0.320)/0.790] + 0.202 \times [(xBtg2-0.262)/0.946] + 0.079 \times [(xCcng1-0.562)/1.166] + (-0.828) \times [(xCdkn1a-0.402)/1.352]$$

marker genes [14, 29]. We first selected the 12 marker genes with eight different mouse hepatocarcinogens and then evaluated them with rat Open TG-GSTEs. Since the hepatocarcinogens that initially chose the marker genes are different, different marker genes were likely chosen.

Next-generation risk assessment of chemical substances is expected to utilize mechanistic information without animal testing. In this regard, toxicogenomics has proven to be a valuable tool for elucidating the mechanisms underlying the adverse effects of toxic substances: 3D liver microtissue, primary human hepatocytes (PHH) [90], human liver tissue: cancer-derived cell lines (HepaRG) [91], and others are under investigation. The current in vivo short-term test will be instrumental in selecting the genes to be analyzed.

Toxicogenomics technology has progressed from DNA microarray through qPCR to RNA-Seq. DNA microarray and qPCR compare relative fluorescence intensities, while RNA-Seq compares digital nucleotide numbers, which is more reliable. The following emerging technology is the spatial transcriptome (spatial biology) [92]. The technique can count the number of mRNA expressions for each cell in a pathological section. Incorporating this method would provide reliable data because this method can count the number of mRNA expressions for each cell type in the organ. Whether for RNA-Seq or spatial biology, the small number of target genes is an advantage. The four genes in this paper will be helpful for future research. Although the current research on these genes focus only in liver, it is worth extending their application to other organs because the genotoxic mechanism involving them can be common in all organs.

Conclusions

The present results unequivocally demonstrate the performance of four genotoxic marker genes (*Bax*, *Btg2*, *Ccng1*, and *Cdkn1a*) and PCA in discriminating GTHCs from NGTHCs and NGTNHCs in Open TG-GATEs. These findings strongly support our recommendation that future rat liver in vivo toxicogenomics tests prioritize these four genotoxic marker genes, as they have proven to be highly effective in discriminating between different types of hepatocarcinogens.

Abbreviations

Chemicals

AA	Allyl alcohol
AAF	2-Acetylaminofluorene
AFB	Aflatoxin B1
α-NFL	Alpha-naphthoflavone
ASP	Aspirin
BAL	Balasalazide
BIS	Bisphenol A

CAF	Caffeine
CHL	Chloroform
CLO	Clofibrate
COR	Cortisone
CPA	Chlorpheniramine
CPP	Chlorpropamide
COR	Cortisone
CRE	P-Cresidine
DEHP	Di(2-ethylhexyl)phthalate
DEN	Diethylnitrosamine
DEX	Dexamethasone
DIA	Diazepam
DMF	N,N-Dimethylformamide
DMN	Dimethylnitrosamine
ETH	Ethanol
FEN	Fenofibrate
FIN	Finasteride
GEN	Gemfibrozil
IFA	Ifosfamide
MCT	Monocrotaline
MP	Methapyrilene
2NF	2-Nitrofluorene
NNM	N-Nitrosomorpholine
PHE	Phenobarbital
PBZ	Phenylbutazone
RAL	Raloxifene
TAA	Thioacetamide
THE	Theophylline
TOL	Tolbutamide

Genes

<i>Aen</i>	Apoptosis enhancing nuclease
<i>Atp6vaf</i>	ATPase H ⁺ transporting V1 subunit F
<i>Bax</i>	BCL2 associated X, apoptosis regulator
<i>Btg2</i>	BTG anti-proliferation factor 2
<i>Ccnf</i>	Cyclin F
<i>Ccng1</i>	Cyclin G1
<i>Cgrrf1</i>	Cell growth regulator with ring finger domain 1
<i>Cdkn1a</i>	Cyclin-dependent kinase inhibitor 1A
<i>Cyria</i>	CYFIP related Rac1 interactor A
<i>Gdf15</i>	Growth differentiation factor 15
<i>Glrx3</i>	Glutaredoxin 3
<i>Lrp1</i>	LDL receptor related protein 1
<i>Mdmb1</i>	Methyl-CpG binding domain protein 1
<i>Mgmt</i>	O-6-methylguanine-DNA methyltransferase
<i>Mok</i>	MOK protein kinase
<i>Mybl1</i>	MYB proto-oncogene like 1
<i>Nudt5</i>	Nudix Hydrolase 5
<i>Phlda3</i>	Pleckstrin homology-like domain, family A, member 3
<i>Plk2</i>	Polo-like kinase 2
<i>Sugt</i>	Succinyl CoA:glutarate-CoA transferase
<i>Tmem47</i>	Transmembrane protein 47
<i>Tubb4b</i>	Tubulin, beta 4B class IVb

Words

EPA	The U.S. Environmental Protection Agency
GTHC	Genotoxic hepatocarcinogen
ICEM	International Conference on Environmental Mutagen
JEMS/MMS	Japanese Environmental Mutagen and Genome Society/Mammalian Mutagenicity Study Group
NGTHC	Non-genotoxic hepatocarcinogen
NGTNHC	Non-genotoxic non-hepatocarcinogen
OECD	The Organisation for Economic Co-operation and Development
PCA	Principal component analysis
PC1	1st principal component
PC2	2Nd principal component
qPCR	Quantitative real-time PCR
RNA-Seq	RNA-sequencing

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s41021-024-00322-8>.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Authors' contributions

CF and TS designed and critically discussed the manuscript. CF and TS read and approved the final manuscript.

Data availability

Not applicable.

Declarations

Ethics approval and consent to participate

For studies involving animals, ethical approvals were obtained from the institutions where the original studies were conducted.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Division of Molecular Target and Gene Therapy Products, National Institute of Health Sciences, 3-25-26 Tonomachi, Kawasaki-Ku, Kawasaki, Kanagawa 210-9501, Japan. ²School of Science and Engineering, Aoyama Gakuin University, Sagami-hara, Sagami-hara, Kanagawa 252-5258, Japan. ³Division of Genome Safety Science, National Institute of Health Sciences, 3-25-26, Tonomachi, Kawasaki-Ku 210-9501, Japan.

Received: 11 September 2024 Accepted: 3 December 2024

Published online: 19 December 2024

References

1. Cancer-Causing Substances in the Environment (<https://www.cancer.gov/about-cancer/causes-prevention/risk/substances>).
2. Hayes AW, Dayan AD, Hall WC, Kodell RL, Williams GM, Waddell WD, Slesinski RS, Kruger CL. A review of mammalian carcinogenicity study design and potential effects of alternate test procedures on the safety evaluation of food ingredients. *Regul Toxicol Pharmacol*. 2011;60(1 Suppl):S1–34.
3. Jacobs A, Jacobson-Kram D. Human carcinogenic risk evaluation, Part III: Assessing cancer hazard and risk in human drug development. *Toxicol Sci*. 2004;81:260–2.
4. Heusinkveld H, Braakhuis KZ, Gommans R, Botham P, Corvaro M, van der Laan JW, Lewis D, Madia F, Manou I, Schorsch F, Wolterink G, Woutersen R, Corvi R, Mehta J, Luijten M. Towards a mechanism-based approach for the prediction of nongenotoxic carcinogenic potential of agrochemicals. *Crit Rev Toxicol*. 2020;50:725–39.
5. Li T, Tong W, Roberts R, Liu Z, Thakkar S. DeepCarc: Deep learning-powered carcinogenicity prediction using model-level representation. *Front Artif Intell*. 2021;4: 757780.
6. Corvi R, Madia F, Guyton KZ, Kasper P, Rudel R, Colacci A, Kleinjans J, Jennings P. Moving forward in carcinogenicity assessment: Report of an EURL ECVAM/ESTIV workshop. *Toxicol In Vitro*. 2017;45:278–86.
7. Jacobs MN, Colacci A, Corvi R, Vaccari M, Aguila MC, Corvaro M, Delrue N, Desaulniers D, Ertych N, Jacobs A, Luijten M, Madia F, Nishikawa A, Ogawa K, Ohmori K, Paparella M, Sharma AK, Vasseur P. Chemical carcinogen safety testing: OECD expert group international consensus on the development of an integrated approach for the testing and assessment of chemical non-genotoxic carcinogens. *Arch Toxicol*. 2020;94:2899–923.
8. Totsuka Y, Watanabe M, Lin Y. New horizons of DNA adductome for exploring environmental causes of cancer. *Cancer Sci*. 2021;112:7–15.
9. LaFave MC, Sekelsky J. Mitotic Recombination: Why? When? How? Where? *Plos Genet*. 2009;5: e1000411.
10. Bevan RJ, Harrison PTC. Threshold and non-threshold chemical carcinogens: A survey of the present regulatory landscape. *Regul Toxicol Pharmacol*. 2017;88:291–302.
11. Zhou T, Chou J, Watkins PB, Kaufmann WK. Toxicogenomics: transcription profiling for toxicology assessment. *EXS*. 2009;99:325–66.
12. Watanabe T, Tobe K, Nakachi Y, Kondoh Y, Nakajima M, Hamada S, Namiki C, Suzuki S, Maeda S, Tadakuma A, Sakurai M, Arai A, Hyogo A, Hoshino M, Tashiro T, Ito H, Inazumi H, Sakaki Y, Tashiro H, Furihata C. Differential gene expression induced by two genotoxic N-nitroso carcinogens, phenobarbital and ethanol in mouse liver examined with oligonucleotide microarray and quantitative real-time PCR. *Genes Environ*. 2007;29:115–27.
13. Watanabe T, Tanaka G, Hamada S, Namiki C, Suzuki T, Nakajima M, Furihata C. Dose-dependent alterations in gene expression in mouse liver induced by diethylnitrosamine and ethylnitrosourea and determined by quantitative real time PCR. *Mutat Res*. 2009;673:9–20.
14. Watanabe T, Suzuki T, Natsume M, Nakajima M, Narumi K, Hamada S, Sakuma T, Koeda A, Oshida K, Miyamoto Y, Maeda A, Hirayama M, Sanada H, Honda H, Ohya W, Okada E, Fujiishi Y, Souto S, Tadakuma A, Ishikawa Y, Kido M, Minamiguchi R, Hanahara I, Furihata C. Discrimination of genotoxic and non-genotoxic hepatocarcinogens by statistical analysis based on gene expression profiling in the mouse liver as determined by quantitative real-time PCR. *Mutat Res*. 2012;747:164–75.
15. Suenaga K, Takasawa H, Watanabe T, Wako Y, T. Suzuki T, Hamada S, Furihata C. Differential gene expression profiling between genotoxic and non-genotoxic hepatocarcinogens in young rat liver determined by quantitative real-time PCR and principal component analysis. *Mutat Res*. 2013;751:73–83.
16. Sakurai M, Watanabe T, Suzuki T, Furihata C. Time-course comparison of gene expression profiles induced by the genotoxic hepatocarcinogen, chrysene, in the mouse liver. *Genes Environ*. 2014;36:54–64.
17. Furihata C, Watanabe T, Suzuki T, Hamada S, Nakajima M. Collaborative studies in toxicogenomics in rodent liver in JEMS-MMS; a useful application of principal component analysis on toxicogenomics. *Genes Environ*. 2016;38:15.
18. NCBI.NLM.NIH, <https://pubchem.ncbi.nlm.nih.gov/compound/2-Acetylaminofluorene>
19. NCBI.NLM.NIH, https://pubchem.ncbi.nlm.nih.gov/compound/2_4-Diaminotoluene.
20. NTP.NIEHS.NIH., https://ntp.niehs.nih.gov/ntp/roc/content/profiles/dimet_hylaminoozobenzene.pdf.
21. Hirao K, Shinohara Y, Tsuda H, Fukushima S, Takahashi M. Carcinogenic activity of quinoline on rat liver. *Cancer Res*. 1976;36:329.
22. NTP.NIEHS.NIH., <https://ntp.niehs.nih.gov/sites/default/files/ntp/roc/content/profiles/nitrosamines.pdf>
23. Hecht SS. Biochemistry, biology, and carcinogenicity of tobacco-specific N-nitrosamines. *Chem Res Toxicol*. 1998;11:559–603.
24. Hecht SS, Lijinsky W, Kovatch RM, Chung FL, Saavedra JE. Comparative tumorigenicity of N-nitroso-2-hydroxymorpholine, N-nitrosodiethanolamine and N-nitrosomorpholine in A/J mice and F344 rats. *Carcinogenesis*. 1989;10:1475–7.
25. NTP.NIEHS.NIH., <https://ntp.niehs.nih.gov/ntp/roc/content/profiles/urethane.pdf>.
26. D'Angiolella V, Donato V, Forrester FM, Jeong Y-T, Pellacani C, Kudo Y, Saraf A, Florens L, Washburn MP, Pagano M. Cyclin F-mediated degradation of ribonucleotide reductase M2 controls genome integrity and DNA repair. *Cell*. 2012;149:1023–34.
27. Van Gool B, Dedieu S, Emonard H, Roebroek AJ. The matricellular receptor LRP1 forms an interface for signaling and endocytosis in modulation of the extra cellular tumor environment. *Front Pharmacol*. 2015;6:271.
28. Sobierajska K, Ciszewski WM, Wawro ME, Wieczorek-Szukala K, Boncela J, Papiewska-Pajak I, Niewiarowska J, Kowalska MA. TUBB4B downregulation is critical for increasing migration of metastatic colon cancer cells. *Cells*. 2019;8:810.
29. Furihata C, Suzuki T. Evaluation of 12 mouse marker genes in rat toxicogenomics public data, Open TG-GATES: Discrimination of genotoxic from non-genotoxic hepatocarcinogens. *Mutat Res Genet Toxicol Environ Mutagen*. 2019;838:9–15.

30. Furihata C, Toyoda T, Ogawa K, Suzuki T. Using RNA-Seq with 11 marker genes to evaluate 1,4-dioxane compared with typical genotoxic and non-genotoxic rat hepatocarcinogens. *Mutat Res Genet Toxicol Environ Mutagen*. 2018;834:51–5.
31. Furihata C, You X, Toyoda T, Ogawa K, Suzuki T. Using FFPE RNA-Seq with 12 marker genes to evaluate genotoxic and non-genotoxic rat hepatocarcinogens. *Genes Environ*. 2020;42:15.
32. Furihata C, Suzuki T. Short-term in vivo testing to discriminate genotoxic carcinogens from non-genotoxic carcinogens and non-carcinogens using next-generation RNA sequencing, DNA microarray, and qPCR. *Genes Environ*. 2023;45:7.
33. Igarashi Y, Nakatsu N, Yamashita T, Ono A, Ohno Y, Urushidani T, Yamada H. Open TG-GATES: a large-scale toxicogenomics database. *Nucleic Acids Res*. 2015;43:D921–7 Database issue.
34. Rooney J, Hill T 3rd, Qin C, Sistare FD, Corton JC. Adverse outcome pathway-driven identification of rat liver tumorigens in short-term assays. *Toxicol Appl Pharmacol*. 2018;356:99–113.28.
35. Agents Classified by the IARC Monographs, Volumes 1–123. <https://monographs.iarc.who.int/wp-content/uploads/2018/09/ClassificationsAlphaOrder.pdf>
36. The carcinogenic potency database <https://files.toxplanet.com/cpdb/pdfs/ChemicalTable.pdf>
37. Mirsalis JC, Tyson CK, Butterworth BE. Detection of genotoxic carcinogens in the in vivo-in vitro hepatocyte DNA repair assay. *Environ Mutagen*. 1982;4:553–62.
38. Butler WH, Greenblatt M, Lijinsky W. Carcinogenesis in Rats by Aflatoxins B₁, G₁, and B₂. *Cancer Res*. 1969;29:2206–11.
39. Beijer B, Möller L. Correlation between induction of unscheduled DNA synthesis in the liver and excretion of mutagenic metabolites in the urine of rats exposed to the carcinogenic air pollutant 2-nitrofluorene. *Carcinogenesis*. 1988;9:1465–70.
40. Brambilla G, Martelli A, Pino A, Robbiano L. Sequential analysis of DNA damage and repair during the development of carcinogen-induced rat liver hyperplastic lesions. *Cancer Res*. 1986;46:3476–81.
41. Cortinovis C, Klimek F, Nogueira E. Rat hepatocarcinogenesis induced by N-nitrosodiethylamine and N-nitrosomorpholine continuously administered at low doses. From basophilic areas of hepatocytes to hepatocellular tumors. *Am J Pathol*. 1991;139:1157–71.
42. Korr H, Botzem B, Schmitz C, Enzmann H. N-Nitrosomorpholine induced alterations of unscheduled DNA synthesis, mitochondrial DNA synthesis and cell proliferation in different cell types of liver, kidney, and urogenital organs in the rat. *Chem Biol Interact*. 2001;134:217–33.
43. Gezer E, Özer C, Şimşek T, Yaprak Bayrak B, Turan G, Çetinarslan B, Selek A, Cantürk Z, Sözen M, Köksalan D. N-Nitrosomorpholine-induced oncogenic transformation in rat endocrine organs. *Eur J Med Res*. 2024;29:64.
44. Takayanagi T, Takashima R, Wako Y, Kawasaki K, Tanaka Y, Hori H, Fujii W. Repeated dose liver micronucleus assay using clofibrate in young adult rats. *Mutat Res Genet Toxicol Environ Mutagen*. 2015;780–781:117–22.
45. Ji Z, Settivar RS, LeBaron MJ. Pilot studies evaluating the nongenotoxic rodent carcinogens phenobarbital and clofibrate in the rat Pig-a assay. *Environ Mol Mutagen*. 2019;60:42–6.
46. Reddy JK, Qureshi SA. Tumorigenicity of the hypolipidaemic peroxisome proliferator ethyl-alpha-p-chlorophenoxyisobutyrate (clofibrate) in rats. *Br J Cancer*. 1979;40:476–82.
47. Clofibrate. IARC monographs on the identification of carcinogenic hazards to humans. 1996;66:391–426.
48. Navasumrit P, Ward TH, Dodd NJ, O'Connor PJ. Ethanol-induced free radicals and hepatic DNA strand breaks are prevented in vivo by antioxidants: effects of acute and chronic ethanol exposure. *Carcinogenesis*. 2000;21:93–9.
49. Nakagawa Y, Toyozumi T, Sui H, Ohta R, Kumagai F, Usumi K, Saito Y, Yamakage K. In vivo comet assay of acrylonitrile, 9-aminoacridine hydrochloride monohydrate and ethanol in rats. *Mutat Res Genet Toxicol Environ Mutagen*. 2015;786–788:104–13.
50. Tawfeeq MM, Suzuki T, Shimamoto K, Hayashi H, Shibutani M, Mitsumori K. Evaluation of in vivo genotoxic potential of fenofibrate in rats subjected to two-week repeated oral administration. *Arch Toxicol*. 2011;85:1003–11.
51. NCBI.NLM.NIH, <https://pubchem.ncbi.nlm.nih.gov/compound/Fenofibrate>
52. Fitzgerald JE, Sanyer JL, Schardein JL, Lake RS, McGuire EJ, de la Iglesia FA. Carcinogen bioassay and mutagenicity studies with the hypolipidemic agent gemfibrozil. *J Natl Cancer Inst*. 1981;67:1105–16.
53. Gemfibrozil. IARC monographs on the identification of carcinogenic hazards to humans. 1996;66:427–44.
54. Hexachlorobenzene. IARC monographs on the identification of carcinogenic hazards to humans. 1999;(Supl 7):219.
55. Phenobarbital. IARC monographs on the identification of carcinogenic hazards to humans. 2001;13:157–182.
56. Rothfuss A, O'Donovan M, De Boeck M, Brault D, Czich A, Custer L, Hamada S, Plappert-Helbig U, Hayashi M, Howe J, Kraynak AR, van der Leede BJ, Nakajima M, Priestley C, Thybaud V, Saigo K, Sawant S, Shi J, Storer R, Struwe M, Vock E, Galloway S. Collaborative study on fifteen compounds in the rat-liver Comet assay integrated into 2- and 4-week repeat-dose studies. *Mutat Res*. 2010;702:40–69.
57. Biegel LB, Hurtt ME, Frame SR, O'Connor JC, Cook JC. Mechanisms of extrahepatic tumor induction by peroxisome proliferators in male CD rats. *Toxicol Sci*. 2001;60:44–55.
58. Cattley RC1, Smith-Oliver T, Butterworth BE, Popp JA. Failure of the peroxisome proliferator WY-14,643 to induce unscheduled DNA synthesis in rat hepatocytes following in vivo treatment. *Carcinogenesis*. 1988;9:1179–83.
59. NCBI.NLM.NIH. Allyl alcohol. <https://pubchem.ncbi.nlm.nih.gov/compound/Allyl-alcohol#section=Toxicity-Summary>.
60. NCBI.NLM.NIH. Aspirin. <https://pubchem.ncbi.nlm.nih.gov/compound/Aspirin>.
61. Coffee. IARC monographs on the identification of carcinogenic hazards to humans. 1991;51:42–206.
62. NCBI.NLM.NIH. Chlorpheniramine. <https://pubchem.ncbi.nlm.nih.gov/compound/Chlorpheniramine>.
63. NCBI.NLM.NIH. Chlorpropamide. <https://pubchem.ncbi.nlm.nih.gov/compound/Chlorpropamide>.
64. NCBI.NLM.NIH. Dexamethasone. <https://pubchem.ncbi.nlm.nih.gov/compound/Dexamethasone>.
65. Diazepam. IARC Monographs on the Identification of Carcinogenic Hazards to humans. 1996;66:37–96.
66. NCBI.NLM.NIH. Indomethacin. <https://pubchem.ncbi.nlm.nih.gov/compound/Indomethacin>.
67. Phenylbutazone. IARC monographs on the identification of carcinogenic hazards to humans. 1987;13:183–200.
68. Theophylline. IARC monographs on the identification of carcinogenic hazards to humans. 1991;51:207–271.
69. TP. NTP toxicology and carcinogenesis studies of theophylline (CAS No. 58-55-9) in F344/N rats and B6C3F1 mice (feed and gavage studies). *Natl Toxicol Program Tech Rep Ser* 1998;473:1–326.
70. NCBI.NLM.NIH. Tolbutamide. <https://ntp.niehs.nih.gov/publications/reports/tr/000s/tr031>.
71. Inoue K, Ochi A, Koda A, Wako Y, Kawasaki K, Doi T. The 14-day repeated dose liver micronucleus test with methapyrilene hydrochloride using young adult rats. *Mutat Res Genet Toxicol Environ Mutagen*. 2015;780–781:123–7.
72. NTP. NTP Hepatotoxicity Studies of the Liver Carcinogen Methapyrilene Hydrochloride (CAS No. 135-23-9) Administered in Feed to Male F344/N Rats. *Toxic Rep Ser*. 2000;46:1–C7.
73. Takashima R, Takasawa H, Wako Y, Yasunaga K, Hattori A, Kawabata M, Nakadate K, Nakagawa M, Hamada S. Micronucleus induction in rat liver and bone marrow by acute vs. repeat doses of the genotoxic hepatocarcinogen monocrotaline. *Mutat Res Genet Toxicol Environ Mutagen*. 2015;780–781:64–70.
74. Petry TW, Sipes IG. Modulation of monocrotaline-induced hepatic genotoxicity in rats. *Carcinogenesis*. 1987;8:415–9.
75. Priestley CC, Regan S, Kevin Park B, Williams DP. The genotoxic potential of methapyrilene using the alkaline Comet assay in vitro and in vivo. *Toxicology*. 2011;290:249–57.
76. Toyoda T, Sone M, Matsushita K, Akane H, Akagi J, Morikawa T Y, Young-Man Cho Y-M, Ogawa K. Early detection of hepatocarcinogens in rats by immunohistoc. *J Toxicol Sci*. 2023;48:323–32.
77. Sui H, Matsumoto H, Wako Y, Kawasaki K. Evaluation of in vivo genotoxicity by thioacetamide in a 28day repeated dose liver micronucleus assay using male young adult rats. *Mutat Res Genet Toxicol Environ Mutagen*. 2015;780–781:81–4.

78. Möller L, Rafter J, Gustafsson JA. Metabolism of the carcinogenic air pollutant 2-nitrofluorene in the rat. *Carcinogenesis*. 1987;8:637–45.
79. North M, Vulpe CD. Functional toxicogenomics: mechanism-centered toxicology. *Int J Mol Sci*. 2010;11:4796–813.
80. Liu Z, Huang R, Roberts R, Tong W. Toxicogenomics: A 2020 vision. *Trends Pharmacol Sci*. 2019;40:92–103.
81. Pandiri AR, Auerbach SS, Stevens JL, Blomme EAG. Toxicogenomics approaches to address toxicity and carcinogenicity in the liver. *Toxicol Pathol*. 2023;51:470–81.
82. Waters MD, Jackson M, Lea I. Characterizing and predicting carcinogenicity and mode of action using conventional and toxicogenomics methods. *Mutat Res*. 2010;705:184–200.
83. Ellsworth SG, Rabatic BM, Chen J, Zhao J, Campbell J, Wang W, Pi W, Stanton P, Matuszak M, Jolly S, Miller A, Kong FM. Principal component analysis identifies patterns of cytokine expression in non-small cell lung cancer patients undergoing definitive radiation therapy. *PLoS ONE*. 2017;12: e0183239.
84. Ram S, Tang W, Bell AJ, Pal R, Spencer C, Buschhaus A, Hatt CR, diMagliano MP, Rehmentulla A, Rodríguez JJ, Galban S, Galban CJ. Lung cancer lesion detection in histopathology images using graph-based sparse PCA network. *Neoplasia*. 2023;42: 100911.
85. Ringnér M. What is principal component analysis? *Nat Biotechnol*. 2008;26:303–4.
86. Watanabe T, Miura T, Degawa Y, Fujita Y, Inoue M, Kawaguchi M, Furihata C. Comparison of lung cancer cell lines representing four histopathological subtypes with gene expression profiling using quantitative real-time PCR. *Cancer Cell Int*. 2010;10:2.
87. Oshima Y, Shinzawa H, Takenaka T, Furihata C, Sato H. Discrimination analysis of human lung cancer cells associated with histological type and malignancy using Raman spectroscopy. *J Biomed Opt*. 2010;15: 017009.
88. Villacampa EG, Larsson L, Mirzazadeh R, Kvstad L, Andersson A, Mollbrink A, Kokaraki G, Monteil V, Schultz N, Appelberg KS, Montserrat N, Zhang H, Penninger JM, Miesbach W, Mirazimi A, Carlson J, Lundberg J. Genome-wide spatial expression profiling in formalin-fixed tissues. *Cell Genom*. 2021;1: 100065.
89. Gi M, Suzuki S, Kanki M, Yokohira M, Tsukamoto T, Fujioka M, Vachiraarunwong A, Qiu G, Wanibuchi H. A novel support vector machine-based 1-day, single-dose prediction model of genotoxic hepatocarcinogenicity in rats. *Arch Toxicol*. 2024;98:2711–30.
90. Gupta R, Schrooders Y, Hauser D, van Herwijnen M, Albrecht W, Braak BT, Brecklinghaus T, Castell JV, Elenschneider L, Escher S, Guye P, Hengstler JG, Ghallab A, Hansen T, Leist M, MacLennan R, Moritz W, Tolosa L, Tricot T, Verfaillie C, Walker P, van de Water B, Kleinjans J, Caiment F. Comparing in vitro human liver models to in vivo human liver using RNA-Seq. *Arch Toxicol*. 2021;95:573–89.
91. Kunnen SJ, Arnesdotter E, Willenbockel CT, Vinken M, van de Water B. Qualitative and quantitative concentration-response modelling of gene co-expression networks to unlock hepatotoxic mechanisms for next generation chemical safety assessment. *Altex*. 2024;41:213–32.
92. Tian L, Chen F, Macosko EZ. The expanding vistas of spatial transcriptomics. *Nat Biotechnol*. 2023;41:773–82.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.