Gene expression profiling in primary mouse hepatocytes discriminates true from false-positive genotoxic compounds

K. Mathijs1,2, K. J. J. Brauers1, D. G. J. Jennen1,2, D. Lizarraga1,2, J. C. S. Kleinjans1,2 and J. H. M. van Delft1,2,*

1Department of Health Risk Analysis and Toxicology, Maastricht University, Universiteitssingel 50, 6229 ER Maastricht, The Netherlands and 2Netherlands Toxicogenomics Center, Universiteitssingel 50, 6229 ER Maastricht, The Netherlands

*To whom correspondence should be addressed. Department of Health Risk Analyses and Toxicology, Faculty of Health, Medicine and Life Sciences, Maastricht University, Universiteitssingel 50, 6229 ER Maastricht, The Netherlands

Well-established in vitro methods for testing the genotoxic potency of chemicals—such as the Ames/Salmonella test, the mouse lymphoma assay, the micronucleus test and the chromosomal aberration test—show a high false-positive rate for predicting in vivo genotoxicity and carcinogenicity. Thus, there is a need for more reliable in vitro assays. We investigated whether gene expression profiling in metabolically competent primary mouse hepatocytes is capable of discriminating true genotoxic (GTX) compounds from false-positive genotoxic (FP-GTX) compounds. Sandwich-cultured primary hepatocytes from male C57Bl6 mice were treated for 24 and 48 h with five true GTX and five FP-GTX compounds. Whole genome gene expression modifications were analysed by means of Affymetrix mouse genome 430 2.0 microarrays. Filtered genes were used for hierarchical clustering and class prediction methods. Classifiers were generated by prediction analysis of microarray using a leave-one-compound-out method and selecting the genes that were common to the 10 training replicates (16). Consequently, our hypothesis is that gene expression profiling in primary mouse hepatocytes is capable of discriminating between true GTX and FP-GTX compounds.

Introduction

Chemical compounds that are capable of causing gene mutations or chromosomal damage are defined as genotoxic (GTX) compounds. In most cases, genotoxicity is initiated by covalent binding of such chemicals or their metabolites to DNA (1,2). If not or incorrectly repaired, the DNA damage may lead to gene or chromosome mutations and—in a worst case—may ultimately induce the formation of tumours (1,2).

Well-established in vitro systems frequently used to identify the genotoxic potency of chemicals are the bacterial Ames test, the mouse lymphoma assay, the micronucleus test and the chromosomal aberration test (3). These classic in vitro genotoxicity tests, however, have been shown to generate an extremely high false-positive rate in terms of erroneously predicting genotoxicity and carcinogenicity in vivo (3,4). The specificity of the bacterial Ames test, for example, is 73.9%, while the mouse lymphoma assay, the micronucleus test and the chromosomal aberration test each demonstrate a specificity even <45% (3,5). A false-positive outcome from the in vitro test battery means that an unnecessary increased number of compounds are subjected to earlier and additional in vivo genotoxicity testing (4), which implies that without reason animals are used. This leads to costs and animal burden, based on the wrong premises. False-negative results, however, are very low in these well-established in vitro systems. Therefore, novel in vitro systems need to be developed that are capable of reliably discriminating genotoxins from non-genotoxins.

The liver is the principal organ involved in the biotransformation of xenobiotic substances, with its capacity to convert hydrophobic compounds into water-soluble products that can be readily secreted from the body. This transformation first involves oxidative activation to intermediates that are often highly reactive and can bind to DNA, causing DNA damage (6). Therefore, the liver represents a major target organ for chemical carcinogens (7,8). The use of in vitro liver systems is therefore of relevance for the screening of compounds for their carcinogenic properties.

The application of global gene expression profiling in toxicology is widely considered to possess the potential to provide new biomarkers for predicting hazard risks of chemicals for humans. In parallel with identifying a characteristic set of genes from a database of reference profiles in order to predict a compound’s unknown toxicity, these methodologies may hypothetically enhance our understanding of molecular mechanisms underlying the compound’s toxic effects (1,9). Gene expression profiling has previously been shown to be a valuable tool for investigating mechanisms of toxicity in several in vitro liver systems and in vivo (10–14). In a previous study, we successfully applied gene expression profiling in primary mouse hepatocytes to discriminate genotoxic from non-genotoxic compounds and this discrimination improved with increasing treatment period (15). After a 42-h recovery period of the cells, baseline gene expression in primary mouse hepatocytes was shown before to have only minor variance between biological replicates (16). Consequently, our hypothesis is that gene expression profiling in primary mouse hepatocytes can also be used for other class discriminations.

In this context, the current study focuses on the application of chemically induced altered gene expression profiling to
primary mouse hepatocytes as an in vitro system to discriminate GTX compounds from false-positive genotoxic (FP-GTX) compounds. FP-GTX compounds are here defined as non-carcinogens, demonstrating genotoxicity only in in vitro tests, but not in vivo in rodent bioassays. We focus on primary mouse hepatocytes because (i) they are metabolically competent, (ii) transgenic mouse models may allow relevant mechanistic follow-up studies and (iii) the complete sequence of the mouse genome is available (17). Cells were challenged with five true GTX compounds and five FP-GTX compounds, and a list of classifiers was generated to discriminate these two classes. Subsequently, this classification was validated by an additional set of two GTX and three FP-GTX compounds. The formation of phosphorylated H2AX (γH2AX) foci, which particularly monitor double-strand breaks as histone H2AX becomes phosphorylated on serine-139 by ataxia telangiectasia mutated (ATM) kinase and accumulates at the breaks (18–20), was assessed in order to find genes which correlate with DNA damage induction and which may be good classifiers to discriminate the two classes.

Materials and methods

Chemicals
Dulbecco’s modified Eagle’s medium, foetal calf serum, Hanks’ calcium- and magnesium-free buffer, Alexa fluor 480 goat anti-mouse IgG antibody, insulin and Trizol were obtained from Invitrogen (Breda, The Netherlands). Glucagon, hydrocortisone, collagenase type IV, benzo[a]pyrene (BaP), aflatoxin B1 (AFB1), 2-acetylaminofluorene (2-AAF), dimethylnitrosamine (DMN), mitomycin C (MitC), o-anthracene, 2-(chloromethyl)pyridine.HCl (2-CP), 4-nitro-o-phenylenediamine (4-NP), quercetin (Q), 8-hydroxyquinoline, curcumin (Cur), resorcinol, ethylacrylate, phenacetin, trypan blue, dimethylsulphoxide (DMSO), bovine serum albumin, 4',6-diamidino-2-phenylindole and Tween-20 were purchased from Sigma–Aldrich (Zwijndrecht, The Netherlands). 7,12-Dimethylbenz[a]anthracene (DMBA) was obtained from MP Biologicals (formerly ICN Biomedicals; Morgan Irvine, CA), Triton X-100, NaCl, Na2HPO4.2H2O and NaH2PO4 were obtained from Merck (Darmstadt, Germany), weighing 20–25 g, were obtained from Charles River GmbH. This mouse strain was chosen because it is frequently used in toxicological and pharmacological investigations, and it provides a common background for transgenic mouse strains. The animals were housed in macrolon cages with sawdust bedding at 22°C and 50–60% humidity. The light cycle was 12 h light–dark. Feed and tap water were available ad libitum.

Isolation of hepatocytes
Hepatocytes were isolated from adult male C57/B6 mice by a two-step collagenase perfusion method according to Seglen and Casciano (21,22), with modifications as described before (16). Cell viability and yield were determined by trypan blue exclusion.

Cell culturing and treatments
Cells with viability >85%, were cultured in a collagen–collagen sandwich formation as described before (16,23,24). Dead cells are removed during this procedure. Prior to treatment, primary cultures of mouse hepatocytes were allowed to recover for 40–42 h at 37°C in a humidified chamber with 95%/5% air/CO2 in serum-free culture medium supplemented with insulin (0.5 U/ml), glucagon (7 ng/ml), hydrocortisone (7.5 μg/ml) and 2% penicillin/streptomycin (5000 U/ml penicillin; 5000 μg/ml streptomycin). Culture medium was refreshed every 24 h. After the recovery period, the culture medium was replaced by culture medium containing one of the selected 10 compounds or with vehicle control. Only non-cytotoxic doses were used for each compound, as determined by the 3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide, a tetrazole assay (~80% viability after 24 h incubation) (see Table I). Cells were incubated for 24 or 48 h before being used for detecting double-stranded (ds) DNA breaks or harvested for RNA isolation by adding Trizol reagent. Three independent replicate biological experiments with hepatocytes from different mice were conducted for each compound.

Detection of dsDNA breaks
dsDNA breaks were detected by immunostaining of the phosphorylated histone H2AX (γH2AX) as described before (15,26). Samples were corrected for their

<table>
<thead>
<tr>
<th>Chemical</th>
<th>Abbreviation</th>
<th>CAS number</th>
<th>Solvent, dose (v/v %)</th>
<th>Dose</th>
<th>GTX in vitro (A)</th>
<th>GTX in vivo (B)</th>
<th>Carcinogen mouse liver</th>
<th>PAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>GTX carcinogens</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benzo[a]pyrene</td>
<td>BaP</td>
<td>50-32-8</td>
<td>DMSO, 0.5</td>
<td>30 μM</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Aflatoxin B1</td>
<td>AFB1</td>
<td>1162-65-8</td>
<td>DMSO, 0.5</td>
<td>15 μM</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2-Acetylaminofluorene</td>
<td>2-AAF</td>
<td>53-96-3</td>
<td>DMSO, 0.5</td>
<td>125 μM</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Dimethylnitrosamine</td>
<td>DMN</td>
<td>62-75-9</td>
<td>2 mM</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Mitomycin C</td>
<td>MitC</td>
<td>50-07-7</td>
<td>Ethanol, 0.5</td>
<td>5 μM</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Dimethylbenzanthracene</td>
<td>DBMA</td>
<td>57-97-6</td>
<td>DMSO, 0.5</td>
<td>500 μM</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Phenacinet</td>
<td>Phen</td>
<td>62-44-2</td>
<td>Ethanol, 0.5</td>
<td>1.5 mM</td>
<td>-</td>
<td>-</td>
<td>+/−</td>
<td>+/−</td>
</tr>
<tr>
<td>False positive GTX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>o-Anthracene acid</td>
<td>ANAC</td>
<td>118-92-3</td>
<td>DMSO, 0.5</td>
<td>2 mM</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2-(Chloromethyl)pyridine.HCl</td>
<td>2-CP</td>
<td>6959-47-3</td>
<td>DMSO, 0.5</td>
<td>125 μM</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>4-Nitro-o-phenylenediamine</td>
<td>4-NP</td>
<td>99-56-9</td>
<td>DMSO, 0.5</td>
<td>2 mM</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Quercetin</td>
<td>Q</td>
<td>117-39-5</td>
<td>DMSO, 0.5</td>
<td>200 μM</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>8-Hydroxyquinoline</td>
<td>8-HQ</td>
<td>148-24-3</td>
<td>Ethanol, 0.5</td>
<td>150 μM</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Curcumin</td>
<td>Cur</td>
<td>458-37-7</td>
<td>DMSO, 0.5</td>
<td>80 μM</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Ethyl acrylate</td>
<td>Ethylacrylate</td>
<td>140-88-5</td>
<td>Ethanol, 0.5</td>
<td>500 μM</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Resorcinol</td>
<td>Resorcinol</td>
<td>108-46-3</td>
<td>Ethanol, 0.5</td>
<td>2 mM</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

MLA, mouse lymphoma assay; MN, micronucleus test; CA, chromosomal aberrations. The carcinogenicity data is derived from the International Agency for Research on Cancer (http://monographs.iarc.fr/), the National Toxicology Program (http://ntp-server.niehs.nih.gov/), the Environmental Protection Agency (http://www.epa.gov/) or the Carcinogenic Potency Database (25). The genotoxicity data is mainly derived from Kirkland et al. (3) and the ntp-server (http://ntp-server.niehs.nih.gov/) and ~20 more references.
time-matched vehicle control and for the statistical analysis of dose-dependent effects, the t-test was used (P-value < 0.05).

RNA isolation

Total RNA was isolated from cultured mouse hepatocytes using Trizol and by means of the RNeasy kit according to the manufacturer’s protocol. RNA concentrations were measured by means of a spectrophotometer and the quality of each RNA preparation was determined by means of a bio-analyser (Agilent Technologies, Amstelveen, The Netherlands). Only samples with a good quality (clear 18S and 28S peaks and RNA integrity number > 6) were used for hybridisation. Extracted RNA was stored at −80°C until further analysis.

Whole genome gene expression analysis

Tissues were prepared according to the Affymetrix protocol, complementary RNA targets were hybridised according to the manufacturer’s recommended procedures on high-density oligonucleotide gene chips (Affymetrix Mouse Genome 430 2.0 GeneChip arrays). The gene chips were washed and stained using an Affymetrix fluidics station and scanned by means of an Affymetrix GeneArray scanner.

A total of 82 GeneChips was run (One chip per RNA sample; 82 RNA samples were generated from three experiments with two time points for 10 compounds and solvent controls). Normalisation quality controls, including scaling factors, average intensities, present calls, background intensities, noise and raw Q values, were within acceptable limits for all chips. Hybridisation controls Bb, Bc, Bd and CreX were identified on all chips and yielded the expected increases in intensities.

The data expression data discussed in this publication have been deposited in ArrayExpress (accession numbers: E-MEXP-2209 and E-MEXP-2539 for the training compounds and E-MEXP-2636 for the test compounds), the European Bioinformatics Institute database (http://www.ebi.ac.uk/arrayexpress/).

Data analysis

Filtering of genes for the training compounds. Eighty-two datasets were obtained from this experiment. Raw data were imported into ArrayTrack (27,28) and normalised using Robust Multi-array Average (RMA, integrated into ArrayTrack) (29).

Present–absent calls were used to identify and omit genes with probe sets (a collection of probes designed for a given gene) of poor quality or with low expression values (30). Genes with two or more absence calls within the three replicates for all the treatments as well as for the vehicle controls were omitted. Subsequently, the remaining genes were logaritically (base 2) transformed, corrected for their time-matched vehicle control and subjected to all further analyses (24 h: 26 100 genes; 48 h: 26 690 genes; total: 27 363 genes). Genes for which expression was up- or down-regulated by at least one compound at a minimum fold change of 1.2-fold in at least two of three experiments with expression altered in the same direction in all replicates and with a mean fold up- or down-regulation of 1.5 were selected (31). This was repeated for each time point and all these genes were combined in one list. The generated list with filtered genes (log 2 ratios) was used for hierarchical clustering analysis (HCA) and prediction analysis of microarray (PAM) (24 h: 10 776 genes; 48 h: 12 180 genes; see supplementary file 1 in ArrayExpress, available at Mutagenesis Online).

Class prediction and functional analysis of classifiers. The software tool ‘PAM’, developed at Stanford University, was used for discriminating true GTX compounds from FP-GTX compounds (32). PAM uses gene expression data to calculate the shrunken centroid for each class and identifies the specific genes that determine the centroid most. Based on the nearest shrunken centroid, PAM is also capable of predicting to which class an unknown sample belongs and gives comparable results with other classification methods (32,33). Class prediction was performed after 24 and 48 h of exposure.

For the PAM analysis, the filtered gene lists were used. Although pre-selection of genes for PAM is not required, in our hands, it improved the performance of PAM. For each exposure period, 10 sets of genes (classifiers) were generated by PAM, each using all nine compounds (thus leaving out one compound), based on the smallest estimated classification error rate (generated by 10-fold cross-validation) and a >80% predicted test probability. The one compound that was left out was used for predicting its class (true GTX or FP-GTX) using the classifiers derived from the nine other compounds in the dataset. For each time point, the classifiers that were common to the 10 training sets were set as the optimal classifier set for that time point (24 h: 215 classifiers; 48 h: 220 classifiers; see supplementary file 2 in ArrayExpress, available at Mutagenesis Online). Five extra GTX and FP-GTX compounds were used to validate the classification model using the selected classifiers (Table I). Forty-eight datasets were obtained from this experiment. Raw data were again imported into ArrayTrack (27,28) and normalised using RMA (integrated into ArrayTrack) (29). The 215 and 220 classifier genes for 24 and 48 h, respectively, were selected and used for validation.

The classifiers from each time point were further analysed for over-representation analysis in MetaCore (GeneGo, San Diego, CA) to identify the involvement of these classifiers in specific cellular pathways, which included genes sets from Gene Ontology terms. The total list of genes, without genes of poor quality, were selected for each time point as described before (24 h: 26 100 genes; 48 h: 26 690 genes; total: 27 363 genes; see supplementary file 3 in ArrayExpress, available at Mutagenesis Online) and used as a background list in MetaCore. Significant pathways were selected by P < 0.05 and three or more classifiers involved in each process.

Correlation analysis. Genes that correlated with γH2AX foci formation were identified using the gene expression profile analysis suite (GEPAS 4.0; CIPF, Valencia, Spain). The values for γH2AX foci formation, corrected by subtracting the values for DMSO control from the log (base 2)-transformed values for the exposed samples, were correlated with the expression changes of filtered genes. Spearman correlation coefficients were calculated and correlating genes (P < 0.05) were selected for PAM.

Results

Fifteen compounds were selected to examine the usefulness of sandwich-cultured primary mouse hepatocytes for discriminating true GTX compounds from their FP-GTX compounds (Table I). All true GTX compounds were previously shown to be genotoxic in vitro and in vivo and to be carcinogenic in mouse liver. The FP-GTX compounds are all genotoxic in at least two different in vitro tests, but not genotoxic in in vivo tests and they are non-carcinogens regarding to mouse liver (3,4).

Hierarchical clustering analyses

To visualise possible similarities and differences between gene expression profiles from the 10 training compounds, HCA was performed by including only the genes filtered out based on consistent fold changes per compound for the 24- and 48-h exposure, respectively (24 h: 10 776 genes; 48 h: 12 180 genes) (Figure 1). After 48 h of exposure, but not after 24 h, replicates from each compound cluster perfectly together. At none of the time points, however, a complete clustering of true or FP-GTX compounds occurs. At 48 h, MitC and DMN cluster separately from all other compounds, and 2-AAF clusters together with the FP-GTX compounds. All the FP-GTX compounds, with the exception of 4-NP, cluster together. The FP-GTX compound 4-NP is even clustered separately from all other compounds.

Identification of classifiers

In order to identify classifier genes whose expression changes best discriminate true GTX compounds from the FP-GTX compounds, the class prediction software tool PAM was used. For each time point, initially 10 lists of classifiers were generated, based on analyses by PAM with the leave-one-out method. The compound left out (consisting of three replicates), was used for validation. Ultimately, the final list of classifiers for each time point consists of those, which were in common between all these 10 training sets. This resulted in 215 classifiers after 24 h of exposure and 220 classifiers after 48 h (supplementary file 2 is available at Mutagenesis Online). Ninety-one classifiers were in common between exposure periods of 24 and 48 h. The classifications of the one compound left out in each analysis, resulted in a perfect classification for 9 of the 10 compounds. Only 2-AAF was misclassified for all experiments at both time points (Table II).
Detection of γH2AX foci

Since DNA damage is expected to be introduced only by exposure to GTX compounds, it can be hypothesised that genes whose expression changes correlate with DNA damage might be a better classifier. Therefore γH2AX foci formation was detected by immunostaining, as these foci are a measure for DNA strand breaks. The γH2AX foci formation in relation to exposure to 10 compounds in comparison to vehicle controls for 24 and 48 h is shown in Figure 2. γH2AX foci in cells treated with true GTX compounds were significantly increased. Cells treated with FP-GTX compounds, with exception of 4-NP, showed no significant increase in γH2AX foci formation. The results for Q-treated cells, however, are less reliable as non-specific staining of the collagen layer was detected in these samples (data not shown).

We hypothesised that genes correlating with the formation of phosphorylated γH2AX foci might be good classifiers, and therefore they were also used for classification using PAM. This resulted in a similar outcome, i.e. a misclassification of 2-AAF (data not shown) and, therefore, did not improve the classification.

The values of γH2AX foci formation [logarithmically (base 2) transformed and corrected for vehicle control] were also added to the list with filtered genes for PAM in order to test whether this γH2AX foci formation would improve the class discrimination. This H2AX turned out to be on the total list of classifiers of 48 h but did not improve the classification.

Fig. 1. Hierarchical clustering of treatments using expression changes of filtered genes in primary mouse hepatocytes after (A) 24-h exposure and (B) 48-h exposure to GTX and FP-GTX compounds.

Table II. Prediction by PAM for GTX versus FP-GTX based on gene expression changes induced by compounds in primary mouse hepatocytes following 24- and 48-h treatments

<table>
<thead>
<tr>
<th>Compound</th>
<th>Genotoxic class</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFB1</td>
<td>GTX</td>
<td>GTX</td>
</tr>
<tr>
<td>BaP</td>
<td>GTX</td>
<td>GTX</td>
</tr>
<tr>
<td>2-AAF</td>
<td>GTX</td>
<td>GTX</td>
</tr>
<tr>
<td>DMN</td>
<td>GTX</td>
<td>GTX</td>
</tr>
<tr>
<td>MitC</td>
<td>GTX</td>
<td>GTX</td>
</tr>
<tr>
<td>2-CP</td>
<td>FP-GTX</td>
<td>FP-GTX</td>
</tr>
<tr>
<td>4-NP</td>
<td>FP-GTX</td>
<td>FP-GTX</td>
</tr>
<tr>
<td>ANAC</td>
<td>FP-GTX</td>
<td>FP-GTX</td>
</tr>
<tr>
<td>Q</td>
<td>FP-GTX</td>
<td>FP-GTX</td>
</tr>
<tr>
<td>8-HQ</td>
<td>FP-GTX</td>
<td>FP-GTX</td>
</tr>
</tbody>
</table>

Data from nine compounds were used for training, followed by testing the remaining compound for correct prediction. This was repeated until all compounds were tested.


detection of γH2AX foci

Since DNA damage is expected to be introduced only by exposure to GTX compounds, it can be hypothesised that genes whose expression changes correlate with DNA damage might be a better classifier. Therefore γH2AX foci formation was detected by immunostaining, as these foci are a measure for DNA strand breaks. The γH2AX foci formation in relation to exposure to 10 compounds in comparison to vehicle controls for 24 and 48 h is shown in Figure 2. γH2AX foci in cells treated with true GTX compounds were significantly increased. Cells treated with FP-GTX compounds, with exception of 4-NP, showed no significant increase in γH2AX foci formation. The results for Q-treated cells, however, are less reliable as non-specific staining of the collagen layer was detected in these samples (data not shown).

We hypothesised that genes correlating with the formation of phosphorylated γH2AX foci might be good classifiers, and therefore they were also used for classification using PAM. This resulted in a similar outcome, i.e. a misclassification of 2-AAF (data not shown) and, therefore, did not improve the classification.

The values of γH2AX foci formation [logarithmically (base 2) transformed and corrected for vehicle control] were also added to the list with filtered genes for PAM in order to test whether this γH2AX foci formation would improve the class discrimination. This H2AX turned out to be on the total list of classifiers of 48 h but did not improve the classification.

Fig. 1. Hierarchical clustering of treatments using expression changes of filtered genes in primary mouse hepatocytes after (A) 24-h exposure and (B) 48-h exposure to GTX and FP-GTX compounds.

Table II. Prediction by PAM for GTX versus FP-GTX based on gene expression changes induced by compounds in primary mouse hepatocytes following 24- and 48-h treatments

<table>
<thead>
<tr>
<th>Compound</th>
<th>Genotoxic class</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFB1</td>
<td>GTX</td>
<td>GTX</td>
</tr>
<tr>
<td>BaP</td>
<td>GTX</td>
<td>GTX</td>
</tr>
<tr>
<td>2-AAF</td>
<td>GTX</td>
<td>GTX</td>
</tr>
<tr>
<td>DMN</td>
<td>GTX</td>
<td>GTX</td>
</tr>
<tr>
<td>MitC</td>
<td>GTX</td>
<td>GTX</td>
</tr>
<tr>
<td>2-CP</td>
<td>FP-GTX</td>
<td>FP-GTX</td>
</tr>
<tr>
<td>4-NP</td>
<td>FP-GTX</td>
<td>FP-GTX</td>
</tr>
<tr>
<td>ANAC</td>
<td>FP-GTX</td>
<td>FP-GTX</td>
</tr>
<tr>
<td>Q</td>
<td>FP-GTX</td>
<td>FP-GTX</td>
</tr>
<tr>
<td>8-HQ</td>
<td>FP-GTX</td>
<td>FP-GTX</td>
</tr>
</tbody>
</table>

Data from nine compounds were used for training, followed by testing the remaining compound for correct prediction. This was repeated until all compounds were tested.

Detection of γH2AX foci

Since DNA damage is expected to be introduced only by exposure to GTX compounds, it can be hypothesised that genes whose expression changes correlate with DNA damage might be a better classifier. Therefore γH2AX foci formation was detected by immunostaining, as these foci are a measure for DNA strand breaks. The γH2AX foci formation in relation to exposure to 10 compounds in comparison to vehicle controls for 24 and 48 h is shown in Figure 2. γH2AX foci in cells treated with true GTX compounds were significantly increased. Cells treated with FP-GTX compounds, with exception of 4-NP, showed no significant increase in γH2AX foci formation. The results for Q-treated cells, however, are less reliable as non-specific staining of the collagen layer was detected in these samples (data not shown).

We hypothesised that genes correlating with the formation of phosphorylated γH2AX foci might be good classifiers, and therefore they were also used for classification using PAM. This resulted in a similar outcome, i.e. a misclassification of 2-AAF (data not shown) and, therefore, did not improve the classification.

The values of γH2AX foci formation [logarithmically (base 2) transformed and corrected for vehicle control] were also added to the list with filtered genes for PAM in order to test whether this γH2AX foci formation would improve the class discrimination. This H2AX turned out to be on the total list of classifiers of 48 h but did not improve the classification.

Fig. 1. Hierarchical clustering of treatments using expression changes of filtered genes in primary mouse hepatocytes after (A) 24-h exposure and (B) 48-h exposure to GTX and FP-GTX compounds.

Table II. Prediction by PAM for GTX versus FP-GTX based on gene expression changes induced by compounds in primary mouse hepatocytes following 24- and 48-h treatments

<table>
<thead>
<tr>
<th>Compound</th>
<th>Genotoxic class</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFB1</td>
<td>GTX</td>
<td>GTX</td>
</tr>
<tr>
<td>BaP</td>
<td>GTX</td>
<td>GTX</td>
</tr>
<tr>
<td>2-AAF</td>
<td>GTX</td>
<td>GTX</td>
</tr>
<tr>
<td>DMN</td>
<td>GTX</td>
<td>GTX</td>
</tr>
<tr>
<td>MitC</td>
<td>GTX</td>
<td>GTX</td>
</tr>
<tr>
<td>2-CP</td>
<td>FP-GTX</td>
<td>FP-GTX</td>
</tr>
<tr>
<td>4-NP</td>
<td>FP-GTX</td>
<td>FP-GTX</td>
</tr>
<tr>
<td>ANAC</td>
<td>FP-GTX</td>
<td>FP-GTX</td>
</tr>
<tr>
<td>Q</td>
<td>FP-GTX</td>
<td>FP-GTX</td>
</tr>
<tr>
<td>8-HQ</td>
<td>FP-GTX</td>
<td>FP-GTX</td>
</tr>
</tbody>
</table>

Data from nine compounds were used for training, followed by testing the remaining compound for correct prediction. This was repeated until all compounds were tested.
Validation of classifiers

For the purpose of validating the class discrimination models with the final classifier sets, gene expression data were generated for two additional true GTX compounds, phenacetin and DMBA, and for three FP-GTX compounds, cur, ethyl-acrylate and resorcinol and the vehicle control for exposure periods of 24 and 48 h. All the independent triplicate treatments of all compounds were classified correctly with a predicted test probability of 100% at both time points, with the exception of phenacetin, which is misclassified as a FP-GTX compound, only at 48 h (Table III). This resulted in a positive prediction value of 100% for both time points and a negative prediction value of 89 and 80% for 24 and 48 h, respectively.

Again the genes correlating with the formation of phosphorylated γH2AX foci were also fed into PAM for the test compounds in an attempt to improve the classification, but this resulted in a worse classification (data not shown).

Functional analyses of classifiers

The final classifiers as selected by PAM from the filtered genes from the each time point were used for over-representation analysis in MetaCore in order to reveal their biological function. Classifiers appeared to have a predominant role in several metabolic and biosynthetic processes, immune and damage responses, cellular organisation and cell cycle and apoptosis at both time points (Table IV; supplementary file 3). Genes which were up-regulated by the GTX compounds and down-regulated by the FP-GTX compounds were involved in apoptotic and immune processes. Genes which were down-regulated by the GTX compounds and up-regulated by the FP-GTX compounds were mainly involved in metabolic and biosynthetic amino acid processes (supplementary file 3 is available at Mutagenesis Online).

Discussion

The well-established in vitro test battery for genotoxicity has an extremely high false-positive rate for predicting in vivo genotoxicity and carcinogenicity (3,4). The predictive value of in vitro experiments therefore is poor and this implies for instance the risk that novel leads for possibly successful drugs are unnecessarily discarded, and that many chemicals are unnecessarily subjected to subsequent in vivo testing in rodent bioassays. Therefore, there is a need for novel in vitro systems for discriminating genotoxic from non-genotoxic compounds, which demonstrate a relatively low false-positive rate, thereby generating a realistic and not longer over-estimated demand for genotoxicity and carcinogenicity testing in long-term rodent bioassays. Therefore, it was the hypothesis for the current study, that gene expression profiles in primary mouse hepatocytes present a better reliable in vitro tool for discriminating FP-GTX compounds from true GTX compounds than currently available assays.

The method used for class prediction based on gene expression data is PAM (32). The use of this method and the study on the set of 10 training compounds, resulted in lists of genes, called classifiers, which can discriminate between true GTX and FP-GTX compounds. These were validated by using a test set of five more compounds. By applying PAM to the training set, surprisingly 2-AAF, a well-known true GTX compound that causes liver tumours in C57/B6 male mice and is genotoxic in mice (34) (http://ntp-server.niehs.nih.gov), is misclassified in our study as a FP-GTX compound, in all experiments and at both time points, while all the other compounds are always correctly classified. In our study, however, 2-AAF is able to induce γH2AX foci formation to a level comparable to that induced by other true GTX compounds, like BaP and AFB1. Apparently, in mouse hepatocytes, differences between 2-AAF and the other true GTX are pronounced only at the gene expression level. Around 16% of the filtered genes and almost one-third of the classifiers generated by PAM appeared to be affected in the opposite direction by 2-AAF as compared to the other GTX compounds; this includes genes involved in apoptosis (Ahr, Bax, Serpinb2 and several Caspases) and cell cycle (Cdkn1a). 2-AAF may have been misclassified by using gene expression profiles

### Table III. Validation of the class prediction model with five additional compounds

<table>
<thead>
<tr>
<th>Compound</th>
<th>Genotoxic class</th>
<th>Prediction</th>
<th>24 h</th>
<th>48 h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Exp 1</td>
<td>Exp 2</td>
</tr>
<tr>
<td>DMBA</td>
<td>GTX</td>
<td>GTX</td>
<td>GTX</td>
<td>GTX</td>
</tr>
<tr>
<td>Phen</td>
<td>GTX</td>
<td>GTX</td>
<td>GTX</td>
<td>GTX</td>
</tr>
<tr>
<td>Cur</td>
<td>FP-GTX</td>
<td>FP-GTX</td>
<td>FP-GTX</td>
<td>FP-GTX</td>
</tr>
<tr>
<td>Ethylacrylate</td>
<td>FP-GTX</td>
<td>FP-GTX</td>
<td>FP-GTX</td>
<td>FP-GTX</td>
</tr>
<tr>
<td>Resorcinol</td>
<td>FP-GTX</td>
<td>FP-GTX</td>
<td>FP-GTX</td>
<td>FP-GTX</td>
</tr>
</tbody>
</table>

The intersection of the classifiers from Table II, for each time point separately, was used for generating the classification model in PAM. The five new compounds were used for validating that model.

Fig. 2. Levels of dsDNA breaks (γH2AX foci) in primary mouse hepatocytes exposed to GTX and FP-GTX compounds for 24 or 48 h (average with SD of three experiments and corrected for solvent control). *Significant (P-value < 0.05) compared to solvent controls. ** Significant (P-value < 0.05) over time.
because 2-AAF induces opposite regulatory effects on genes involved in apoptosis and cell cycle, and precisely these genes are important for the classification of GTX compounds. It was observed in rat liver cells that 2-AAF is able to enhance cell replication and to suppress apoptosis and this might be the reason for the misclassification of 2-AAF in our study (35,36). Also reversed effects on genes involved in DNA repair (base excision, nucleotide excision, mismatch repair and DNA polymerases) are observed. This apparent contradiction, induction of DNA damage (γH2AX foci) but not of DNA damage response genes, by 2-AAF is not understood. We suggest that 2-AAF or one of its metabolites, is able to inactivate ATM or ATM and Rad3-related, which are both proximal components of DNA damage-induced cell cycle checkpoint pathways, as has been shown for caffeine (37).

Class prediction based on PAM with the test set resulted in a correct classification of all compounds at 24 h and only one misclassified compound, phenacetin, at 48 h, demonstrating that the obtained classifiers can indeed discriminate between true and false-positive GTX compounds. In view of the apparent misclassification of phenacetin, it is of interest that some doubts have recently risen about the genotoxicity of phenacetin (38). The most important metabolites of phenacetin are acetaminophen, phenacetin-3,4 epoxide, N-hydroxyphenacetin and 2-hydroxyphenetidine (38,39). Phenacetin is mainly metabolised to acetaminophen by cytochrome P450 1A2, but it is rapidly converted into a reactive metabolite, N-acetyl-p-benzoquinonimine and detoxified by conjugation with glutathione before excretion in the urine (39). Also, this acetaminophen was never associated with cancer formation (38). The other metabolites of phenacetin, however, are reactive intermediates and have been shown to contribute to the toxic and carcinogenic effects of phenacetin (40). But the metabolites of phenacetin are rapidly transferred from the liver to the kidney and are then further metabolised by the kidney for excretion (38). Therefore phenacetin may act as a weak GTX in the liver.

The classifiers generated by PAM were used for overrepresentation analysis in MetaCore for biological interpretation. It is not surprising that many of these classifiers are involved in apoptosis-related processes, especially those genes that were up-regulated by the GTX compounds and down-regulated by the FP-GTX compounds. This is not surprising as it is well documented that DNA damage responses involve induction of apoptosis (41). Many in vivo and in vitro studies on gene expression profiles of GTX compounds in multiple species have also found genes involved in apoptosis, albeit different from individual genes related to apoptosis as identified in our study (14,42–47). Furthermore, additional groups of genes from our study appear involved in several metabolism processes. Genes that were down-regulated by treatment with GTX compounds and up-regulated by treatment with FP-GTX compounds appeared to be involved in amino acid metabolism and biosyntheses. This is not startling because it was shown before that a disturbance in amino acids is related to apoptosis (48,49). Many similar metabolism-related genes were previously found in rat in vivo studies on gene expression modifications by genotoxic and non-genotoxic compounds, but also these were not the same genes as in our study (45,46). However, the classifiers seem to be involved in the same pathways as in our in vitro model, indicating that our classifiers are specific and reliable and thus are suitable for discrimination of true GTX and FP-GTX compounds. Not all pathways were similarly modified at 24 and 48 h of exposure, but most of these pathways were involved in the same cell processes. Thus, in general, the classifiers and pathways may not be completely overlapping between the two time points, but the overall effects are comparable. Furthermore, not all of these pathways are
directly involved in DNA damage responses, which might explain why genes correlating with γH2AX do not provide better classifiers.

Our study proves the principle that gene expression profiling in primary mouse hepatocytes is capable of classifying compounds according to their true genotoxic potential. Both treatment periods of 24 and 48 h seem to be suitable for this purpose, although at 48 h, one compound is misclassified. This confirms results previously found in in vivo and in vitro class prediction studies based on gene expression profiles. These studies were based on discriminating genotoxic from non-genotoxic carcinogens, DNA damaging from non-DNA damaging genotoxic compounds or discriminating compounds according to their carcinogenic potential (14,15,42–47). For the well-established in vitro test battery for genotoxicity, specificity appears to be extremely low. The Ames test has a reasonable specificity within 73.9%, but all mammalian cell tests show a specificity <45% (3.5). The current study can, therefore, be the basis for an additional predictive screen, which can be applied in chemical safety procedures as a novel extension of current genotoxicity testing models. Before it can be implemented, future implementation and validation steps need to focus on including more chemicals, preparing a Standard Operation Procedure and submitting to European Centre for the Validation of Alternative Methods for further test development and validation.

In summary, this study demonstrates that true GTX compounds can be reliably discriminated from false-positive GTX compounds in primary mouse hepatocytes by gene expression profiling, when using PAM. Herewith, we provide proof of the principle that toxicogenomics-based approaches in primary mouse hepatocytes may be suitable for screening compounds for their true genotoxic potential, in that possibly outperforming the standard battery of in vitro genotoxicity tests. This may result in fewer lead drugs being stopped unnecessarily and in less usage of animals for in vivo testing of genotoxicity.

Supplementary data
Supplementary data and files 1–3 are available at Mutagenesis Online.

Funding
Netherlands Genomics Initiative (050-060-510).

Acknowledgements
We would like to thank M. Tsamou for collecting the in vivo and in vitro genotoxicity data for the selected compounds.

Conflict of interest statement: None declared.

References

Discrimination of true from false-positive genotoxins