SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
MICROARRAYS
CANCER GENOMICS
CANCER MODELS
CANCER SCREENING

Received
29 May 2013

Accepted
6 September 2013

Published
3 October 2013

Correspondence and requests for materials should be addressed to S.W.K. (swkwon@snu.ac.kr)

Distinguishing between genotoxic and non-genotoxic hepatocarcinogens by gene expression profiling and bioinformatic pathway analysis

Seul Ji Lee[1], Young Na Yum[2], Sang Cheol Kim[3], Yuneung Kim[4], Johan Lim[4], Won Jun Lee[1], Kyung Hye Koo[2], Joo Hwan Kim[2], Jee Eun Kim[2], Woo Sun Lee[2], Soojung Sohn[2], Sue Nie Park[2], Jeong Hill Park[1], Jeongmi Lee[5] & Sung Won Kwon[1]

[1]College of Pharmacy and Research Institute of Pharmaceutical Sciences, Seoul National University, Seoul, 151–742, Korea, [2]Toxicological Evaluation and Research Department, National Institute of Food and Drug Safety Evaluation, Korea Food and Drug Administration, Osong, 363–700, Korea, [3]Korean Bioinformation Center (KOBIC), KRIBB, Daejeon, 305–701, Korea, [4]Department of Statistics, Seoul National University, Seoul, 151–742, Korea, [5]School of Pharmacy, Sungkyunkwan University, Suwon, 440–746, Korea.

A rapid and sensitive method to determine the characteristics of carcinogens is needed. In this study, we used a microarray-based genomics approach, with a short-term *in vivo* model, in combination with insights from statistical and mechanistic analyses to determine the characteristics of carcinogens. Carcinogens were evaluated based on the different mechanisms involved in the responses to genotoxic carcinogens and non-genotoxic carcinogens. Gene profiling was performed at two time points after treatment with six training and four test carcinogens. We mapped the DEG (differentially expressed gene)-related pathways to analyze cellular processes, and we discovered significant mechanisms that involve critical cellular components. Classification results were further supported by Comet and Micronucleus assays. Mechanistic studies based on gene expression profiling enhanced our understanding of the characteristics of different carcinogens. Moreover, the efficiency of this study was demonstrated by the short-term nature of the animal experiments that were conducted.

Carcinogens can be categorized as either genotoxic (GTX) or non-genotoxic (NGTX), according to their specific pathogenic mechanism. Most GTX carcinogens are electrophiles that interact directly with DNA through the formation of covalent bonds, resulting in DNA-carcinogen complexes (DNA adducts). These complexes lead to various types of DNA damage, including the formation of cross-links between the two helices, chemical bonds between adjacent bases, removal of DNA bases (hydration) and cleavage of the DNA strands, all of which result in modifications to the information stored within the DNA. Such mutations are typically fixed by DNA repair mechanisms; however, if DNA replication occurs prior to the action of a repair mechanism, mutations can become permanent and may eventually cause tumors. Conversely, NGTX carcinogens have no direct interaction with DNA; they are believed to cause tumors by disrupting cellular structures and by changing the rate of either cell proliferation or of processes that increase the risk of genetic error.

These types of differences in the sub-mechanisms of carcinogenicity also affect the gene expression patterns of cells exposed to carcinogens, which encourages genomic approaches in toxicological studies[1–3]. Previous studies have found that GTX carcinogens activate p53 tumor suppressor gene products in response to DNA damage, which leads to the initiation of sub-mechanisms, including the activation of cell cycle arrest, apoptosis and DNA repair processes, and which results in changes in the expression of specific genes, such as Cdkn1a, Mdm2 and Bcl2[4,5]. NGTX carcinogens display complicated and varying mechanisms that are not completely understood. However, these mechanisms have been associated with an alteration in oxidative stress, modulation of metabolizing enzymes, induction of peroxisome proliferation, alteration of intercellular communication and disruption of the balance between proliferation and apoptosis[5,6]. Therefore, distinguishing GTX from NGTX carcinogens by gene expression profiling is possible, and indeed, many studies have attempted to classify unknown carcinogens using this method[7–11].

Table 1 | Carcinogens used. Three GTX carcinogens and three NGTX carcinogens were used as the training set, and four carcinogens were used as the test set

Chemical	Abbreviation	CAS-number	Genotoxicity	Dose	Vehicle
GTX carcinogens					
2-Acetamidofluorene	2-AAF	53-96-3	+	50 mg/kg	corn oil
3′ Methyl dimethylaminoazobenzene	3′MeDAB	3732-90-9	+	800 mg/kg	corn oil
N-nitroso-diethylamine	DEN	55-18-5	+	50 mg/kg	saline
NGTX carcinogens					
Clofibrate	CFB	637-07-0	-	600 mg/kg	saline
DL-Ethionine	ET	67-21-0	-	1000 mg/kg	corn oil
1,4-Dioxane	DIOX	123-91-1	-	1000 mg/kg	saline
Test					
1,3-Dichloro-2-propanol	DCP	96-23-1		50 mg/kg	saline
Urethane	URT	51-79-6		500 mg/kg	saline
Methyleugenol	ME	93-15-2		300 mg/kg	corn oil
Sodium nitrite	SN	7632-00-0		50 mg/kg	saline

To study the toxicology of a realistic biological condition, we used an *in vivo* rat model. Importantly, we selected the liver as the target organ because it metabolizes many different compounds and it also increases the toxicity of compounds by activating cytochrome P450, which promotes the electrophilicity of pro-carcinogens. We used microarrays to conduct gene expression profiling and traditional statistical methods for data analysis. Furthermore, we considered the influence of the administration time during the classification and analysis of the pathways affected by GTX and NGTX carcinogens.

This study was performed using 6 training carcinogens, including three known GTX carcinogens, 2-AAF (2-acetamidofluorene), 3′MeDAB (3′ methyl dimethyl–amino-azobenzene) and DEN (N-nitroso-diethylamine); and three known NGTX carcinogens, clofibrate, DL-ethionine and 1,4-dioxane. Differentially expressed genes (DEGs) were selected and mechanistically studied. Additionally, the selected DEGs were tested using four carcinogens, DCP (1,3-dichloro-2-propanol), urethane, methyleugenol and sodium nitrite (Table 1). *In vivo* liver Comet assays and *in vivo* micronucleus assays were used to validate the findings (Table 2).

Results

Preprocessing of microarray data. For each treatment type (single and multiple), 10 compounds were administered, and the experiment was repeated in triplicate. For DEG selection and mechanistic studies, a total of 41 data points (three GTX carcinogens, three NGTX carcinogens and their controls) were used, and the rest (24 data points, four test carcinogens) of the data points were used only in hierarchical clustering as a test set (Figure 1).

To confirm the pattern of overall data, a PCA (principal component analysis) was performed (Figure 2), and 27,458 genes of 41 data points were imported to ensure correction through normalization and filtering. Overall, there was no clustering between the GTX and NGTX groups, which indicated the importance of the DEG selection process. Because an organism maintains homeostasis, there are only a small number of genes for which the expression is changed significantly in an experiment. Consequently, it is difficult to discover expression traits caused by the injection of a carcinogen when taking the pattern of the whole genome into account because the few DEGs, when masked by many equally expressed genes (EEGs), become difficult to detect. Thus, statistical analysis, which enables the

Table 2 | Evaluation of *in vivo* genotoxicity in rat liver treated with hepatocarcinogens. The *in vivo* liver Comet assay and *in vivo* liver micronuclei assay using young rats (n=3)

Substances	Dose	*in vivo* liver Comet *(% tail DNA)*	*in vivo* liver micronuclei *(% MN-HEPs frequencies)*
GTX carcinogens			
	Vehicle Control	6.11 ± 4.62	0.03 ± 0.06
2-Acetamidofluorene	50 mg/kg	*16.32 ± 15.12	*0.45 ± 0.23
3′ Methyl dimethylaminoazobenzene	800 mg/kg	*18.72 ± 14.95	*1.02 ± 0.28
N-nitroso-diethylamine	50 mg/kg	*29.66 ± 20.74	*0.60 ± 0.05
NGTX carcinogens			
	Vehicle Control	5.38 ± 4.34	0.03 ± 0.06
Clofibrate	600 mg/kg	8.02 ± 6.56	0.02 ± 0.03
DL-Ethionine	1000 mg/kg	6.82 ± 5.94	0.10 ± 0.05
1,4-Dioxane	1000 mg/kg	6.68 ± 6.40	0.08 ± 0.08
Test			
	Vehicle Control	5.21 ± 7.39	0.01 ± 0.03
1,3-Dichloro-2-propanol	50 mg/kg	6.08 ± 7.97	0.05 ± 0.07
	25 mg/kg	4.49 ± 6.26	0.01 ± 0.03
	Vehicle Control	6.51 ± 8.21	0.01 ± 0.03
Urethane	500 mg/kg	13.60 ± 12.21	0.05 ± 0.04
	250 mg/kg	9.17 ± 10.03	0.08 ± 0.06
	Vehicle Control	6.97 ± 7.05	0.01 ± 0.03
Methyleugenol	300 mg/kg	7.64 ± 7.86	0.03 ± 0.03
	150 mg/kg	7.02 ± 6.31	0.03 ± 0.05
	Vehicle Control	6.27 ± 6.52	0.01 ± 0.03
Sodium nitrite	50 mg/kg	7.05 ± 8.47	0.06 ± 0.06
	25 mg/kg	7.41 ± 8.30	0.04 ± 0.05

*$P<0.01$, significantly different from the concurrent solvent control.

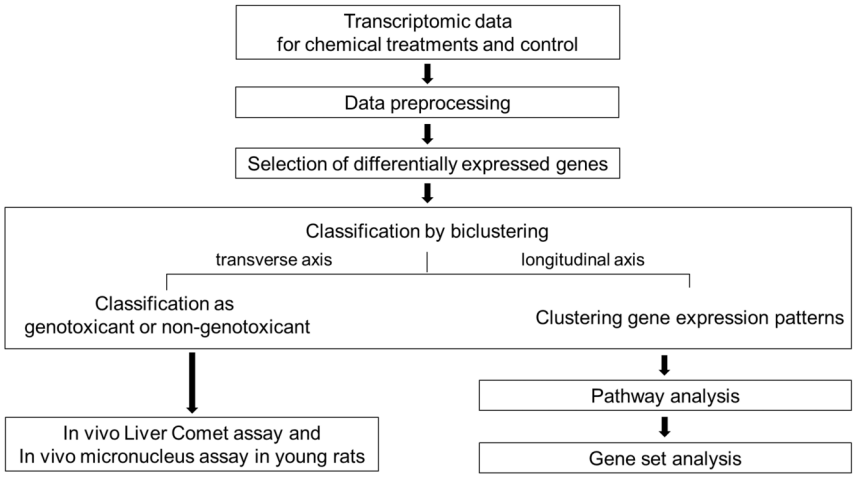

Figure 1 | **Schematic outline of the research protocol.**

selection of only significantly changed variables in genomics, is required. Importantly, only the multiple-treatment DL-ethionine data (light blue) showed a different PC2 value from the other treatments. Because all three data sets showed the same result, it was assumed that the unique quality of the compound may have caused the outcome. This difference was taken into consideration when analyzing and interpreting the results.

Selection of differentially expressed genes (DEGs). Selection of DEGs was conducted based on the results of multiple t-tests. We tried to maintain a level of confidence above 95%. The t-tests were performed in two steps. The first step profiled changes in expression between the control and each test group, and the second step selected DEGs with factors characteristic of GTX and NGTX carcinogens for single and multiple treatments. Ultimately, we compiled two lists of DEGs: single treatment, 91 DEGs, including Igfbp2, Cd36 and Ifrd1; and multiple treatments, 176 DEGs, including Hspb1, Ccng1 and Ifrd1. These genes were categorized on the basis of their major functions (Supplementary Table 1 and Table 2).

Classification by clusters and assays. Prior to the mechanistic study of DEGs, whether the applicability of classifying other carcinogens with the selected DEGs was tested (Figure 3). In the single treatment, all test data formed clusters with GTX. This suggests that under this treatment modality, test carcinogens show a gene expression pattern similar to GTX. Conversely, with multiple treatments, the test group, including DCP, methyleugenol, sodium nitrite and urethane, formed a cluster with the NGTX group, which suggests that the test carcinogens showed a gene expression pattern similar to that of the NGTX carcinogens employed under this multiple-treatment modality. The difference between single and multiple treatments suggests that the innate body response to single treatments was strong enough to diminish the unique expression caused by the compounds. The heat map of DEGs in Figure 3 distinguishes the GTX compounds from the NGTX compounds and allows for hierarchical clustering. The red color represents high expression, and the blue color represents low expression, within the −2.9∼2.9 (Figure 3A) and −3.5∼3.5 (Figure 3B) ranges.

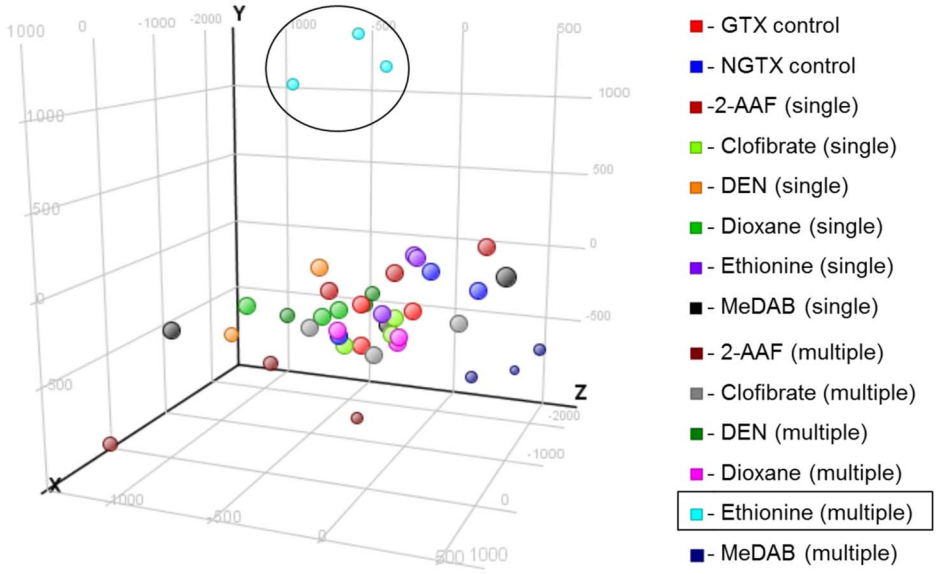

Figure 2 | **PCA (Principal Component Analysis).** The results are depicted 3-dimensionally with PC1 (35%), PC2 (12%) and PC3 (7%) as the X, Y and Z axes, respectively. Each color represents a different compound in single and multiple treatments, and one chip data are shown as a large circle, with 41 training data points on the graph. The data before the statistical analysis show no gathering between the GTX and NGTX groups; therefore, a selection process to detect the significant gene is needed. Only DL-ethionine (multiple), light blue, showed a different PC2 value from the other data; this was found when interpreting the results after the statistical analysis.

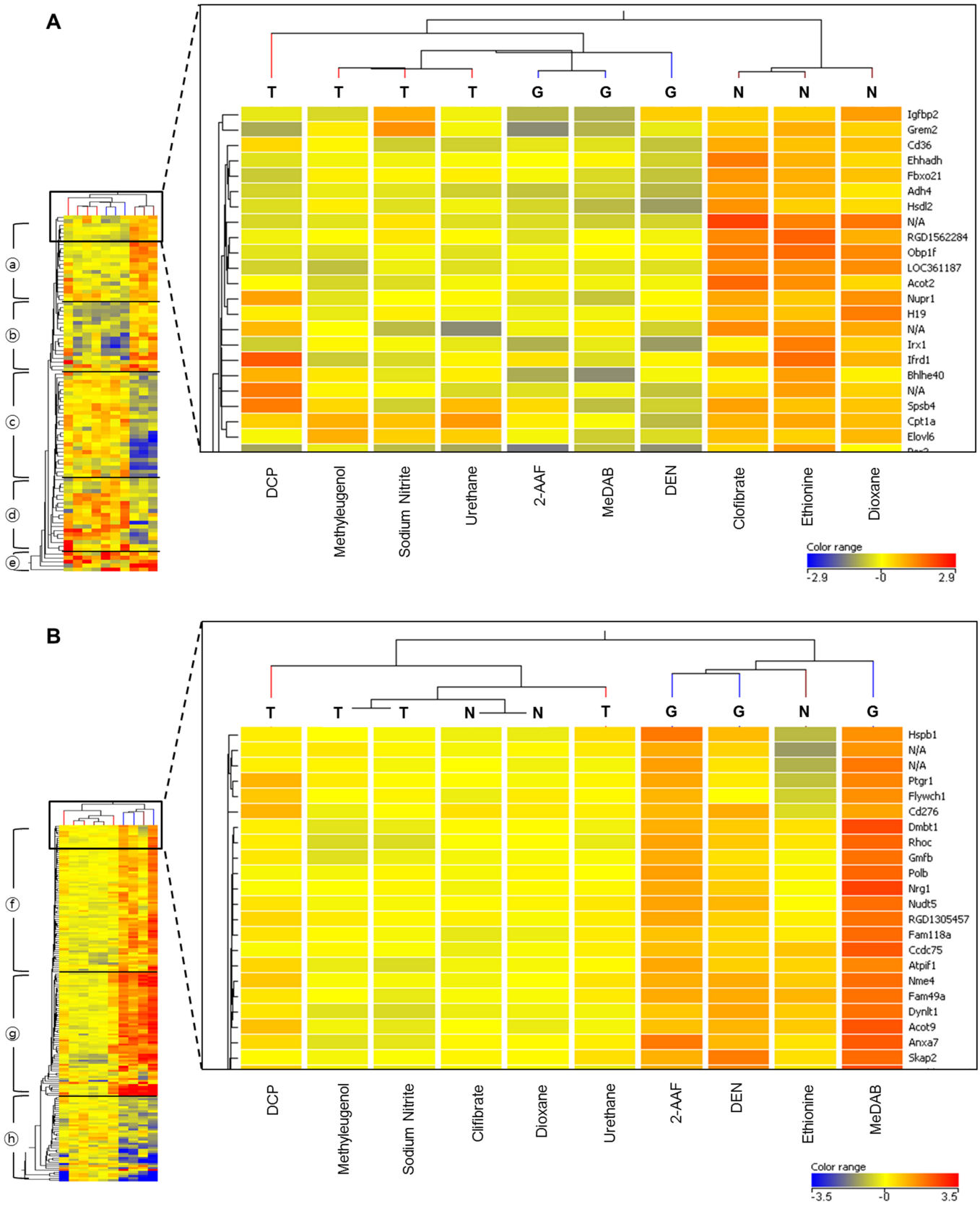

Figure 3 | DEG (differentially expressed gene) heat map and hierarchical clustering results distinguishing GTX from NGTX compounds. These data are from the training set of three GTX (G; 2-AAF, DEN, MeDAB), three NGTX (N; clofibrate, dioxane, DL-ethionine) carcinogens and the test set of four unknown carcinogens (T; DCP, methyleugenol, sodium nitrite, urethane). The expression patterns for the DEGs selected from the overall heat map on the left side are shown in color. The red color represents high expression and the blue color low expression, within the −2.9~2.9 (Figure 3A) and −3.5~3.5 (Figure 3B) ranges, respectively. The magnified section on the right shows parts of DEG lists and hierarchical clustering results of the upper part. (a): single-treatment results, (b): multiple-treatment results.

To evaluate the classification results, we performed Comet and micronucleus assays. The results showed that the training carcinogens, AAF, DEN and MeDAB, exhibited positive (genotoxic) results, whereas DL-ethionine, dioxane and clofibrate exhibited negative (non-genotoxic) results. Four compounds from the test set were confirmed as negative, regardless of dose (Table 2), which supported the classification results.

Pathway analysis. Pathway analysis was performed with 91 DEGs in the single treatment group and 176 DEGs in the multiple treatment group. Gene interactions were demonstrated based on the literature references in the database, and we uncovered the cause of changed expression by evaluating the relationships among the DEGs. Thus, we searched for the entire range of information on interactions between the DEGs, including regulation, expression, binding and protein modification. The information from this study was ultimately used to formulate a map (Figure 4), and the cellular processes for which there was a change in expression when a carcinogen was administered were revealed by the selected DEGs (Table 3). The groups in Table 3 are annotated as single/multiple and GTX/NGTX to better evaluate these two different variables (carcinogen and administration period).

Discussion

In this study, the GTX and NGTX carcinogenicity of the chemicals were predicted using the response mechanism of initial exposure. The DEGs, which were directly affected by the chemicals, were identified, and the cluster and pathway map showed the relationships between them.

Figure 3 is a bicluster, that is, a cluster of compounds widthwise and a cluster of genes lengthwise. On the right side of the magnified section, there are enumerated clustered genes lengthwise; on the left side of the magnified section, there is another clustering, which represents the clustering of selected DEGs, according to expression value. With respect to the gene-clustering results, the DEGs were divided into five categories for single treatments and three for multiple treatments. In the case of single exposure (Figure 3A), the five results were: ⓐ expression increases in NGTX, ⓑ decreases in GTX, ⓒ decreases in NGTX, ⓓ increases in GTX and ⓔ increases in both GTX and NGTX. Because the expression trend among the genes was likely to be inconsistent, the genes were divided into five groups. In addition, because there was no harmony in the expression trends in any of the five groups, effective clustering did not occur. In comparison with and following the verification of pathway analysis (Figure 4A), a specific pathway was not assigned to each group: thus, a change in expression by mechanism unit was not observed.

On the other hand, the biclustering results from multiple exposures (Figure 3B) revealed that all of the test data formed clusters with the NGTX data, indicating that the gene expression patterns of the test data were similar to those of the NGTX carcinogens. However, in this process, DL-ethionine was estimated to be a genotoxin despite being classified as a nongenotoxic chemical in the database. This is because DL-ethionine gains hepatotoxicity during strong oxidative stress[6] despite of having a low affinity for DNA[12]. The secondary gene damage caused by strong oxidative stress can induce the genotoxic response process. Thus, the gene expression patterns were similar to those of the genotoxic carcinogens. These results reflected the unique characteristics of DL-ethionine, suggesting the reliability of the classifications in this study. The classification results and the data from the Comet assays and micronucleus assays were all consistent.

Additionally, after the clustering of genes, it became clear that there was an expression trend according to the characteristics of the injected carcinogen. This group was divided into three categories: ⓕ expression increases moderately, ⓖ expression increases markedly and ⓗ expression decreases with the injection of the GTX carcinogen. Interestingly, with additional time following the injection of an NGTX carcinogen, the expression of the gene returned to the original level. However, with GTX carcinogens, increases or decreases in gene expression were more apparent as time passed. The differences between the characteristics of GTX and NGTX carcinogens, as observed through the clustering of the heat map and the genes, became more evident as time passed.

Because the multiple-exposure data showed a distinct expression pattern, an attempt was made to verify this pattern by comparing it with the results of the pathway analysis. The genes included in the pathway analysis in Figure 4B were divided into three clusters (Figure 3B): ⓕ 11, ⓖ 13 and ⓗ 5 each. The inclusion of the majority of genes, especially Ccng1, Mdm2, Fas and Cdkn1a, in the ⓖ group was significant. Because these genes are representatives of the p53 signaling pathway, they have been used as markers for genotoxicity in a number of different studies.

The reliability of DEGs as a tool for classifying treatments was confirmed by a series of individual pathway analyses using selected DEGs. The aim of this process was to explain why the selected DEGs were differentially expressed. Some DEGs altered one another's expression by direct interaction (Figure 4A), and this was observed in the map centered on Cdkn1a (p21), which is the target for p53 in the p53 signaling pathway that is expressed when DNA is damaged. In a series of responses to carcinogen treatment, a number of genes showed altered expression. In the map, the transcription factor Arntl, which is central to the formation of circadian rhythms, regulate the expression of Cdkn1a and controls cell growth[13]. Arntl as well as Npas2 and Per2 belong to the circadian rhythm pathway. Npas binds to Arntl and forms a hetero-dimer, which then activates expression of the *per* and *cry* genes; these genes are negative regulatory components of the circadian rhythm[14]. The Npas/Arntl hetero-dimer also up-regulates Bhlhb2 (Dec1), which exhibits a circadian rhythm[15]. Furthermore, among the DEGs, Cdkn1a is influenced by Klf6, Igfbp2, Rhob and Egr1. When over-expressed, Klf6 and SV1 down-regulate Cdkn1a and up-regulate a number of genes, including the pro-survival gene Bcl-2 and the oncogene *c-myc*[16]. Igfbp2 binds to p21Cip1/Waf1 (Cdkn1a) and participates in the control of cell proliferation. Rhob also regulates p21Cip1/Waf1 (Cdkn1a) and promotes tumor progression[17]. Egr1 is reported to interact directly with a specific sequence found in the gene promoter of Cdkn1a and to regulate its expression[18].

Understanding the expression of DEGs after multiple exposures (Figure 4B) is more complicated but can be achieved by referencing Cdkn1a. In the map, Cdkn1a, Bax, Mdm2, Fas and Ccng1 are the main proteins (genes) composing the p53 pathway. This pathway is essential to the regulation of growth and apoptosis during mutagenic stress. When p53 is activated, it induces the expression either of p21 (Waf1, Cip1), which participates in cellular arrest at the G1-S transition, or of Bax, PIGs, IGF-BP3, Fas, Fas-L and DR5[19]. In addition to the p53 pathway, many genes interact with Cdkn1a. Nrg1 induces Cdkn1a and promotes cancer cell proliferation[20]; Itgav, which is independent of p53, reduces the expression of Cdkn1a, leading to the inhibition of cell-death pathways[21]. Tyms also suppresses the expression of Cdkn1a via Mdm2, an oncogene, and demonstrates its influence in a cancer model[22]. Conversely, Dcn increases the expression of Cdkn1a and allows the cell cycle to remain in the G1 phase[23]. Many proteins (genes), including Hspb1, Hmox1 and Cyr61, increase the expression of Cdkn1a and induce cell cycle arrest when carcinogens are administered[24,25]. Most of the DEGs that do not belong to the p53 pathway are related to xenobiotic clearance or to the metabolism of endobiotics. Cyp1a1 and Lpl are involved in aromatic amino acid metabolism and triacylglycerol degradation, respectively; Cpt1a is involved in fatty acid oxidation, whereas Abcb1 (P-GP) affects xenobiotic clearance[26,27].

The information gained by the analysis of such pathways is abundant and detailed, but the patterns of gene expression cannot be easily understood. However, a gene set analysis reveals the entire

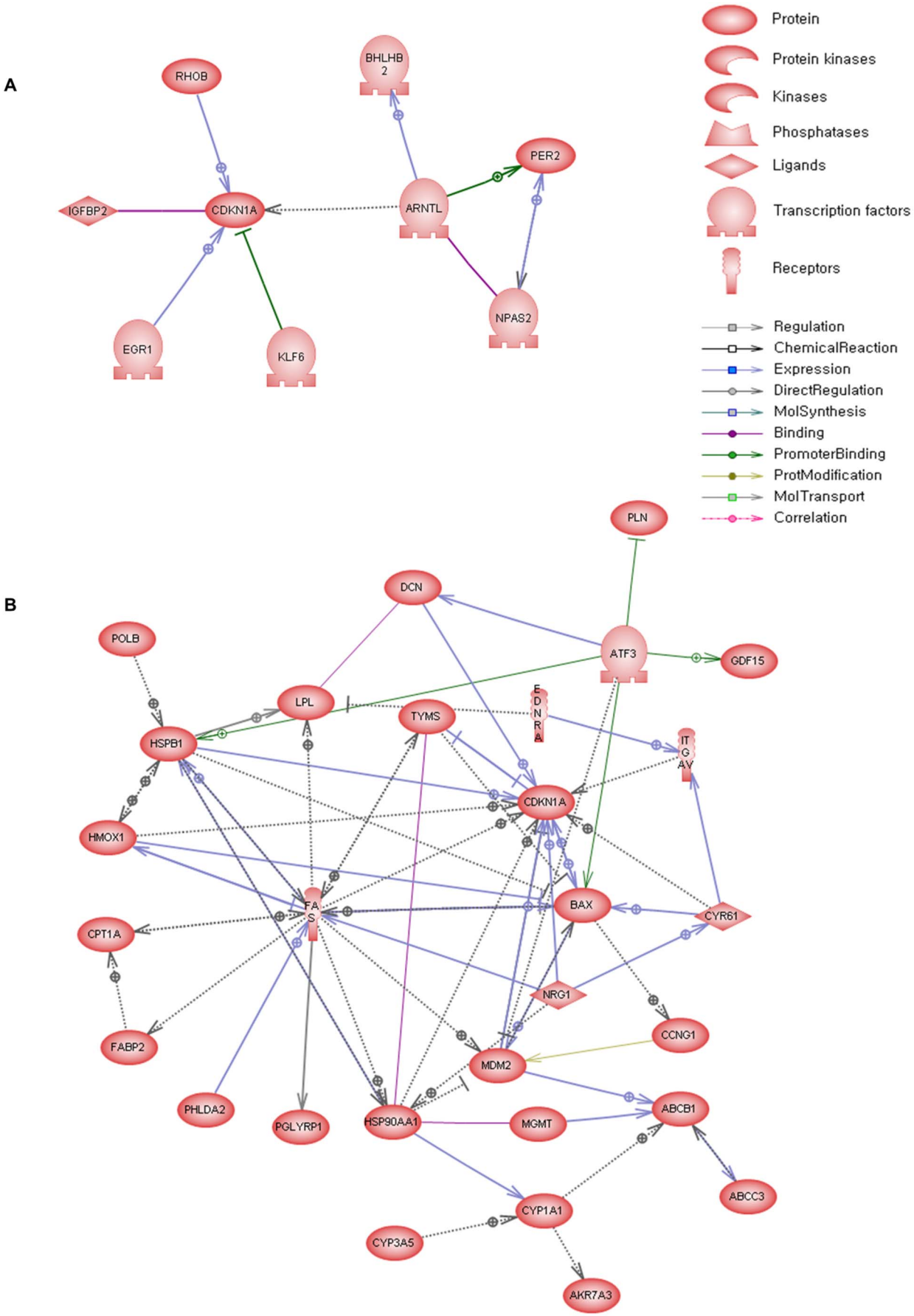

Figure 4 | **Pathway mapping among DEGs.** Each entity represents a protein transcribed from the selected DEG, and the arrows indicate the connections. (a) DEG pathway analysis results between GTX carcinogens and NGTX carcinogens in single treatment. (b) DEG pathway analysis results between GTX carcinogens and NGTX carcinogens in multiple treatments.

Table 3 | Pathways that changed expression according to each condition (based on DAVID). The EASE Score associated with each annotation term inside each cluster is identical to the meaning/value of the p-value (Fisher Exact/EASE Score) that is shown in the regular chart report for the same terms

Single Exposure	Count*	Percentage*	EASE Score*
GTX carcinogen			
Glycine, serine and threonine metabolism	3	12.5	5.0E-3
NGTX carcinogen			
Circadian rhythm	3	6.1	1.6E-3
PPAR signaling pathway	4	8.2	5.1E-3
Fatty acid metabolism	3	6.1	1.8E-2
Glycine, serine, threonine metabolism	3	6.1	1.9E-2
Metabolism of xenobiotics by cytochrome p450	3	6.1	2.3E-2
Multiple Exposures	**Count***	**Percentage***	**EASE Score ***
GTX carcinogen			
p53 signaling pathway	5	3.9	3.1E-3
Pathways in cancer	8	6.2	2.8E-2
NGTX carcinogen			
-	-	-	-

*Count: genes from the list involved in this annotation category.
Percentage: involved genes/total genes.
EASE Score: modified Fisher Exact P-value.

expression picture. Thus, the two methods used in this study complement one another.

GTX carcinogens with single exposure (Table 3) significantly changed the expression of the genes in glycine, serine and threonine metabolism pathway, which is a pathway that is responsible for the synthesis and breakdown of several amino acids, such as glycine, serine and threonine[28]. Glycine repairs damaged tissues and promotes healing, whereas serine is important for RNA and DNA functions and for cell formation[29] and threonine is involved in fatty acid metabolism. Therefore, this pathway can play an important role in maintaining the biological functions that protect against stress caused by xenobiotics in the early stages of the administration of a carcinogen, but these responses are not prompted by specific circumstances, such as DNA damage. This theory is supported by the concurrent increase in activity of this pathway after treatment with either NGTX or GTX carcinogens and indeed under all conditions of xenobiotic administration.

After single exposure, NGTX carcinogens induced changes in the expression of the following pathways: circadian rhythm; PPAR signaling; fatty acid metabolism; metabolism of xenobiotics by cytochrome p450; and glycine, serine and threonine metabolism. The circadian rhythm pathway is involved in biological rhythms; and previous studies have found that when these rhythms are disrupted, melatonin (a hormone that inhibits the occurrence of tumors) decreases, while the occurrence of cancer increases[30]. Moreover, a recent study showed that in normal organisms, circadian rhythms and the cell cycle are strongly connected, as the cell cycle is under the control of circadian rhythms. In cancer cells, however, the two rhythms are separated, and because the cell cycle is not controlled by circadian rhythms, cell division can be stimulated[31]. In the case of the PPAR signaling pathway, the induction of PPAR by the administration of a peroxisome proliferator in rodents is related to the presence of hepato-carcinogens[32] and associated with H_2O_2 generation and lipid peroxidation[6]. The clofibrate used in this study is an example of a peroxisome proliferator, and many other NGTX carcinogens are thought to act via this mechanism. Additionally, the induction of PPAR through the administration of a peroxisome proliferator influences the β-oxidation of fatty acids and thus it is logical that fatty acid metabolism would also increase expression[33]. Cytochrome P450 produces oxygen free radicals during the process of xenobiotic chemical metabolism, which leads to oxidative stress. It is natural for P450 to be expressed during the initial stage of chemical exposure.

After multiple exposures to GTX carcinogens, the expression of the p53 signaling pathway and the pathways in cancer are altered, as would be expected, given that the former is a representative response pathway that is induced by DNA damage.

The data in Table 3 indicate that NGTX carcinogens cause a short-term effect, whereas GTX carcinogens exert a long-term influence after administration. These results are supported by the number of DEGs selected for each condition: for the GTX carcinogens, the number of DEGs increased from 30 after single exposure to 170 as the exposure time increased, but for the NGTX carcinogens, the number of DEGs decreased from 66, after a single exposure, to eight, after multiple exposures. This study indicates, therefore, that NGTX carcinogens produce their characteristic profile after a short-term exposure and that GTX carcinogens do so after a long-term exposure. Thus, time is a critical factor in the classification of these compounds. For the classification based on the microarray data, a long-term exposure model (3 days) can be helpful for increasing the accuracy of a clustering result-based classification.

This study utilized microarrays to classify GTX and NGTX carcinogens by gene expression profiling and employed a bioinformatic pathway analysis to evaluate the carcinogens. The DEG-based mechanistic study demonstrated that it is possible to understand the distinctive characteristics of GTX and NGTX compounds by using a small number of carcinogens.

Methods

A flowchart of the study procedure is provided in Figure 1.

Experimental design and compounds. To study the tumor inducing mechanism of the chemicals, typical genotoxic carcinogens, such as 2-acetylaminofluorene, diethylnitrosamine and 4-dimethylamino-3′-methyl azobenzene, and nongenotoxic carcinogens, such as clofibrate, 1,4-dioxane and DL-ethionine, were separately administered to the rats to observe the expression changes of the internal mechanism. Each experimental compound was assigned to two models, single treatment and multiple treatment models, and in each model, there were three Sprague-Dawley rats[34]. For the single treatment group, each carcinogen was administered once, and an autopsy was performed after 24 h. For the multiple treatment group, carcinogens were administered every 24 h for a total of three times, and an autopsy was performed 24 h after the last dose. The dose and negative control (vehicle solvents) used for each compound are summarized in Table 1. Additionally, four undefined chemicals, 1,3-dichloro-2′-propanol (96-23-1), urethane (51-79-6), sodium nitrite (7632-00-0) and methyleugenol (93-15-2), were tested in the same way to predict classifications based

on the mechanism. All compounds were obtained from Sigma-Aldrich (St. Louis, MO).

Animals and treatments. For this study, 6–7 weeks old male SD rats were obtained from the Department of Laboratory Animal Resources at the National Institute of Food and Drug Safety Evaluation (NIFDS), Seoul, Korea. The animals were housed with hardwood chips in polycarbonate cages in a room under 12/12 h light/dark cycles and maintained at a controlled temperature. They were given food and water *ad libitum*. All procedures were approved by the Institutional Animal Care and Use Committee of NIFDS (0901KFDA029).

RNA isolation and microarray protocol. Liver specimens were processed for the purposes of RNA extraction. RNA was isolated using TRIzol (Invitrogen, Carlsbad, CA), and quality control was performed using RNA 6000 Nano chips on an Agilent 2001 Bioanalyzer (Agilent Technologies, Germany). Total RNA was extracted from frozen tissue using TRIzol and was purified with an RNeasy mini kit (QIAGEN, Hilden, Germany). Total RNA (1 μg) was amplified using the Affymetrix one-cycle cDNA synthesis protocol. For each array, 15 μg of amplified biotin-cDNA was fragmented and hybridized to the Affymetrix Rat Genome 230 2.0 GeneChip array (Affymetrix, Santa Clara, CA) for 16 h at 45°C in a rotating hybridization oven. Slides were stained with streptavidin/phycoerythrin and washed to amplify the antibodies. Arrays with hybridized targets were scanned using an Affymetrix GeneChip Analysis System (GC450 Fluidics Stations, a Hybridization Oven 640, a GC3000 7G scanner), and the scanned images were analyzed using the GeneChip® Operating Software v1.4 (GCOS) (Affymetrix). Spots that were determined by visual inspection to be of poor quality were excluded from further analysis. Additional information and raw data are available in GEO [Gene Expression Omnibus]; http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE31307. These data are available to the public and we published another paper that dealt with statistical methods from this result[35].

Preprocessing. For each treatment type (single and multiple), 10 compounds were administered, and the experiments were repeated in triplicate. Additionally, there were three control groups each for GTX and NGTX. Thus, there were 66 data points (10 (carcinogens) × 2 (treatment types) × 3 (repeats) + 3 (GTX control) + 3 (NGTX control) = 66). However, one data point had a technical error and was therefore excluded. As a result, sixty-five microarray data points were used in this study. These data are available at the GEO [Gene Expression Omnibus]; http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE31307. For the DEG selection and mechanistic studies, a total of 41 data points (3 GTX, 3 NGTX carcinogens and their controls) were used, and the remaining 24 data points were only used in hierarchical clustering as a test set. Preprocessing, PCA, DEG selection and clustering were performed by GeneSpring GX (Agilent Technologies). Each '.CEL' file obtained as a raw data file was standardized using the RMA (Robust Multi-array Average) algorithm. By using the RMA algorithm and the probe-level data, we were able to perform background signal correction, normalization and summarization of the PM values. Filtering was based on the expression values of the raw data: probes with one or more values in the percentile range of 20–100% (27,458 of 31,099 total chip probes) were selected. We then performed a PCA of the samples; a 3-dimensional plot of the PC1, PC2 and PC3 results (on the X, Y and Z axes, respectively) is provided in Figure 2. Each color represents a single compound yielded by single or multiple treatments, and each sphere indicates the data from a single chip. There are a total of 41 data points (6 carcinogens and controls) on the graph.

Selection of differentially expressed genes (DEGs). We performed data quality control and selected genes that had a greater than a two-fold increase/decrease in expression relative to the controls for each condition as DEGs. We performed t-tests for each GTX and NGTX carcinogen and selected genes that were significantly changed. The cut-off for significance was considered to be below 0.05. The following four gene lists were selected for further analysis: (i) control: single exposure to GTX group; (ii) control: single exposure to NGTX group; (iii) control: multiple exposures to GTX group; and (iv) control: multiple exposures to NGTX group. These groups were coupled together for analysis to select the final DEGs. To evaluate the single-treatment DEGs for GTX and NGTX carcinogens, we performed t-tests and fold-change analyses for the sum of the (i) and (ii) lists. To evaluate the multiple-treatment DEGs for GTX and NGTX carcinogens, we performed t-tests and fold-change analyses for the sum of the (iii) and (iv) lists. The two resulting lists (one for single treatment and one for multiple treatments) are presented in Supplementary Table 1 and Table 2, respectively.

Classification by biclustering. Based on the expression changes in the profiled gene, carcinogen classification was attempted. Grouping was performed by using hierarchical cluster and the centroid linkage rule was used. Therefore, the distance between two clusters is the average distance between their respective centroids. Additionally, a Euclidean distance measurement algorithm was applied. Figure 3 shows a heat map and biclustering (2-dimensional hierarchical clustering) for the single treatment (Figure 3A) and multiple treatment (Figure 3B) DEG lists. The clusters are 2-AAF, MeDAB and DEN in the GTX group; clofibrate, DL-ethionine and dioxane in the NGTX group; and methyleugenol, sodium nitrite and urethane, in the test group.

In vivo liver comet assay. Animals were euthanized 3 h after chemical treatment (Table 2), and liver tissues were placed in ice-cold mincing buffer (20 mM EDTA and 10% DMSO in HBSS, pH 7.5). The tissues were rinsed with mincing buffer to remove any residual blood and were then minced with a pair of fine scissors to release the cells. The cell suspension was stored on ice for 15–30 seconds to allow large clumps to settle, and the supernatant cells were mixed with LMO agarose and loaded onto CometSlides™ (Trevigen, MD). The prepared slides were immersed in lysis solution (Trevigen, MD) at 4°C, followed by rinsing in purified water. The slides were then subjected to electrophoresis in an alkaline solution (300 mM NaOH, 1 mM EDTA, pH>13) at 20 volts and 4–9°C for 30 minutes. Following electrophoresis, the slides were immersed in neutralization buffer for 5 minutes, dehydrated by immersion in absolute ethanol, and allowed to air dry. The coded slides were stained with ethidium bromide and examined under fluorescence microscopy. Images were analyzed using the Comet assay program (Komet 5.5 Andor Technology, Belfast, UK) to calculate the percentage of Tail DNA.

In vivo micronucleus assay in young rats. Four-week-old male F344 rats were treated with each hepatocarcinogen, according to Suzuki's protocol (Table 2)[36]. Four days after the second administration, hepatocyte suspensions were prepared by liver perfusion with a collagenase solution, suspended in 10% neutral buffered formalin, and stored under refrigeration. Immediately prior to evaluation, the hepatocytes were fluorescently stained using the AO-DAPI method. In calculating the incidence of micronucleated hepatocytes (MNHEPS) and the number of mitotic cells, 2,000 cells per animal were examined under a fluorescence microscope.

Pathway analysis and gene set analysis. The DEG lists produced by the statistical analysis were imported by PathwayStudio 7.0 and were then transformed into a suitable protein identity and mapped with the ResNet 7.0 database. We selected the 'add direct interactions' algorithm to identify direct relationships between the selected DEGs, and we showed all interactions, including the regulation, expression, binding and protein modification of genes, with literature references.

Cellular processes were analyzed using DAVID (Database for Annotation, Visualization and Integrated Discovery). To obtain results that were significantly relevant to the KEGG pathway, we confirmed the EASE Score (< 0.05). Confirmation of the list of resulting pathways and of individual information on DEG identities was performed using the Entrez gene.

1. Zhao, Y., Xie, P. & Fan, H. Genomic profiling of microRNAs and proteomics reveals an early molecular alteration associated with tumorigenesis induced by MC-LR in mice. *Environ Sci Technol* **46**, 34–41 (2012).
2. Benninghoff, A. D. *et al.* Promotion of hepatocarcinogenesis by perfluoroalkyl acids in rainbow trout. *Toxicol Sci* **125**, 69–78 (2012).
3. Silva Lima, B. & Van der Laan, J. W. Mechanisms of nongenotoxic carcinogenesis and assessment of the human hazard. *Regul Toxicol Pharmacol* **32**, 135–43 (2000).
4. Ellinger-Ziegelbauer, H., Stuart, B., Wahle, B., Bomann, W. & Ahr, H. J. Characteristic expression profiles induced by genotoxic carcinogens in rat liver. *Toxicol Sci* **77**, 19–34 (2004).
5. van Delft, J. H. *et al.* Discrimination of genotoxic from non-genotoxic carcinogens by gene expression profiling. *Carcinogenesis* **25**, 1265–76 (2004).
6. Uehara, T. *et al.* A toxicogenomics approach for early assessment of potential non-genotoxic hepatocarcinogenicity of chemicals in rats. *Toxicology* **250**, 15–26 (2008).
7. Sano, Y. *et al.* Trichloroethylene liver toxicity in mouse and rat: microarray analysis reveals species differences in gene expression. *Arch Toxicol* **83**, 835–49 (2009).
8. Jonker, M. J. *et al.* Finding transcriptomics biomarkers for in vivo identification of (non-)genotoxic carcinogens using wild-type and Xpa/p53 mutant mouse models. *Carcinogenesis* **30**, 1805–12 (2009).
9. Seidel, S. D., Stott, W. T., Kan, H. L., Sparrow, B. R. & Gollapudi, B. B. Gene expression dose-response of liver with a genotoxic and nongenotoxic carcinogen. *Int J Toxicol* **25**, 57–64 (2006).
10. Zeller, J. *et al.* Assessment of genotoxic effects and changes in gene expression in humans exposed to formaldehyde by inhalation under controlled conditions. *Mutagenesis* **26**, 555–61 (2011).
11. Matsumoto, H. *et al.* Discrimination of carcinogens by hepatic transcript profiling in rats following 28-day administration. *Cancer Inform* **7**, 253–69 (2009).
12. Ellinger-Ziegelbauer, H., Gmuender, H., Bandenburg, A. & Ahr, H. J. Prediction of a carcinogenic potential of rat hepatocarcinogens using toxicogenomics analysis of short-term in vivo studies. *Mutat Res* **637**, 23–39 (2008).
13. Llamas, B., Verdugo, R. A., Churchill, G. A. & Deschepper, C. F. Chromosome Y variants from different inbred mouse strains are linked to differences in the morphologic and molecular responses of cardiac cells to postpubertal testosterone. *BMC Genomics* **10**, 150 (2009).
14. Uchida, T. *et al.* CO-dependent activity-controlling mechanism of heme-containing CO-sensor protein, neuronal PAS domain protein 2. *J Biol Chem* **280**, 21358–68 (2005).
15. Noshiro, M. *et al.* Liver X receptors (LXRalpha and LXRbeta) are potent regulators for hepatic Dec1 expression. *Genes Cells* **14**, 29–40 (2009).
16. Wu, J. & Lingrel, J. B. KLF2 inhibits Jurkat T leukemia cell growth via upregulation of cyclin-dependent kinase inhibitor p21WAF1/CIP1. *Oncogene* **23**, 8088–96 (2004).

17. Terrien, X. *et al*. Intracellular colocalization and interaction of IGF-binding protein-2 with the cyclin-dependent kinase inhibitor p21CIP1/WAF1 during growth inhibition. *Biochem J* **392**, 457–65 (2005).
18. Choi, B. H. *et al*. p21 Waf1/Cip1 expression by curcumin in U-87MG human glioma cells: role of early growth response-1 expression. *Cancer Res* **68**, 1369–77 (2008).
19. Mendoza-Rodriguez, C. A. & Cerbon, M. A. [Tumor suppressor gene p53: mechanisms of action in cell proliferation and death]. *Rev Invest Clin* **53**, 266–73 (2001).
20. Klein, E. A. & Assoian, R. K. Transcriptional regulation of the cyclin D1 gene at a glance. *J Cell Sci* **121**, 3853–7 (2008).
21. Erdreich-Epstein, A. *et al*. Endothelial apoptosis induced by inhibition of integrins alphavbeta3 and alphavbeta5 involves ceramide metabolic pathways. *Blood* **105**, 4353–61 (2005).
22. Geller, J. I., Szekely-Szucs, K., Petak, I., Doyle, B. & Houghton, J. A. P21Cip1 is a critical mediator of the cytotoxic action of thymidylate synthase inhibitors in colorectal carcinoma cells. *Cancer Res* **64**, 6296–303 (2004).
23. Reed, C. C., Gauldie, J. & Iozzo, R. V. Suppression of tumorigenicity by adenovirus-mediated gene transfer of decorin. *Oncogene* **21**, 3688–95 (2002).
24. Park, S. H., Lee, Y. S., Osawa, Y., Hachiya, M. & Akashi, M. Hsp25 regulates the expression of p21(Waf1/Cip1/Sdi1) through multiple mechanisms. *J Biochem* **131**, 869–75 (2002).
25. Lee, T. S., Chang, C. C., Zhu, Y. & Shyy, J. Y. Simvastatin induces heme oxygenase-1: a novel mechanism of vessel protection. *Circulation* **110**, 1296–302 (2004).
26. Bandyopadhyay, S. *et al*. Mechanism of apoptosis induced by the inhibition of fatty acid synthase in breast cancer cells. *Cancer Res* **66**, 5934–40 (2006).
27. Suzuki, M., Sugimoto, Y. & Tsuruo, T. Efficient protection of cells from the genotoxicity of nitrosoureas by the retrovirus-mediated transfer of human O6-methylguanine-DNA methyltransferase using bicistronic vectors with human multidrug resistance gene 1. *Mutat Res* **401**, 133–41 (1998).
28. McNeil, J. B. *et al*. Glycine metabolism in Candida albicans: characterization of the serine hydroxymethyltransferase (SHM1, SHM2) and threonine aldolase (GLY1) genes. *Yeast* **16**, 167–75 (2000).
29. Beliveau, G. P. & Freedland, R. A. Metabolism of serine, glycine and threonine in isolated cat hepatocytes Felis domestica. *Comp Biochem Physiol B* **71**, 13–8 (1982).
30. Garcia-Saenz, J. A. *et al*. Circulating tumoral cells lack circadian-rhythm in hospitalized metastasic breast cancer patients. *Clin Transl Oncol* **8**, 826–9 (2006).
31. Yeom, M., Pendergast, J. S., Ohmiya, Y. & Yamazaki, S. Circadian-independent cell mitosis in immortalized fibroblasts. *Proc Natl Acad Sci U S A* **107**, 9665–70 (2010).
32. Holden, P. R. & Tugwood, J. D. Peroxisome proliferator-activated receptor alpha: role in rodent liver cancer and species differences. *J Mol Endocrinol* **22**, 1–8 (1999).
33. Gonzalez, F. J. The peroxisome proliferator-activated receptor alpha (PPARalpha): role in hepatocarcinogenesis. *Mol Cell Endocrinol* **193**, 71–9 (2002).
34. Burczynski, M. E. An introduction to toxicogenomics (CRC Press, Boca Raton, FL, 2003).
35. Kim, S. C. *et al*. Stouffer's Test in a Large Scale Simultaneous Hypothesis Testing. *PLoS One* **8** (2013).
36. Suzuki, H., Shirotori, T. & Hayashi, M. A liver micronucleus assay using young rats exposed to diethylnitrosamine: methodological establishment and evaluation. *Cytogenet Genome Res* **104**, 299–303 (2004).

Acknowledgements

This research was supported by grants 08161KFDA565, 09161KFDA534 and09152KFDA645 from the KFDA. This research was also supported by the Bio & MedicalTechnology Development Program of the National Research Foundation (NRF) funded bythe Korean government (MEST) (No. 2011-0019639) and the Korean Ministry of Science,ICT & Future Planning (MSIP) under grant number NRF-2011-0019745 and NRF-2011-0030810.

Author contributions

Conceived and designed the experiments: S.C.K., J.H.P., J.L., J.Lee, S.W.K. Performed the experiments: S.J.L., S.C.K., Y.N.Y., K.H.K., J.H.K., J.E.K., W.S.L., S.W.K. Analyzed the data: S.J.L., S.C.K., W.J.L., Y.K., J.L., S.W.K. Contributed reagents/materials/analysis tools: S.J.L., Y.N.Y., S.S., S.N.P., S.W.K. Wrote the manuscript: S.J.L., S.C.K., J.L., S.W.K.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Lee, S.J. *et al*. Distinguishing between genotoxic and non-genotoxic hepatocarcinogens by gene expression profiling and bioinformatic pathway analysis. *Sci. Rep.* **3**, 2783; DOI:10.1038/srep02783 (2013).