Toxicogenomics directory of chemically exposed human hepatocytes

Marianna Grinberg · Regina M. Stöber · Karolina Edlund · Eugen Rempel · Patricio Godoy · Raymond Reif · Agata Widera · Katrin Madjar · Wolfgang Schmidt-Heck · Rosemarie Marchan · Agapios Sachinidis · Dimitry Spitkovsky · Jürgen Hescheler · Helena Carmo · Marcelo D. Arbo · Bob van de Water · Steven Wink · Mathieu Vinken · Vera Rogiers · Sylvia Escher · Barry Hardy · Dragana Mitic · Glenn Myatt · Tanja Waldmann · Adil Mardinoglu · Georg Damm · Daniel Seehofer · Andreas Nüssler · Thomas S. Weiss · Axel Oberemm · Alfons Lampen · Mirjam M. Schaap · Mirjam Luijten · Harry van Steeg · Wolfgang E. Thasler · Jos C. S. Kleinjans · Rob H. Stierum · Marcel Leist · Jörg Rahnenführer · Jan G. Hengstler

Abstract A long-term goal of numerous research projects is to identify biomarkers for in vitro systems predicting toxicity in vivo. Often, transcriptomics data are used to identify candidates for further evaluation. However, a systematic directory summarizing key features of chemically influenced genes in human hepatocytes is not yet available. To bridge this gap, we used the Open TG-GATES database with Affymetrix files of cultivated human hepatocytes incubated with chemicals, further sets of gene array data with hepatocytes from human donors generated in this study, and publicly available genome-wide datasets of human liver tissue from patients with non-alcoholic steatohepatitis (NASH), cirrhosis, and hepatocellular cancer (HCC). After a curation procedure, expression data of 143 chemicals were included into a comprehensive biostatistical analysis. The results are summarized in the publicly available toxicotranscriptomics directory (http://wiki.toxbank.net/toxicogenomics-map/) which provides information for all genes whether they are up- or downregulated by chemicals and, if yes, by which compounds. The directory also informs about the following key features of chemically influenced genes: (1) *Stereotypical stress response.* When chemicals induce strong expression alterations, this usually includes a complex but highly reproducible pattern named 'stereotypical response.' On the other hand, more specific expression responses exist that are induced only

Marianna Grinberg and Regina M. Stöber shared first authorship.
Jörg Rahnenführer and Jan G. Hengstler shared senior authorship.

M. Grinberg · E. Rempel · K. Madjar · J. Rahnenführer
Department of Statistics, TU Dortmund University, Dortmund, Germany

R. M. Stöber · K. Edlund · P. Godoy · R. Reif · A. Widera · R. Marchan · J. G. Hengstler (✉)
Leibniz Research Centre for Working Environment and Human Factors at the Technical University of Dortmund (IfADo), Ardeystrasse 67, 44139 Dortmund, Germany
e-mail: hengstler@ifado.de

W. Schmidt-Heck
Leibniz Institute for Natural Product Research and Infection Biology, Hans Knoell Institute, Jena, Germany

A. Sachinidis · D. Spitkovsky · J. Hescheler
Center of Physiology and Pathophysiology, Institute of Neurophysiology, University of Cologne (UKK), Cologne, Germany

H. Carmo · M. D. Arbo
Laboratório de Toxicologia, Departamento de Ciências Biológicas, Universidade do Porto, Porto, Portugal

B. van de Water · S. Wink
Division of Toxicology, Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands

M. Vinken · V. Rogiers
Department of Toxicology, Dermato-Cosmetology and Pharmacognosy (FAFY), Center for Pharmaceutical Research (CePhaR), Vrije Universiteit Brussel (VUB), Brussels, Belgium

S. Escher
Fraunhofer Institute for Toxicology and Experimental Medicine (ITEM), Hannover, Germany

B. Hardy
Douglas Connect and OpenTox, Zeiningen, Switzerland

by individual compounds or small numbers of compounds. The directory differentiates if the gene is part of the stereotypical stress response or if it represents a more specific reaction. (2) *Liver disease-associated genes.* Approximately 20 % of the genes influenced by chemicals are up- or downregulated, also in liver disease. Liver disease genes deregulated in cirrhosis, HCC, and NASH that overlap with genes of the aforementioned stereotypical chemical stress response include CYP3A7, normally expressed in fetal liver; the phase II metabolizing enzyme SULT1C2; ALDH8A1, known to generate the ligand of RXR, one of the master regulators of gene expression in the liver; and several genes involved in normal liver functions: CPS1, PCK1, SLC2A2, CYP8B1, CYP4A11, ABCA8, and ADH4. (3) *Unstable baseline genes.* The process of isolating and the cultivation of hepatocytes was sufficient to induce some stress leading to alterations in the expression of genes, the so-called unstable baseline genes. (4) *Biological function.* Although more than 2,000 genes are transcriptionally influenced by chemicals, they can be assigned to a relatively small group of biological functions, including energy and lipid metabolism, inflammation and immune response, protein modification, endogenous and xenobiotic metabolism, cytoskeletal organization, stress response, and DNA repair. In conclusion, the introduced toxicotranscriptomics directory offers a basis for a rationale choice of candidate genes for biomarker evaluation studies and represents an easy to use source of background information on chemically influenced genes.

D. Mitic
Cambridge Cell Networks, Cambridge, UK

G. Myatt
Leadscope, Columbus, OH, USA

T. Waldmann · M. Leist
Department of Biology, University of Konstanz, Konstanz, Germany

A. Mardinoglu
Department of Chemical and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden

G. Damm · D. Seehofer
Department of General-, Visceral- and Transplantation Surgery, Charité University Medicine Berlin, Berlin, Germany

A. Nüssler
Department of Traumatology, Eberhard Karls Universität Tübingen, Tübingen, Germany

T. S. Weiss
Department of Pediatrics and Juvenile Medicine, University of Regensburg Hospital, Regensburg, Germany

A. Oberemm · A. Lampen
Federal Institute for Risk Assessment, Berlin, Germany

M. M. Schaap · M. Luijten · H. van Steeg
Center for Health Protection, National Institute for Public Health and the Environment, Bilthoven, The Netherlands

W. E. Thasler
Department of Surgery, Grosshadern Hospital, Munich, Germany

J. C. S. Kleinjans
Department of Health Risk Analysis and Toxicology, Maastricht University, Maastricht, The Netherlands

R. H. Stierum
TNO Quality of Life, Zeist, The Netherlands

Introduction

Transcriptomics is an emerging field in chemical hazard identification. The patterns of up- and downregulated genes that are obtained as readouts of chemically exposed cells and tissues provide initial evidence of the involved toxic mechanisms. For example, oxidative stress, induction or inhibition of enzymes, activation of nuclear receptors, or genotoxicity each induces a signature gene expression pattern (Godoy 2013; Hewitt et al. 2007). Gene array analysis performed on the livers of rats facilitated the differentiation between genotoxic and non-genotoxic carcinogens (Ellinger-Ziegelbauer et al. 2008; Godoy 2013). These successful toxicogenomics studies in vivo prompted several ambitious research programs, including the EU-funded SEURAT-1 network and the ESNATs project (Waldmann et al. 2014; Krug et al. 2013; Schug et al. 2013; Weng et al. 2014; Balmer et al. 2014; Campos et al. 2014), with the aim to identify biomarkers of toxicity in vitro. One long-term goal of these projects is to identify biomarkers in the in vitro systems that can predict certain mechanisms of toxicity in vivo. Moreover, the pharmacokinetic models that are developed should be able to predict the doses that result in critical compound concentrations at target cells in organisms. The proposed projects are ambitious and require time as their success necessitates that all mechanisms leading to adverse effects are known and that in vitro systems are available that reliably reflect these mechanisms. However, currently the link between gene expression alterations and adverse effects in vivo is not completely understood. Moreover, there are insufficient studies demonstrating which of the responses of in vitro systems are relevant to the in vivo situation. For example, primary cultivated hepatocytes have been shown to become apoptosis resistant in culture, thereby possibly suppressing certain in vivo

relevant responses (Godoy et al. 2009, 2010a, b). Moreover, cultivated hepatocytes are known to upregulate clusters of genes as a response to the isolation procedure and cultivation stress (Zellmer et al. 2010). This response can be suppressed by certain compounds, but consequently represents an effect that can be interpreted as a pure in vitro artifact. On the other hand, a systematic comparison between rat livers in vivo and cultivated rat hepatocytes in vitro has shown a good correlation for some cellular stress, DNA damage, and metabolism-associated genes (Heise et al. 2012). These examples illustrate that successful application of toxicogenomics requires an understanding of the role of each individual up- or down-regulated gene, and the knowledge whether a response in an in vitro system is relevant to the in vivo situation.

Primary human hepatocytes are frequently used as a model in vitro system in both pharmacology and toxicology studies. Despite its popularity, a comprehensive analysis of the genes altered by chemicals has not yet been performed. However, a large set of gene array data 'Open TG-GATEs' generated by treating cultivated primary human hepatocytes with 158 chemicals was recently made publicly available (Xing et al. 2014; Fijten et al. 2013; Zhang et al. 2014; Uehara et al. 2011; Hirode et al. 2009; Kiyosawa et al. 2009). The chemicals included hepatotoxic and non-hepatotoxic drugs, all acting via various mechanisms, as well as several experimental hepatotoxic compounds. Many of the compounds were tested using three concentrations, the highest of which was close to cytotoxic levels. Due to the variability in mechanism of actions caused by the different categories of drugs tested, it was assumed that the database encompassed a large fraction of all genes that could be altered in human hepatocytes after chemical exposure. Therefore, we analyzed the structure of the chemically induced gene expression alterations and categorized the altered genes using the following key principles. (1) *Stereotypical stress response.* Chemicals inducing strong expression alterations are usually accompanied by a complex but highly reproducible 'stereotypical' response. This stereotypical response is generally observed with numerous compounds when close to cytotoxic concentrations are used, even if the compounds act by different mechanisms. Conversely, more specific expression responses exist that are induced only by individual compounds or small numbers of compounds. (2) *Liver disease-associated genes.* Approximately 20 % of the genes influenced by chemicals are also up- or downregulated in liver disease, such as steatohepatitis, liver cirrhosis, and hepatocellular cancer. For humans, a direct validation of the influence of chemicals in hepatocytes in vivo is not possible, but can be circumvented by using sets of gene array data obtained from liver tissue of patients. An obvious course of study would be to investigate whether a gene influenced by a chemical in cultivated hepatocytes in vitro is also deregulated in human liver disease. Of course, the molecular mechanisms leading to cell

stress in liver disease and after exposure to chemicals may differ. However, if the expression of a certain gene is influenced by the microenvironment of a diseased liver, its deregulation by chemicals in vitro is less likely to represent a pure in vitro artifact. (3) *Unstable baseline genes.* Simply the process of isolating and cultivating hepatocytes has been shown to induce stress leading to expression alterations of genes, so-called unstable baseline genes. (4) *Biological function.* Although more than 2,000 genes are transcriptionally influenced by chemicals, the genes can be assigned to a relatively small group of biological functions.

This article focuses on the Open TG-GATEs dataset, several publicly available gene array data of liver diseases and gene expression data of primary human hepatocytes generated in the present study. One challenge working with large sets of transcriptomics data, especially when generated by research consortia with several independent contributors over an extended period of time, is to identify artifacts and eliminate errors. Differences that arise from the combination of several analytical batches, as well as experimental errors in subsets of samples, are almost unavoidable, but can to some extent be identified and controlled. In case of the Open TG-GATEs database, exclusion of implausible data by a number of curation steps has the potential to improve the reliability of the identified genes. Here, we present an in silico characterization and curation approach to identify and control batch effects, assess data reproducibility across replicates, and pinpoint compounds that display an implausible concentration progression. A major challenge to result interpretation is that only a relatively small subset of the 148 analyzed chemicals caused deregulation of many genes and was able to induce high fold changes, whereas the majority of compounds deregulated only small numbers of genes and caused low fold changes.

The present study begins with the generation of a curated Open TG-GATEs dataset, which represents a relatively specialized, bioinformatics-based workflow. However, the initial analysis is highly relevant since some problematic subsets of the database would otherwise compromise the analysis. Subsequent analysis introduces a toxicotranscriptomics directory where for each chemically influenced gene in hepatocytes, the basic information of 'stereotypical' versus 'specific,' 'disease gene,' 'unstable baseline gene,' and biological function is made available. We believe that the information provided by this directory will facilitate more accurate interpretation of toxicogenomics data from human hepatocytes in the future.

Materials and methods

Download and preprocessing of Open TG-GATEs data

The freely available database Open TG-GATEs (Toxicogenomics Project—Genomics-Assisted Toxicity Evaluation

System) (NIBIO 2013) compiles Affymetrix HG U133 Plus 2.0 gene expression microarray data (54,675 probe sets, corresponding to 19,945 uniquely annotated Gene Symbol IDs) from monolayer cultured primary human hepatocytes exposed to 158 compounds and corresponding untreated controls. A subset of the compounds ($n = 52$) was tested using three concentrations and three incubation periods (2 h, 8 h, and 24 h). For the additional compounds ($n = 106$), the concentration and time sets are incomplete. The compounds were tested either for only one or two exposure periods, or with only two concentrations as shown in Tables S1 and S2. If possible, the highest tested concentration (or the only tested concentration if only one was analyzed) was chosen as the concentration yielding an 80–90 % relative survival based on the LDH release, indicating concentrations where the first signs of cytotoxicity become detectable. For non-cytotoxic compounds, the high concentration was defined as the concentration of highest solubility or 10 mM at a maximum. The solvent DMSO was routinely used at 0.1 %. This concentration was increased to 0.5 % for compounds with limited solubility. In total, six batches of human hepatocytes were used, whereby hepatocytes from 'male' and 'female' were specified by the columns 'sex_type' in the 'Attribute.tsv' file, which are given together with the gene expression raw data in Open TG-GATEs. Two replicate experiments were available for 155 of the 158 compounds (Table S2); the experiments without replication were not considered in this study. The raw microarray data (CEL files) for all analyzed compounds and conditions were downloaded from the Open TG-GATEs website (http://toxico.nibio.go.jp/). For the normalization of the entire set of expression arrays, the Robust Multi-Array Average (RMA) algorithm was used that applies background correction, log2 transformation, quantile normalization, and a linear model fit to the normalized data to obtain a value for each probe set (PS) on each array (Krug et al. 2013; Harbron et al. 2007). The difference in gene expression (fold change) between treated samples and corresponding untreated controls was calculated for each compound, and for each concentration and incubation time, based on the average of replicate values. These values were used for all subsequent analyses. Data preprocessing and all subsequent analyses were performed using the statistical programming language R, version 3.0.1 (R Development Core Team 2013).

Visualization of high-dimensional gene expression data

Heatmaps were generated using unsupervised hierarchical clustering to visualize matrices of gene expression values, ranging from blue (low expression) to red (high expression). Principal component analysis (PCA) was used to visualize expression data in two dimensions, representing the first two principal components, i.e., the two orthogonal directions of the data with the highest variance. Both heatmaps and PCA were performed on the basis of the 100 top-ranking genes with highest fold change (absolute values) across all compounds. This gene selection was performed separately for all nine combinations of concentration and incubation time.

Gene set enrichment methods

Gene ontology enrichment was performed with the topGO package (Alexa and Rahnenführer 2010), using Fisher's exact test, and only results from the biological process ontology were considered. The cutoff for the enrichment p value was set to 0.001. Transcription factor binding site (TFBS) enrichment was performed using the PRIMA algorithm (http://acgt.cs.tau.ac.il/prima/) (Elkon et al. 2003) provided in the Expander software suite (version 6.04; 43 http://acgt.cs.tau.ac.il/expander/) (Ulitsky et al. 2010), and the cutoff for the enrichment p value was set to 0.01.

Definition of indices for concentration progression

For analyzing progression of gene alterations over increasing concentrations, two indices were introduced—the 'progression profile index' and the 'progression profile error indicator.' Both indices were calculated for each compound and for each pair of adjacent concentrations. The progression profile index was determined as the proportion of genes deregulated exclusively (at least twofold up or down compared to control) at a higher concentration compared to a respective lower concentration. A value close to zero indicates only few additional genes deregulated at the next higher concentration, a value close to one indicates many additional genes. The 'progression profile error indicator' interchanges the roles of the lower and higher concentration and specifies the proportion of genes deregulated exclusively at a lower compared to a respective higher concentration. Values above 0.5 indicate an implausible concentration progression. However, if only a few genes in total are altered at the respective lower concentration, they can be interpreted as outliers. Thus, the 'modified progression profile error indicator' is an adjustment of the 'progression profile error indicator' for such cases, setting the index to zero, if the original value is larger than 0.5 and at most 20 genes are altered at the respective lower concentration. All three progression indices were calculated separately for the three exposure periods of 2 h, 8 h, and 24 h.

Principle for differentiating between stereotypic and compound-specific gene expression responses

The selection value introduced in this study is a method to differentiate between stereotypic and compound-specific

gene expression responses. Genes that are deregulated by many compounds reflect a stereotypic response, in contrast to compound-specific response of genes that are regulated only by a few compounds. For a given probe set, the selection value determines the number of compounds that induces an expression change of at least threefold. Selection value 20 (SV20) then yields a list of genes deregulated at least threefold by at least 20 compounds, whereas SV3 gives a list of genes deregulated threefold by at least three compounds. For a specific concentration and incubation time, compounds are ranked in order of fold change for each probe set, i.e., for the upregulated probe sets, compounds are ranked from high to low fold change and for the downregulated probe sets in reverse order. Thus, the selection value x (short SVx) for a single probe set is the compound with rank x, meaning that the probe set is deregulated threefold in at least x compounds. The threshold of threefold for the selection value method and for other analyses in this study is chosen arbitrarily. In principle, also, e.g., a 2- or 1.5-fold threshold would be possible. The rationale for choosing a threefold threshold is to keep the number of false-positive genes relatively low.

Primary human hepatocyte isolation and cultivation

Liver tissue from eight patients undergoing surgical liver resection was used to isolate primary human hepatocytes. Patients' characteristics and the assignment of the individual donors to the respective experiments are summarized in Table S3. The cells were used for the time-dependent cultivation experiments to determine the 'unstable baseline genes' and for quantitative real-time PCR (qRT-PCR) experiments. The experiments were approved by the local ethical committees, and all patients provided their written consent. The donors' cells were obtained by a two-step isolation procedure developed by Seglen (1976), and processed and cultivated as recently described by Godoy (2013). Briefly, during the first perfusion step, the liver was rinsed with an EGTA-containing buffer to prevent coagulation and to remove residual blood and calcium from the vessels. Secondly, the liver was perfused with a collagenase-containing buffer, which gradually digested the liver tissue until the cells could be easily released from the liver capsule into a suspension buffer. The cells were transported overnight from the surgical department as cold stored suspensions on ice. Upon arrival, the cells were resuspended in fresh cultivation medium (William's E including 2 mM stable glutamine, 100 U/mL penicillin, 0.1 mg/mL streptomycin, 10 μg/mL gentamicin, 100 nM dexamethasone, 2 ng/mL insulin plus 10 % fetal calf serum (FCS) 'Sera Plus' during the first 3–4 h of cultivation), and the viability was determined using the trypan blue exclusion method. Cells were seeded in conventional six-well plates between two soft gel layers of collagen, using 350 μL of 1 mg/mL collagen gel per layer per well. 10^6 cells/well were seeded in 2 mL FCS containing cultivation medium and kept in the incubator for at least 3 h to allow attachment of cells to the collagen matrix before the second layer of collagen was applied. Upon polymerization, cells were incubated in FCS-free cultivation medium.

RNA extraction

To extract RNA from cultivated primary hepatocytes, the cultivation medium was aspirated and 1 mL of QIAzol (Qiagen, Hilden, Germany) was added immediately. Samples were sonicated for 30 s (alternating 5 s pulse, 2 s pause) on ice, and further processing was performed according to the manufacturer's instructions.

Gene expression microarray analysis for the analysis of 'unstable baseline genes'

RNA extracted from freshly isolated primary human hepatocytes (FH) and primary human hepatocytes cultivated in collagen sandwich (CS) for 1, 2, 3, 5, 7, 10, and 14 days was analyzed on Affymetrix HG U133 Plus 2.0 arrays in triplicates (hepatocytes isolated from three donors on three different occasions). Two of the samples were excluded from the array analysis due to poor RNA quality (RIN value <8), as assessed by the Agilent 2100 Bioanalyzer system and the RNA 6000 Nano LabChip Kit (Agilent Biotechnologies, Palo Alto, USA). Microarray data preprocessing and normalization was performed as described above for the Open TG-GATEs dataset. Replicates were averaged, and for each time point (CS), the fold change was calculated as compared to FH, and up- and downregulation were defined as at least a threefold difference.

Quantitative real-time PCR

Primary human hepatocytes from five donors were cultivated overnight and treated for 24 h with valproic acid, ketoconazole, galactosamine, acetaminophen, and isoniazide (all purchased from Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany) at concentrations corresponding to the highest reported for the respective compound in the Open TG-GATES database, followed by extraction of RNA as described above. The number of donors tested per compound is specified in Table S4. The High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Darmstadt, Germany) was used to reversely transcribe RNA into cDNA. Quantitative real-time PCR (qRT-PCR) with TaqMan probes was performed on the ABI 7500 Fast Real-Time PCR system (Applied Biosystems) to determine gene expression levels.

Glyceraldehyde-3-phosphate-dehydrogenase (GAPDH) was used as the endogenous control. The applied assays were 4352934E (GAPDH), Hs00166169_m1 (G6PD), Hs01037712_m1 (PDK4), Hs01650979_m1 (INSIG1), and Hs01572978_g1 (PCK1) (Applied Biosystems). Furthermore, one additional gene, THRSP (Hs00930058_m1), was analyzed after 24 h treatment at all three concentrations reported in the Open TG-GATES database for three compounds (valproic acid, ketoconazole, isoniazide). Twenty-five nanograms cDNA was used in each PCR, and the conditions were set according to the standard specifications recommended by Applied Biosystems. For calculation of relative gene expression, the $\Delta\Delta CT$ method was used. Time-matched, untreated cells cultivated on the same six-well plate as those that were treated were used as controls. Between two and five biological replicates (hepatocytes isolated from different donors) were analyzed for each gene, depending on the final amount of RNA available.

Liver disease dataset analysis

Microarray datasets that investigated global gene expression changes in liver disease were retrieved from public data repositories ArrayExpress (E-MEXP-3291) and Gene Expression Omnibus (GSE25097). E-MEXP-3291 (Lake et al. 2011) was analyzed on Affymetrix GeneChip Human 1.0 ST arrays and was used to compare non-alcoholic steatohepatitis (NASH) ($n = 16$) to healthy liver tissue ($n = 19$). GSE25097 (Tung et al. 2011) was analyzed on Human RSTA Affymetrix 1.0 Custom CDF microarrays and was used to compare cirrhotic liver ($n = 40$) to non-tumor liver tissue ($n = 243$). Moreover, we retrieved normalized RNA sequencing (RNA-Seq) data, analyzed on the Illumina HiSeq platform, from The Cancer Genome Atlas (TCGA) (http://cancergenome.nih.gov/) to study gene expression changes in hepatocellular carcinoma (HCC) ($n = 163$) as compared to matched non-tumor liver tissue ($n = 49$). Microarray gene expression data was processed and quantile normalized using the Piano R package (Väremo et al. 2013). Differential expression analysis was also carried out using the Piano package, and p values were corrected for multiple testing by the method of Benjamini and Hochberg (Benjamini and Hochberg 1995). Differential expression analysis of RNA-Seq data was performed using the R package DESeq (Anders and Huber 2010). Genes with a fold change of at least 1.3 and a false discovery rate (FDR) adjusted p value ≤ 0.05 in the pairwise comparison of healthy/non-tumor tissue to diseased tissue were considered differentially expressed. To enable a direct comparison of differentially expressed genes to genes deregulated after chemical exposure to human hepatocytes in vitro, which were originally analyzed on different platforms, probe sets included on the Affymetrix arrays were converted

into uniquely annotated Ensembl Gene IDs. This resulted in 18,809 genes for the Open TG-GATEs dataset (originally 54,675 probe sets), 19,477 genes for E-MEXP-3291 (originally 32,321 probe sets), and 25,426 genes for GSE25097 (originally 37,582 probe sets). In a next step, only genes contained in both the Open TG-GATEs dataset and E-MEXP-3291 (17,663 genes) or GSE25097 (16,514 genes), respectively, were included in the final comparison. To enable a direct comparison to the TCGA dataset (20,471 genes, as recognized by a unique Entrez Gene ID), the Affymetrix probe sets that were included in the analysis of the Open TG-GATEs dataset were converted to Entrez IDs using manufacturer mapping after duplicate removal, resulting in 19,944 uniquely annotated genes and 17,895 genes included in the final comparison.

Results

Data structure and curation

Database description and data subset selection

The Open TG-GATEs database (NIBIO 2013) compiles global gene expression data from cultured primary human hepatocytes exposed to 158 compounds and corresponding untreated controls. Fifty-two of the 158 compounds were tested using three concentrations ('high,' 'middle,' and 'low') with the highest concentration approximately representing the EC10, and with a dilution factor in the range 2–7 (factor 5 in more than 80 % of cases). Three incubation times (2 h, 8 h, and 24 h) were investigated, and each condition was assessed using two replicate experiments. Only a subset of the conditions tested for the 52 compounds was analyzed for the additional 106 compounds (Tables S1 and S2). Raw expression data were available for 2,605 array experiments (1,879 exposed and 726 control samples), indicating incomplete concentration sets, time sets, and/or replicate sample pairs for a proportion of the analyzed compounds. For three compounds, bromoethylamine (BEA), lipopolysaccharide (LPS), and trimethadione (TMD), only one replicate experiment was available investigating the highest tested concentration and the 24-h time point. For phorone (PHO), no data at all were available for that test condition. With respect to that test condition, these compounds were therefore excluded from further analyses. A summary of the available data for the different compounds, including concentrations, time points, and replicates, is presented in Table S1, and a detailed compound-specific summary is provided in Table S2. Seven of the tested compounds were cytokines (interferon alpha, hepatocyte growth factor, interleukin 1 beta, interleukin 6, transforming growth factor beta 1, tumor necrosis factor) and LPS, which were all excluded from further analyses resulting

Fig. 1 Principle component analysis of gene expression data ▶ obtained from human hepatocytes after incubation with 148 chemicals (*green symbols*) and 7 cytokines (*red symbols*). Data of the high concentration and 24-h incubation are shown. All other incubation conditions are summarized in Fig. S1. **a** Overview of all samples and replicates. The *dark* and *light green symbols* illustrate the controls and exposed samples, respectively. **b** Connecting lines between replicates illustrate the degree of variability. **c** Mean values of the replicates. **d** Connecting lines between controls (*dark green*) and corresponding compound exposed (*light green*) samples. **e** Subtraction of the controls from the corresponding compound exposed samples (color figure online)

in a total of 151 small molecular chemical compounds that were included in the subsequent analyses.

Batch effects

To obtain an overview of gene expression alterations induced by the tested compounds, PCA was performed using the 100 top-ranking genes with the highest fold change (absolute values) across all compounds, where the nine combinations of concentration and incubation time were independently considered (Fig. S1). This is best demonstrated by the 24-h high concentration subset (Fig. 1a) which illustrates the location of controls within two main clusters—the lower cluster subdivided into several sub-clusters and the majority of the treated samples that move in the direction of the first principal component. To visualize the contribution of technical variability to the observed pattern, all replicates were connected by lines. The resulting pattern, upon visual inspection, indicated that the majority of the replicate pairs were found located close to each other (Fig. 1b). This prompted us to continue our analyses using the mean values of the replicates (Fig. 1c). Connecting lines drawn between controls and the corresponding treated samples illustrate that individual control–treatment pairs were located within only one of the two main clusters, for the incubation condition of 24 h and the high concentration (Fig. 1d), suggesting that the difference between the clusters is a consequence of experimental variability. Therefore, we subtracted the controls from the corresponding compound exposed samples. In the resulting scatter plot (Fig. 1e), the above described clusters are no longer observed, suggesting that batch effects were removed by this procedure. In conclusion, it appears that no special procedures are needed to correct batch effects within the data, as simply subtracting the corresponding controls seems to be sufficient.

Reproducibility

As a next step, the reproducibility between replicates was quantitatively analyzed. For this purpose, the distributions of the Euclidean distances between all replicate sample
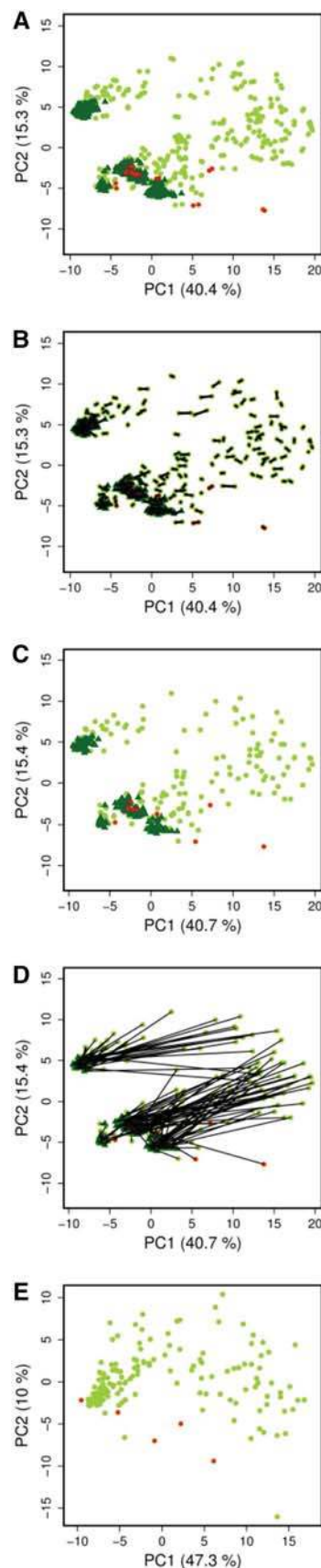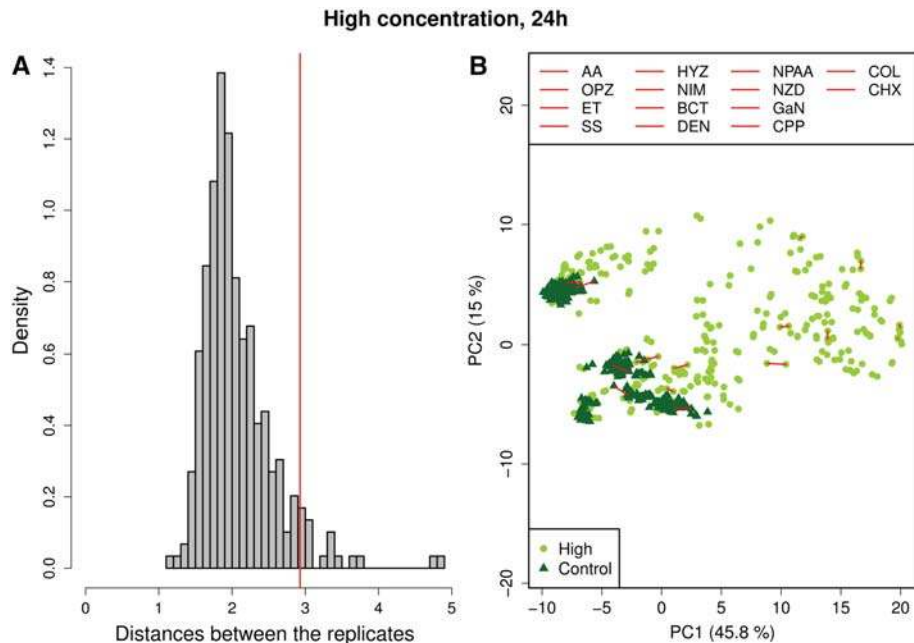
pairs and between control–treatment sample pairs were compared. The median distance between replicates was 4.9-fold lower than the median distance between control–treatment pairs for the 24-h, high concentration samples. The Euclidean distances of the replicate pairs are shown in a frequency distribution (Fig. 2a). The red line in the histogram in Fig. 2a separates the 5 % largest observed distances from the main distribution, representing 14 (9.5 %) of the compounds tested in the 24-h, high concentration subset. These replicate pairs with the largest observed distances were illustrated by connecting lines in a PCA plot (Fig. 2b). The result suggests that even the worst replicates had a relatively small variability in relation to the much larger compound-induced effects. The same principle applies to the other time–concentration subsets. Therefore, reproducibility between replicates is within an acceptable range.

Number of deregulated genes per compound

A situation where a relatively small subset of tested compounds is responsible for the majority of gene expression effects, i.e., where the majority of compounds causes no or only small expression alterations, warrants further investigation in order to exclude false-negative findings, especially when the test compounds were selected based on the probable or potential toxic effects. In the Open TG-GATEs dataset, it is surprising that well-documented hepatotoxic compounds, such as carbon tetrachloride, are among the tested compounds which show only a very weak effect on gene expression. In Fig. 3, the time- and concentration-dependent increase in the number of upregulated genes

per compound is shown for fold changes of 1.5, 2.0, and 3.0, indicating substantial differences between compounds. For example, at 24-h, high concentration, cycloheximide induced the largest number of gene expression alterations with 5,124; 2,547, and 887 upregulated probe sets (Fig. 3) and 5,506; 2,621; and 903 downregulated probe sets (Fig. S2) for fold changes 1.5, 2.0, and 3.0, respectively. Under the same conditions, triazolam deregulated only 37 genes at least 1.5-fold (6 up, 31 down) and only one gene at least twofold (down). Table S5 gives an overview of the compounds that deregulated (twofold up or down compared to control) at most 20 genes when administered at the low, middle, and high concentrations for each incubation period and for both up- and downregulated genes. Eleven of the 48 compounds that were tested at all concentrations for all time periods appear in all six lists, i.e., they deregulated less than 20 genes in total, independently from direction (up or down) and time point. These were clofibrate, hexachlorobenzene, phenytoin, coumarin, gemfibrozil, bromobenzene, amiodarone, sulfasalazine, cimetidine, haloperidol, and glibenclamide.

Determining the 100 strongest deregulated genes across all compounds and assigning these genes to the compound with the most extreme fold change showed that only 32 and 23 of the 148 analyzed compounds were responsible for the 100 most up- and downregulated genes, respectively (Fig. 4; Fig. S3). This is a situation that requires careful consideration because it may indicate that a relatively large fraction of the studied compounds causes only weak expression alterations. Alternatively, it could mean that the highest concentration tested, although intended to be close to cytotoxic, was not high enough. In previous gene array
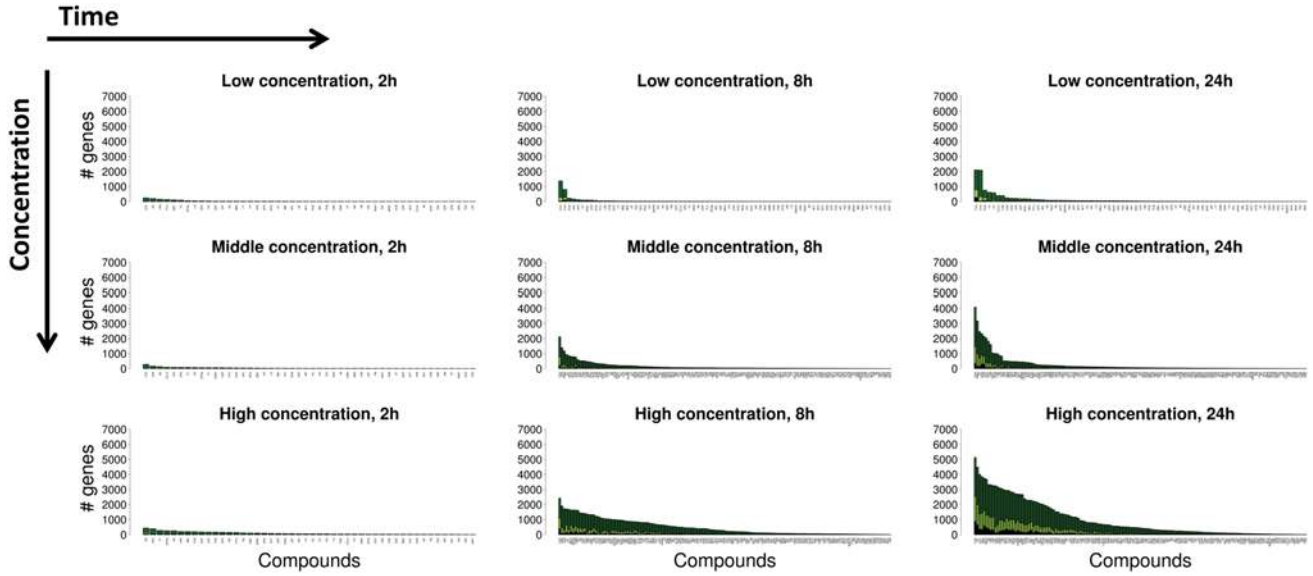
Fig. 3 Number of significantly upregulated genes. The *x* axis lists all chemicals that were tested at the indicated concentration for the corresponding period. The *y* axis gives the number of upregulated genes with at least 1.5-, 2.0-, and 3.0-fold change. The result shows that the number of deregulated genes differs strongly between the chemicals.

The corresponding data for downregulated genes is shown in Fig. S2. *Dark green* more than 1.5-fold upregulated; *light green* more than twofold upregulated; *black* more than threefold upregulated (color figure online)
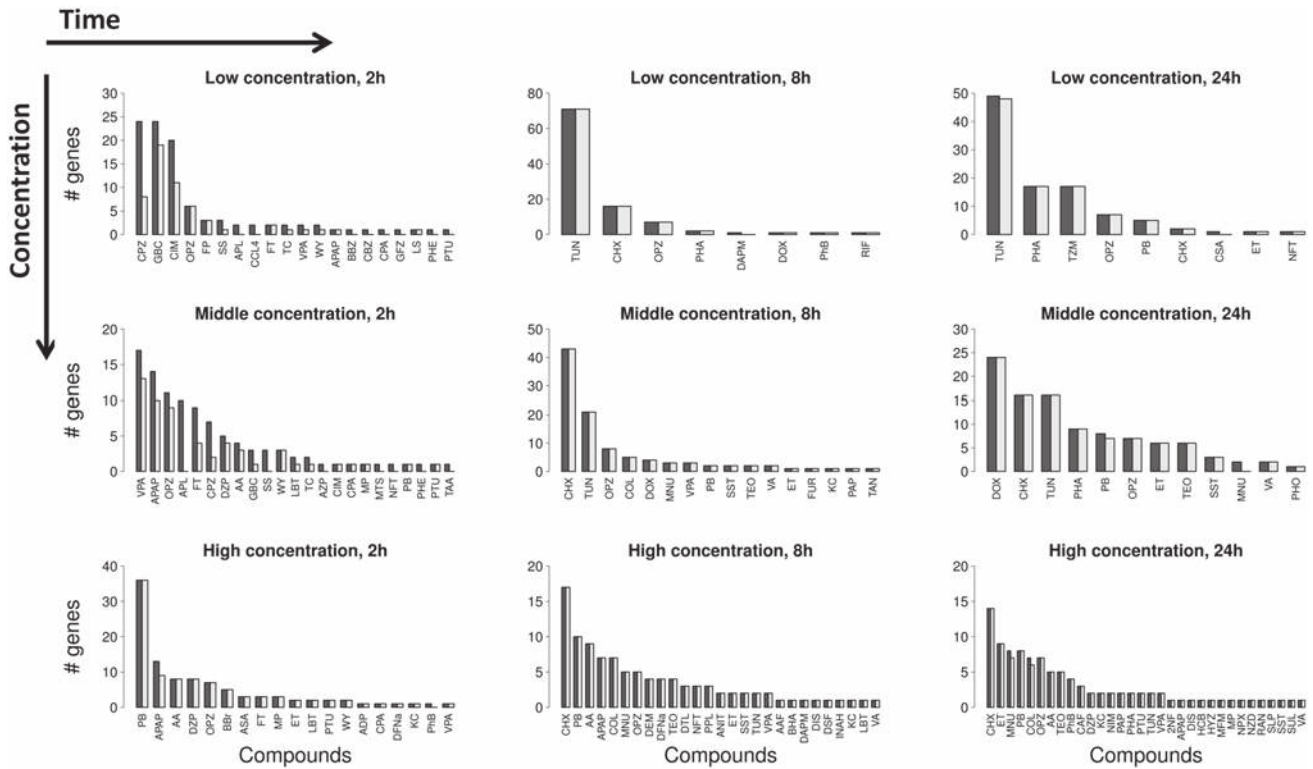

Fig. 4 'Exclusivity analysis' of the upregulated genes. This analysis first determines the 100 strongest upregulated genes across all compounds. Next, these genes are assigned to the compound with the

most extreme fold change. The corresponding analysis for the downregulated genes is shown in Fig. S3

studies, we have seen that identification of 'close to cytotoxic' concentrations is challenging, especially for compounds with steep concentration effect curves (Krug et al. 2013; Waldmann et al. 2014). Also the method used to determine cytotoxicity is of importance. In addition, a lack of differential genes could simply be due to experimental errors. One possibility to identify the latter is to analyze the deviations from monotonous concentration progression, as described in the following paragraph.
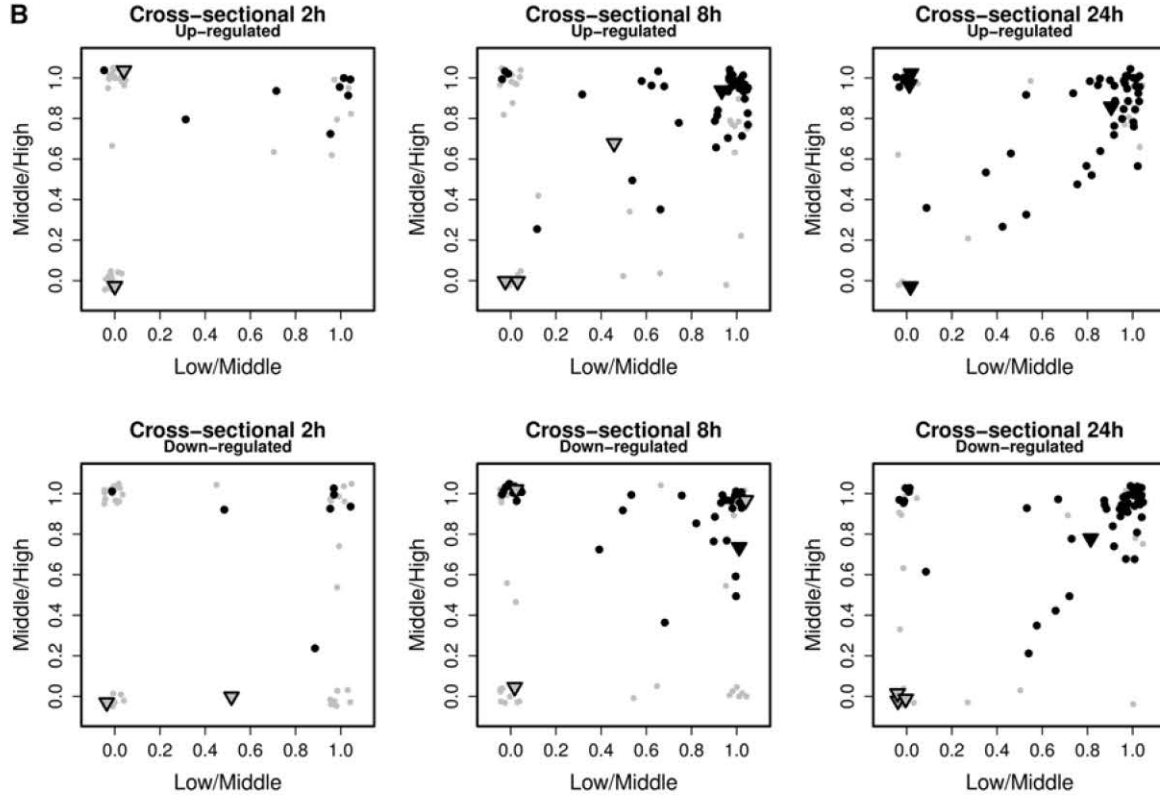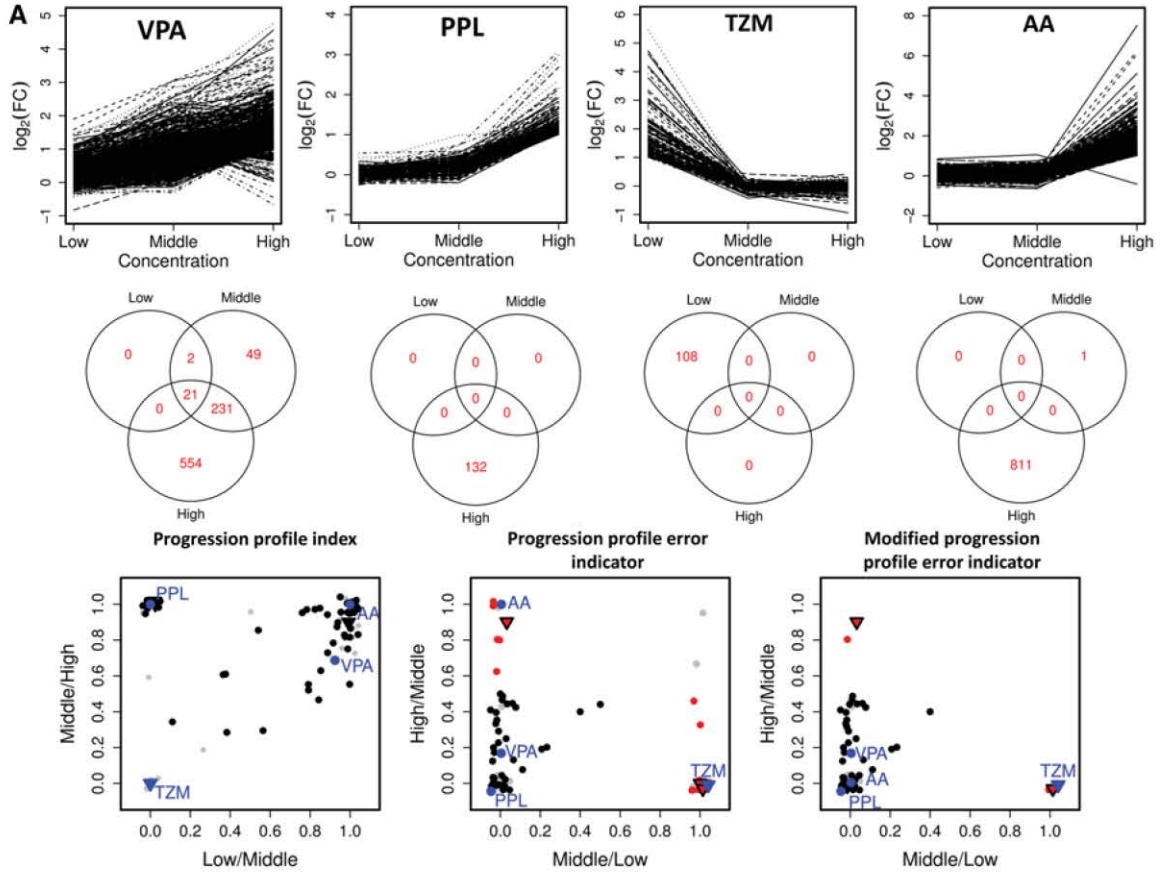
Concentration progression

In concentration-dependent gene expression studies, genes that are up- or downregulated after treatment with low concentrations of a particular compound are usually also deregulated at higher concentrations. When a compound deregulates a large number of genes at a low concentration, but gene expression remains unaltered at a higher concentration, this may be indicative of low-quality data and may be a consequence of experimental errors. Of course a non-monotonous dose–response relationship cannot automatically be interpreted as a consequence of low data quality. However, it may help to identify subsets of data that require more detailed evaluation. To analyze concentration progression across the database, two different types of analyses were performed. First, the 'progression profile index' was introduced to describe at which concentration—low, middle, or high—deregulation of genes occurs. Second, the 'progression profile error indicator' was calculated to identify compounds with an unusual concentration progression of gene expression alterations, i.e., an expression profile with a large fraction of genes deregulated at a lower but not at higher concentrations.

The 'progression profile index' was determined as the fraction of genes that were at least twofold up- or downregulated at a higher concentration (middle vs. low and high vs. middle), but unchanged at the respective lower concentration. A 'progression profile index' value close to zero thus indicates that only few additional genes are deregulated at a higher concentration, whereas a value close to one points toward a large fraction of genes exclusively deregulated at the higher concentration. The concept is illustrated for four examples in Fig. 5a (only for upregulated genes). The upper panel of Fig. 5a illustrates the expression of the individual genes at the three concentrations. The panel in the middle gives the corresponding Venn diagrams that count the overlaps between genes upregulated at least twofold. The middle, but not the low concentration of valproic acid (VPA) resulted in a high fraction of upregulated genes, which therefore cluster to the right side of the 'progression profile index' plot (lowest panel in Fig. 5a). Moreover, a relatively high fraction of additional genes are upregulated with the high compared to the middle concentration, giving

Fig. 5 Concentration progression analysis. **a** Principles of the 'progression profile index' and the 'progression profile error indicator' illustrated for the four compounds valproic acid (VPA), propranolol (PPL), triazolam (TZM), and allyl alcohol (AA). Only the upregulated genes after 24-h exposure were considered in this example. The *upper panel* shows the levels of the individual upregulated genes at three concentrations (low, middle, high). The *panel* in the *middle* summarizes the upregulated genes by Venn diagrams. The *lowest panel* shows the resulting positions of the four compounds in the respective profilers indicated in *blue*. **b** Overview of the 'progression profile indices' for all compounds tested at three concentrations (low, middle, high) after three exposure periods (2 h, 8 h, and 24 h). First, the genes up- or downregulated at a higher concentration were determined (twofold up or down compared to control). Next, the fraction of these genes that are not deregulated at the lower concentration was calculated. These calculations were performed comparing the low versus the middle (*x* axis) and the middle versus the high (*y* axis) concentrations. A value close to zero means that only few additional genes were deregulated at a higher concentration; whereas, a value close to one indicates a large fraction of genes deregulated only at the higher concentration. *Each symbol* represents an individual compound. The *triangles* represent the later excluded compounds. *Black* or *gray symbols* indicate that more than or less than (or equal to) 20 genes, respectively, were deregulated in total (color figure online)

VPA a relatively high position on the vertical axis of the plot. Propranolol (PPL), which upregulates genes only at the high concentration, is positioned in the upper left of the 'progression profile index' plot. Only low concentrations of triazolam (TZM) and high concentrations of allyl alcohol (AA) resulted in the upregulation of genes that clustered to the lower left and the upper right corner of the plot, respectively. Figure 5b illustrates the 'progression profile indices' for each of the 151 compounds for the comparison of the middle versus low (*x* axis) and the high versus middle (*y* axis) concentration. For the 24-h time period, most compounds clustered in the upper right corner of the diagram (Fig. 5b). This reflects the situation where with each concentration step, from low to middle and from middle to high, additional genes become deregulated. The second most abundant compounds cluster to the upper left corner. For these compounds, genes were only deregulated with the high compared to the middle concentration, but not with the middle compared to the low concentration. Compounds that clustered in the lower left corner deregulated additional genes that were not observed for the middle or high concentrations, meaning that the deregulated genes were already up or down with the low concentration. The gray color used in the 'progression profile indices' represents the compounds that deregulated less than 20 genes and is indicative of the 'weak compounds' summarized in Table S5. A relatively high fraction of these compounds clustered to the lower left, suggesting a concentration progression profile with deregulated genes that starts with the low concentration with only a few further genes up- or downregulated with the middle and high concentrations. The observation that this particular concentration progression pattern

A VPA, PPL, TZM, AA concentration-response profiles with Venn diagrams for Low, Middle, High concentrations, and scatter plots for Progression profile index, Progression profile error indicator, and Modified progression profile error indicator.

B Cross-sectional 2h, 8h, 24h Up-regulated and Down-regulated scatter plots.
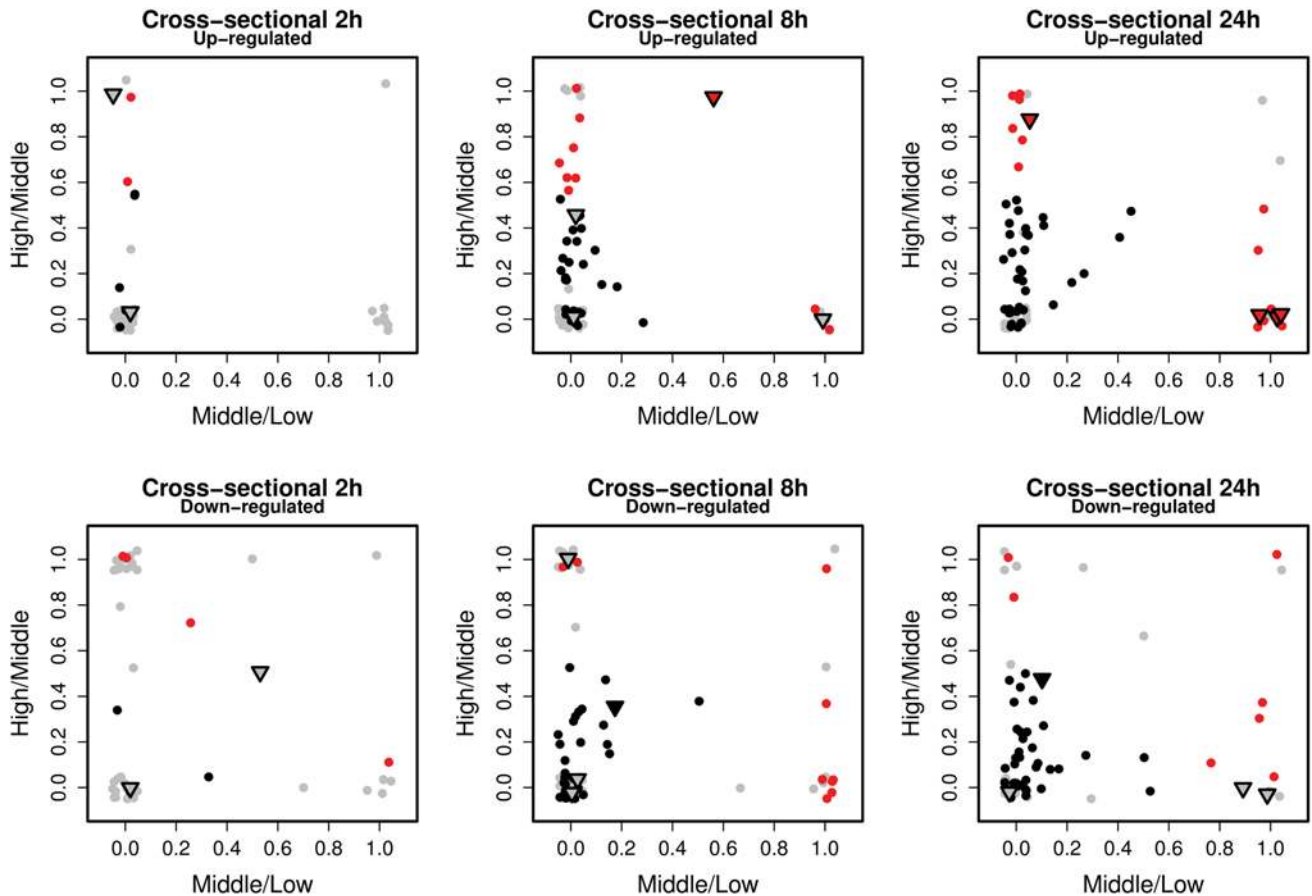
Fig. 6 'Progression profile error indicator.' A high value means that a high fraction of the genes were deregulated exclusively at a lower compared to a respective higher concentration. Each *symbol* represents an individual compound. The *triangles* present the excluded compounds. *Gray symbols* indicate that less than or equal to 20 genes were deregulated in total. *Black symbols* indicate that more than 20 genes were deregulated in total, and both values are smaller than or equal to 0.5. *Red symbols* indicate that more than 20 genes were deregulated in total and that at least one of the error indicator values is greater than 0.5 (color figure online)

was preferentially observed for the 'weak compounds' led us to test their reliability using a 'progression profile error indicator.'

The 'progression profile error indicator' was designed to identify compounds that cause a non-monotonous concentration progression of gene expression alterations. It recognizes genes that are deregulated when treated with the lower but not the respective higher concentration (Fig. 6). The principle is illustrated with four compounds in Fig. 5a. One hundred and eight genes were upregulated using a low-concentration TZM, but none when the middle concentration was used. Therefore, TZM yields a high value on the 'middle to low' axis of the 'progression profile error indicator' (Fig. 5a). This indicator seems useful, because a compound with an ambiguous concentration progression such as TZM requires careful reevaluation. On the other hand, VPA and PPL show predominantly increasing concentration progressions, and therefore, the low 'progression profile error indicators' are plausible. The results obtained with AA are more complicated as a high error indicator on

the 'middle to high' axis is observed (Fig. 5a). However, the corresponding Venn diagram shows that the high 'error indicator' was caused by a single gene that was upregulated with the middle, but not the high concentration. This example illustrates that a certain number of non-monotonous genes should be exceeded before they contribute to the error indicator. In the present study, this number was set to 20, which finally results in the 'modified progression profile error indicator' (Fig. 5a, lowest panel, right side). The results obtained for AA and other compounds demonstrate the usefulness of this error indicator. Therefore, the 'modified progression profile error indicator' was applied to evaluate all 151 compounds. It considered the two error indicators of the comparison of the low versus middle and middle versus high concentration, as well as the number of deregulated genes (twofold up or down compared to control).

On the basis of the modified progression profile error indicator, a 'progression error profile' was defined. Therefore, each compound was assigned to labels annotating the concentration progression for each time period (with respect

to the up- and downregulated genes). Hence, the 'progression error profile' comprises a set of six labels 'NA,' 'OO,' 'o,' '+,' '−' indicating 'the compound was not tested,' 'the number of differentially expressed genes at all concentrations was zero,' 'the number of differentially expressed genes at all concentrations was at most 20,' 'the number of differentially expressed genes was at least 20 and both error indicator values were below (or equal to) 0.5' or 'the number of differentially expressed genes was at least 20 and one of the two error indicator values was above 0.5' for that time period. Combining the annotations for the compound-wise concentration progression at all time points, each compound is provided with a labeling of the following design: '2 h Up| 8 h Up| 24 h Up| 2 h Down| 8 h Down| 24 h Down'. For a more detailed description, the reader is referred to Table S6. The profiles can be used to assign the compounds to groups which are characterized by their concentration progression. All in all, 63 different profiles can be observed (Table S6). Most of the compounds, 35 in total, had the profile NA|+|+|NA|+|+, meaning that these compounds show a plausible concentration progression for the incubation periods 8 h and 24 h. No data were available for the 2-h incubation period. The following compounds were excluded from further analyses of the curated database due to their 'progression error profile': carbon tetrachloride ('o|o|+|o|o|o'), doxorubicin ('NA|−|−|NA|+|+'), triazolam ('NA|o|−|NA|OO|OO'), tetracycline ('o|o|−|OO|o|o'), ticlopidine ('NA|o|o|NA|o|−'). Except for doxorubicin, all compounds belong to the category 'weak compounds.' In addition, three further compounds (aspirin, indomethacin, and methyltestosterone) showed the same profile as carbon tetrachloride, even though they are known to be less directly hepatotoxic. Although the results from all three compounds should be treated with caution, they were not removed, because without further experiments it cannot be excluded that the observed 'weak' result and unusual progression profile are real. Table S7 gives an overview of the compounds that deregulated more than 20 genes and yielded a 'progression profile error indicator' value above 0.5. The top 32 (23) compounds which contribute to the 100 most up (down)-regulated genes for the 24-h, high concentration subset (Fig. 4, Fig. S3) yielded on average, error indicator values in the range of 0.1–0.4 across all time periods. The highest *error indicator* values were attained by compounds that belong to the 'weak compounds,' which deregulated less than 20 genes in total (Table S5). They clustered to the upper right of the 'progression profile error indices' (Fig. 6). In conclusion, unusual concentration progression profiles were mainly identified among the compounds with weak expression responses. Therefore, the following part of the manuscript focuses on the compounds with strong expression responses. The RMA normalized, in silico curated version of the Open TG-GATEs database, where compounds with an unusual concentration progression have been removed, is available under http://wiki.toxbank. net/toxicogenomics-map/. These data represent the basis for all further calculations.

In vitro reproduction of reported compound-gene effects

To assess the reproducibility of the gene expression response to chemical exposure reported in the Open TG-GATEs database, primary human hepatocytes from five donors were isolated and cultivated, as described above. The in vitro response to five compounds (valproic acid, ketoconazole, galactosamine, acetaminophen, isoniazide) under identical conditions, as described in the Open TG-GATEs database (highest concentration and 24-h incubation), was determined by qRT-PCR for four selected genes (G6PD, PDK4, PCK1, INSIG1) (Table S4). The in vitro response to three of the five compounds (valproic acid, ketoconazole, isoniazide) was additionally analyzed for thyroid hormone-inducible hepatic protein (THRSP), using a wide concentration range for the 24-h treatment time point (Table S8). We tested whether a qualitative agreement could be obtained between the data in the Open TG-GATEs database and our qRT-PCR analyses. The strong induction of G6PD by valproic acid, ketoconazole, and acetaminophen (23.7-, 3.4-, and 7.3-fold, respectively) reported by Open TG-GATEs was qualitatively confirmed in our data (Table S4). However, the relatively weak induction of G6PD (2.4-fold) by isoniazide was not confirmed, and galactosamine resulted in large differences between the tested donors. Induction of PDK4 (4.2-fold) by valproic acid in the Open TG-GATEs data was confirmed, but the 3.4-fold increase by acetaminophen was not reproduced (Table S4). For this limited set of genes, the results for hepatocytes from independent donors illustrate that more than half of the positive observations from the database can be reproduced, but independent confirmation will be required in the future to obtain even higher reliability. THRSP should be highlighted as a special case as the gene array data could not be confirmed. THRSP expression was upregulated by many compounds, including VPA in Open TG-GATEs, and has previously been reported to play a role in the pathogenesis of liver steatosis (Wu et al. 2013). However, induction of THRSP in cultivated human hepatocytes could not be confirmed after incubation with valproic acid, ketoconazole, and isoniazide (Table S8).

Warning flags for unstable baseline genes

It is well known that isolation and cultivation of primary hepatocytes causes up- and downregulation of numerous genes, named 'unstable baseline genes' (Godoy 2013; Zellmer et al. 2010). A relatively high number of genes

Fig. 7 Unsupervised clustering of the 100 most deregulated genes across all compounds tested at the highest concentration for 24 h of incubation. The *lines* represent the compounds, while each column stands for a gene. *Red color* indicates up and *blue color* downregulated genes as indicated by the code in the *upper left*. Moreover, the compounds have been classified with respect to their genotoxicity, human hepatotoxicity, and BSEP inhibiting capacity. These properties are indicated in the *columns left* of the heatmap. Unsupervised clustering results in three clusters that can be associated with biological motifs, proliferation, cytochrome P450 (CYP), and stress response as indicated below the heatmap (color figure online)
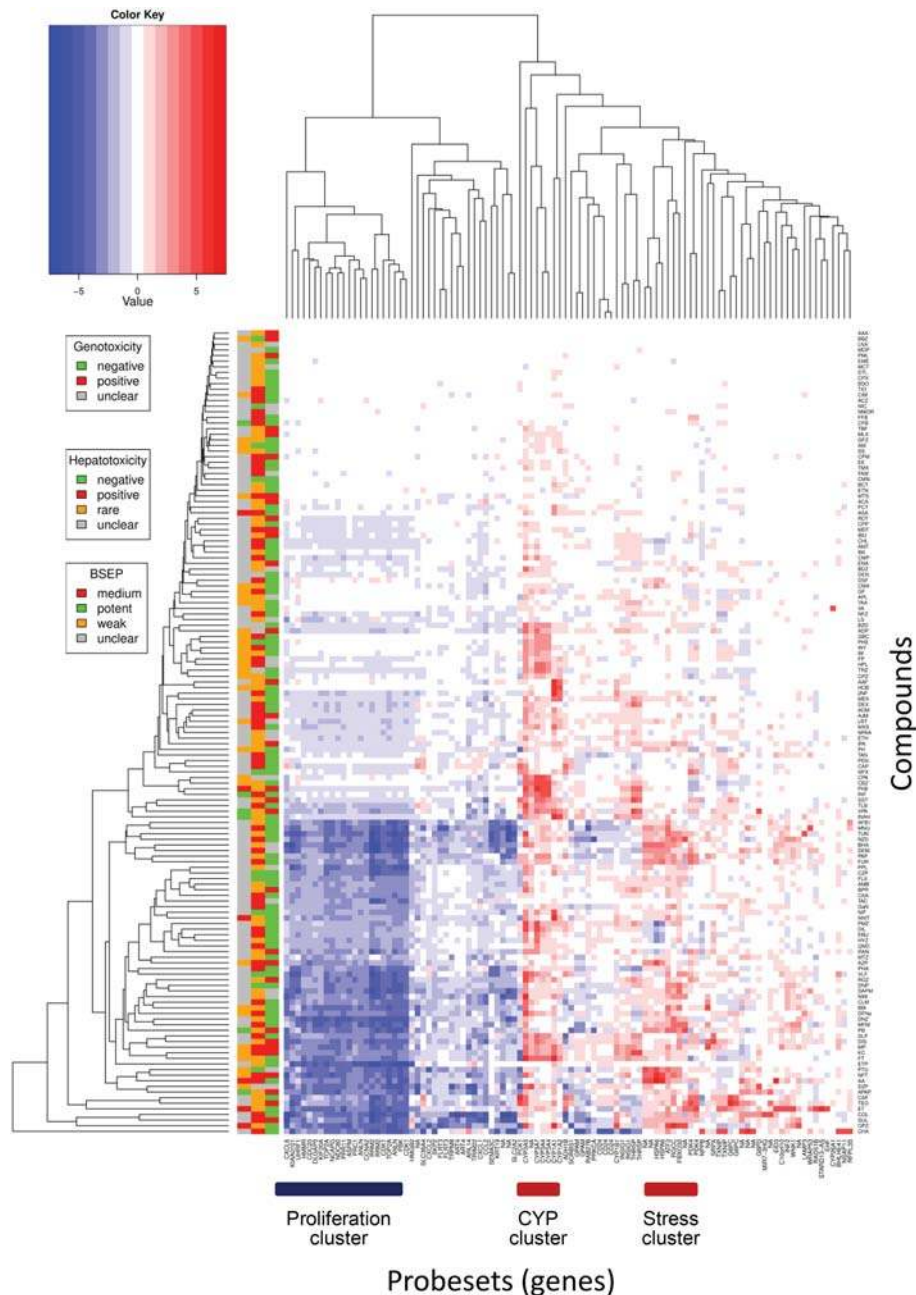
associated with xenobiotic and endogenous metabolism are known to be downregulated during cultivation. In contrast, inflammation-associated genes have been reported to be induced. These gene expression alterations occur as a response to the isolation stress and to the culture conditions. They are independent from compound exposure and might thus give rise to false-positive findings or cloud true findings as a consequence of opposing effects. To identify unstable baseline genes, expression profiles in freshly isolated primary human hepatocytes (FH) were analyzed by gene arrays and compared to the expression of corresponding genes after 1, 2, 3, 5, 7, 10, and 14 days in collagen sandwich (CS) culture. The number of probe sets upregulated at least threefold after one or more time points, as compared to FH, was 1,509 (1,086 genes); for the at least threefold downregulated probe sets, the corresponding number was 1,754 (988 genes) (Table S9). Categorization as unstable baseline genes does not necessarily exclude biological relevance. It is possible that an identical set of genes are influenced by certain chemicals and by cell stress induced in response to hepatocyte isolation and cultivation. However, an unstable baseline may render the identification of the effects of chemicals technically more difficult, because they have to be differentiated from a second factor of influence—isolation and cultivation stress. Therefore, in the toxicotranscriptomics directory, the column CS indicates whether the

gene is up- or downregulated during collagen sandwich culture of primary hepatocytes (Table S13).

Detection of biological motifs

To characterize the genes that were deregulated by the tested compounds at the highest concentration after an incubation period of 24 h, unsupervised cluster analysis was performed. Only the 143 compounds that remained after the aforementioned data curation step (i.e., removal of compounds with an unusual concentration progression profile) were included to generate a heatmap of the 100 strongest deregulated genes (horizontal) across the remaining compounds (vertical) (Fig. 7; Table S10, Table S19). A pattern was obtained with compounds that caused relatively strong expression alterations clustering in the lower part of the heatmap. In contrast, chemicals inducing relatively weak expression alterations clustered at the upper part. The genes formed several clusters; three of the clusters could be manually associated with distinct biological motifs, as indicated by the blue and red bars in Fig. 7: (1) proliferation, (2) cytochrome P450 enzymes, and (3) cell stress-associated clusters. The proliferation-associated genes clustered to the left and were exclusively downregulated (genes: CXCL6 to CDK1; compounds: AFB_1 to CHX). They include well-characterized genes, such as cyclin A (CCNA2), cyclin-dependent kinase 1 (CDK1), and topoisomerase II alpha (TOP2A). Approximately in the center of the heatmap, a cluster with predominantly upregulated cytochrome P450 isoenzymes was observed (CYP3A7, CYP3A4, CYP1A1, CYP1A2). More to the right, a further cluster of upregulated genes that are associated with different types of cell stress could be seen, including heat-shock 70-KD protein 6 (HSPAG); endoplasmic reticulum stress inducible ATF3 (which is also known to be regulated by heat-shock proteins); RGCC which is induced by TP53 in response to DNA damage; FBX032, a TGF beta target gene involved in regulating cell survival; and pyruvate dehydrogenase kinase 4 (PDK4), which is known to be upregulated in response to starvation or hypoxia.

Stereotypic versus compound-specific gene expression responses

Introduction of the selection value concept

As described in the previous paragraph, unsupervised clustering identified sets of genes that were affected by large numbers of compounds, i.e., the proliferation cluster (Fig. 7). In contrast, the expression levels of other genes were influenced only by individual compounds. To systematically analyze stereotypic versus compound-specific gene expression responses, the *selection value* concept was introduced

(Fig. 8). Based on the sample subset that was treated with the highest concentration for 24 h, a list was generated for each probe set, with compounds ranked in order of fold change. Thus, for the upregulated probe sets, compounds were ranked from high to low fold change, and for the downregulated probe sets, compounds were ranked in the reverse order. Selection value *x* (short SV*x*) was then defined to deliver a list of genes at least threefold up- or downregulated by at least *x* compounds. Selection value 20 (SV20) accordingly delivers a list of genes deregulated by 20 or more different compounds; whereas, lists based on SV5, SV3, and SV1 comprise genes deregulated by at least five, three, and one single compound, respectively (Table S11). Genes selected by SV20 exhibited a stereotypic expression response of hepatocytes exposed to chemicals. Applying this concept to the curated Open TG-GATEs data, SV20 delivered a list of 31 upregulated probe sets, which increased to 531; 1,101; and 4,135 for SV5, SV3, or SV1, respectively (Fig. 8b; Fig S4, Table S12). The corresponding numbers for the downregulated probe sets were 179; 857; 1,713; and 4,479 for SV20, SV5, SV3, and SV1, respectively (Fig. 9).

The SV20 genes were chosen for further detailed analysis as they represent a stereotypical or consensus response of hepatocytes exposed to chemicals (Table 1). It should be considered that 20 compounds represent a high fraction, because only 32 of all tested compounds induced strong expression responses (Fig. 4). Next, the functions of all SV20 genes were studied individually and a manual assignment of function was performed, based on the literature, in order not to rely only on computationally based GO analysis which will be described later. Most of the upregulated genes were associated with biological functions, such as phase I and II metabolism, differentiation and development, protein modification and degradation, stress response, as well as energy and lipid metabolism (Table 1A). In contrast, most of the downregulated SV20 genes represented genes involved in cell cycle progression (Table 1B; Table S11). A smaller fraction of the downregulated SV20 genes is associated with DNA synthesis and repair, immune response, cytoskeleton and intracellular trafficking, as well as metabolism. A more comprehensive overview is provided in Table S14, which lists the top 100 upregulated consensus probe sets (69 genes) that were upregulated by at least 20 compounds.

Besides the SV20 genes, a detailed analysis was also performed for the SV3 genes (Table S11). The intention was to consider more individual expression responses, i.e., genes up- or downregulated by at least three compounds. In principle, an even more individual response could be studied for single compounds using, in the terminology of the *selection value* concept, the SV1 genes. However, since the individual compounds were tested only in two replicates, the probability of false positives due to multiple testing would be relatively high. Therefore, the analysis of
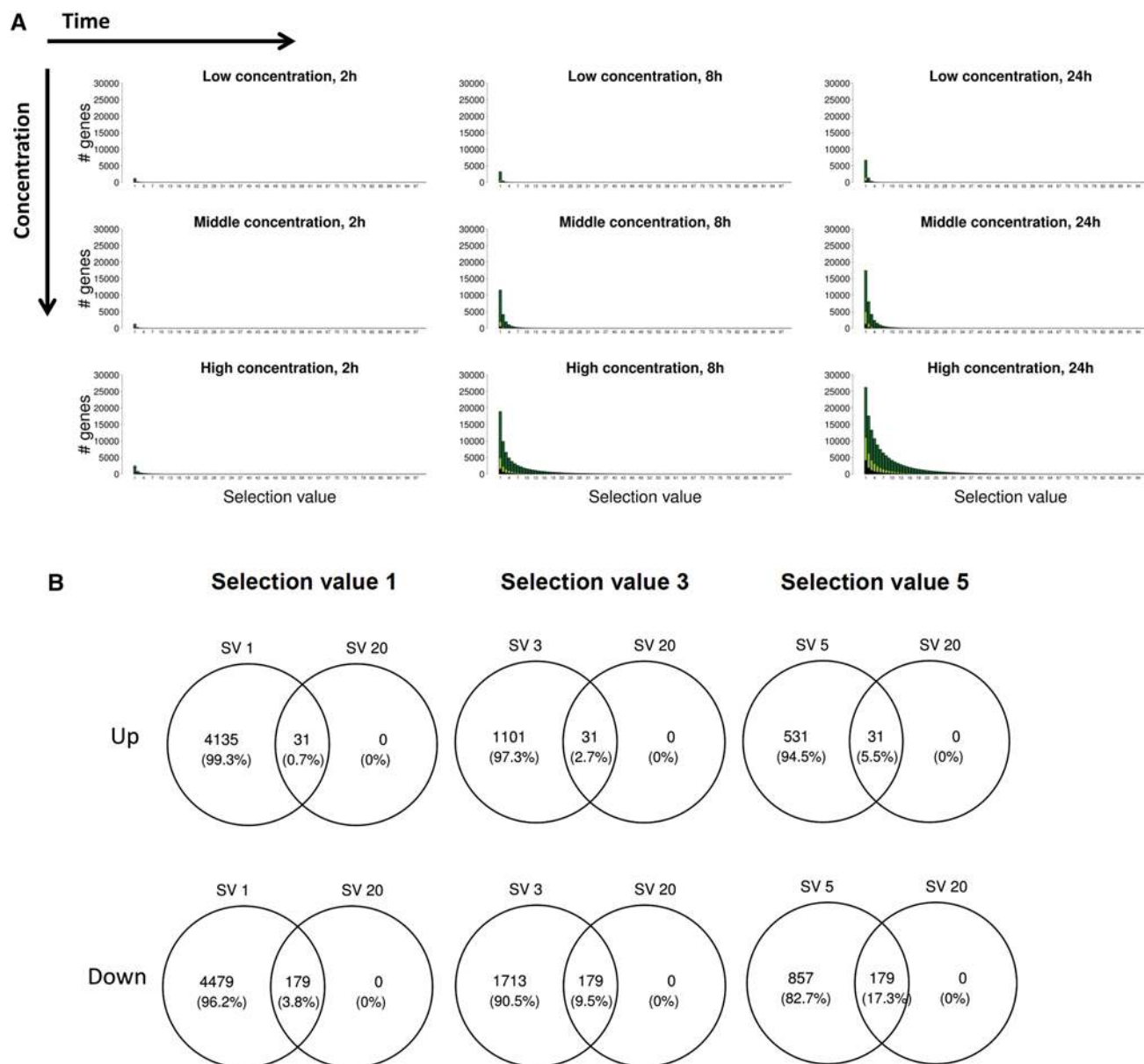
Fig. 8 a Selection values for the upregulated genes. A selection value of, e.g., five means that at least five compounds upregulate (>three-fold) the indicated gene. The corresponding data for downregulated genes are in Figure S4. **b** Overview of the number of selection value 1, 3, 5, and 20 genes. For example, 31 genes are upregulated (>three-fold) by at least 20 compounds (SV20)

SV3 genes for the identification of compound-specific gene expression responses represents a compromise between individuality and reliability. Manual inspection of the SV3 genes revealed a more diverse pattern of biological functions (Table S15). As observed for the SV20 genes, energy and lipid metabolism were frequently observed functions among the upregulated SV3 genes (Table S15A). Further categories were inflammation, development and differentiation, protein degradation as well as regulation of transcription, metabolism, stress response and apoptosis, membrane transporters, and cytoskeletal factors. Typical functions of the downregulated genes are factors involved in differentiation, endogenous and xenobiotic metabolism, cytoskeletal organization, immune response, transporters, energy and lipid metabolism, and apoptosis (Table S15B).

Overrepresented gene ontology groups and transcription factor binding sites

Analysis of overrepresented GO groups resulted in metabolism of xenobiotics and endogenous substrates as the predominating biological functions (Table 2A), thereby confirming the conclusions of the aforementioned manual categorization. To obtain an overview of overrepresented
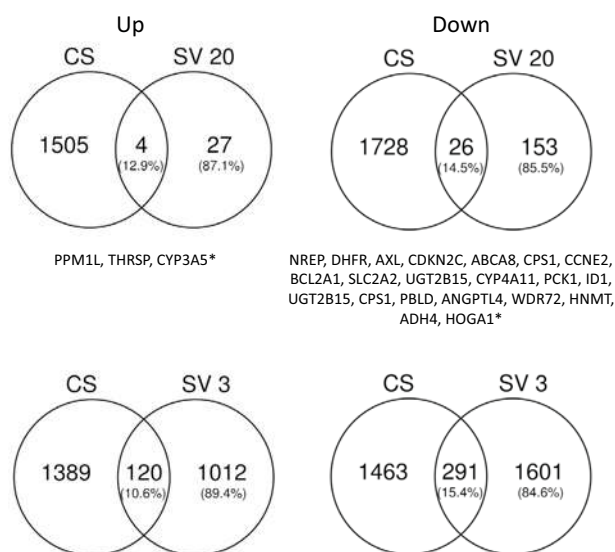
Fig. 9 Overlap between 'unstable baseline genes' (CS) and the SV20 (SV3) genes. For example, 4 of the 31 SV20 genes belong to the unstable baseline genes, meaning that their expression levels were altered by the hepatocyte isolation and cultivation procedure. The uniquely annotated genes in the overlap of the SV20 genes are listed below the corresponding Venn diagrams (the asterisk refers to probe sets that are not annotated). The genes in the overlap of the SV3 genes are listed in Table S9. All specific genes are listed in Table S9 as well

transcription factor binding sites (TFBS), the SV20 genes were analyzed by the PRIMA software. Among the down-regulated genes (Table S11), TFBS of proliferation-associated genes were overrepresented, such as E2F1 and ATF (Table 2B). This corresponds to the results of the GO group analysis where the top overrepresented GO groups all are associated with cell cycle progression and proliferation (Table 2). In the set of upregulated genes (Table S11), the well-established transcription factor of hepatocyte differentiation and a central regulator of liver function, HNF4, (Watt et al. 2003; Kamiya et al. 2003), was overrepresented (Table 2). Table S16 provides an overview of the overrepresented GO groups (unadjusted p value ≤ 0.001) and overrepresented transcription factor binding sites (unadjusted p value ≤ 0.01) for the SV3 genes.

Gene alterations common to both chemical exposure in vitro and human liver disease

It is of high interest to know whether genes deregulated by exposure to chemicals in vitro would respond similarly under conditions of in vivo exposure. Systematic in vitro versus in vivo comparison between cultivated human hepatocytes and human liver tissue after exposure to a test compound is not possible. Human liver tissue after chemical exposure is only available under exceptional and

usually not precisely defined conditions, such as liver tissue obtained from patients with acetaminophen intoxication. To nevertheless approach this question, publicly available whole-transcriptome gene expression datasets were explored to define genes that are differentially expressed in patients with NASH, liver cirrhosis, and HCC, based on the comparison to healthy/non-tumor liver tissue (Table S17). Secondly, the overlap between 'differentially expressed liver disease genes' and chemically deregulated stereotypic genes in vitro, determined by SV20, was plotted (Fig. 10a).

For both liver cirrhosis and hepatocellular carcinoma, the SV20 overlap for downregulated genes ranged from 13 to 16 %. For non-alcoholic steatohepatitis, the overlap was smaller for the downregulated genes, which could be explained by the generally smaller number of differentially expressed genes in this set of data. The phase II metabolizing enzyme SULT1C2, which was upregulated by at least 20 chemicals in vitro (fold change >3) and revealed significantly increased expression levels in diseased liver tissue, provides one example of a gene that is deregulated in both the in vitro and the in vivo situation. Similarly, CYP3A7, the predominant cytochrome P450 in human fetal liver (Pang et al. 2012) and the p53-induced gene RGCC (Huang et al. 2009; Saigusa et al. 2007) were increased by chemical exposure in vitro and in at least two of the studied human liver conditions. Genes that were downregulated by at least 20 chemicals (more than threefold compared to controls) and showed significantly lower expression levels in at least two liver diseases include the aldehyde dehydrogenase family members ALDH8A1 and ADH4, the sterol- and fatty acid-metabolizing cytochrome P450 isoenzymes CYP8B1 and CYP4A11, the urea cycle enzyme CPS1, the gluconeogenesis key enzyme PCK1, the membrane-associated ATP-binding cassette transporter ABCA8, and the glucose transporter SLC2A2. Similarly as for SV20, a considerable overlap was observed between the disease genes and the genes identified by SV3 (Fig. 10b), with the individual genes summarized in Table S18.

Discussion

How to use the toxicogenomics directory

Currently, numerous studies are performed to identify biomarkers of hepatotoxicity, and transcriptomics data are frequently used to identify candidates for further evaluation in biomarker evaluation studies. However, a systematic directory summarizing key features of chemically influenced genes is not available. To bridge this gap and to establish a systematic and comprehensive strategy for the identification of candidate genes, we used the Open TG-GATES database with Affymetrix files of cultivated human

Table 1 Consensus genes deregulated in human hepatocytes by chemical exposure. The listed genes are at least threefold A up- or B downregulated by at least 20 of the 143 studied chemicals (selection value 20)

Symbol	Gene	Probe set	Function of the gene product	
A: Upregulated consensus genes				
Category: Metabolism, xenobiotics				
CYP1A1	Cytochrome P450, sub-fam. 1A, polypeptide 1	205749_at	Metabolic enzyme in the ER	Phase I enzymes
CYP2C9	Cytochrome P450, sub-fam. 2C, polypeptide 9	217558_at		
CYP3A4	Cytochrome P450, sub-fam. 3A, polypeptide 4	205999_x_at 208367_x_at		
CYP3A5[2]	Cytochrome P450, sub-fam. 3A, polypeptide 5	214235_at 243015_at		
CYP3A7[1]	Cytochrome P450, sub-fam. 3A, polypeptide 7	205939_at 211843_x_at		
SULT1C2[1]	Sulfotransferase 1C2	205342_s_at	Cytosolic enzyme; catalyzes sulfonation	Phase II enzymes
SULT2A1	Sulfotransferase 2A1	206292_s_at 206293_at		
Category: Differentiation and development				
FGF21	Fibroblast growth factor 21	221433_at	Secreted growth factor; mitosis and survival	Growth factor
GDF15[1]	Growth/differentiation factor 15	221577_x_at	Secreted growth factor; inflammation and apoptosis	
IFRD1	Interferon-related developmental regulator 1	202147_s_at	Nuclear protein; regulation of gene expression in proliferative and differentiative pathways	Interferon-related signaling
EFNA1[1]	Ephrin-A1	202023_at	Receptor tyrosine kinase; migration and adhesion	Other
Category: Protein modification and degradation				
CBX4[1]	E3 SUMO-protein ligase CBX4	227558_at	Protein ligase; SUMO1 conjugation and proteasomal degradation	
FBXO32[1]	F-box protein 32	225803_at	Cytosolic protein; ubiquitination and proteasomal degradation	
KLHL24	Kelch-like protein 24	221985_at 221986_s_at	Cytosolic protein; role in protein degradation	
Category: Stress response				
ATF3	Activating transcription factor 3	202672_s_at 1554980_a_at	Transcription factor; stress response, further involved in cell cycle regulation, DNA repair, apoptosis	Cell cycle arrest
RGCC[1]	Regulator of cell cycle	218723_s_at	Cytosolic protein; induced by p53 modulates the activity of cell cycle specific kinases in response to DNA damage	
CREBRF	CREB3 regulatory factor	225956_at	Nuclear protein; regulates transcription, negative regulator of the ER stress response	Stress response/ER stress
PPM1E	Phosphoprotease 1E	205938_at	Serine/threonine-protein phosphatase; negative regulator of cell stress response pathways	
Category: Energy and lipid metabolism				
PDK4[1]	Pyruvate dehydrogenase kinase	225207_at	Mitochondrial membrane enzyme; increased PDK4 leads to enhanced gluconeogenesis	Glucose metabolism/homeostasis
PPM1L[2]	Protein phosphatase 1L	228108_at	Membrane bound enzyme; regulation of blood-glucose	
Category: Other				
SLC7A11[1]	Solute carrier fam. 7 member 11	217678_at	membrane anchored protein; cysteine and glutamate transport	
ZCCHC6[1]	Terminal uridylyltransferase 7	242776_at	Enzyme involved in RNA processing	
[3]236542_at[1], 237031_at				
B: Downregulated consensus genes. The 100 probe sets (74 genes) with the highest fold change are given below				
Category: Cell cycle progression and regulation				
ASPM	Abnormal spindle protein homolog	219918_s_at	Cytosolic protein; role in mitotic spindle regulation and coordination of mitotic processes	
AURKA	Aurora kinase A	208079_s_at	Cytosolic kinase; regulation of cell cycle progression	

Table 1 continued

Symbol	Gene	Probe set	Function of the gene product
AURKB	Aurora kinase B	209464_at	Cytosolic kinase; regulation of cell cycle progression
BIRC5	Baculoviral IAP repeat-containing protein 5	202095_s_at	Cytosolic protein; chromosome alignment and segregation during mitosis, further role as apoptotic factor
CENPK	Centromere protein K	222848_at	Nuclear protein; assembly of kinetochore proteins, mitotic progression and chromosome segregation
FAM83D	Protein FAM83D	225687_at	Cytosolic protein; chromosome alignment on the spindle
MLF1IP	Centromere protein U	218883_s_at	Component of a nucleosome-associated complex; assembly of kinetochore proteins
NCAPG	Condensin-2 complex subunit G2	218662_s_at 218663_at	Nuclear protein; regulation of mitotic chromosome architecture
OIP5	Opa-interacting prot. 5	213599_at	Nuclear protein; chromosome segregation
TTK	Dual specificity protein kinase TTK	204822_at	Protein kinase; centrosome duplication and mitosis progression, associated with cell proliferation
CCNA2	Cyclin A2	203418_at 213226_at	Nuclear protein; cell cycle control at the G1/S and G2/M transitions
CCNB1	Cyclin B1	214710_s_at 228729_at	Cytosolic and nuclear protein; cell cycle control at the G2/M transition
CCNB2	Cyclin B2	202705_at	
CCNE2[2]	Cyclin E2	205034_at 211814_s_at	Nuclear protein; controls the cell cycle at the late G1 and early S phase
CDC20	Cell division cycle 20 homolog	202870_s_at	Cytosolic protein; regulation of anaphase initiation and mitotic exit
CDC6	Cell division control protein 6 homolog	203967_at 203968_s_at	Cytosolic and nuclear protein; control and initiation of DNA replication
CDCA3	Cell division cycle-associated protein 3	223307_at	Cytosolic protein; required for entry into mitosis
CDK1	Cyclin-dependent kinase 1	203213_at 203214_x_at 210559_s_at	Kinase, cell cycle control by modulation of the centrosome cycle and mitosis initiation
CDKN3	Cyclin-dependent kinase inhibitor 3	209714_s_at 1555758_a_a	Cytosolic protein; cell cycle regulation
DLGAP5	Disks, large homolog-associated protein 5	203764_at	Cytosolic and nuclear protein; cell cycle regulator
DTL	Denticleless protein homolog	218585_s_at 222680_s_at	Cytosolic and nuclear protein; cell cycle control, DNA repair
HMMR	Hyaluronan-mediated motility receptor	207165_at 209709_s_at	Cell surface receptor; required for entry and regulation of mitosis
MELK	Maternal embryonic leucine zipper kinase	204825_at	Serine/threonine-protein kinase; modulator of intracellular signaling, further role in apoptosis
TRIP13	Thyroid receptor-interacting protein 13	204033_at	Transcription factor interacting protein; checkpoint arrest, chromosome recombination and structure development during meiosis, role in DNA double strand break repair
BUB1	Serine/threonine-protein kinase BUB1	209642_at	Cytoplasmic and nuclear kinase; mitotic checkpoint, required for normal mitosis progression
BUB1B	Serine/threonine-protein kinase BUB1b	203755_at	
CASC5	Cancer susceptibility candidate 5	228323_at	Nuclear protein; spindle-assembly checkpoint signaling and chromosome alignment
INSC	Protein inscuteable homolog	237056_at	Cytosolic protein; spindle orientation
MAD2L1	Mitotic arrest-deficient 2L1	203362_s_at 1554768_a_at	Cytosolic and nuclear protein; spindle-assembly checkpoint
NDC80	Kinetochore protein NDC80 homolog	204162_at	Nuclear protein; chromosome segregation and spindle checkpoint activity
Symbol	Gene	Probe set	Function of the Gene Product
Category: Cell cycle progression and regulation			
NEK2	Never in mitosis A-related kinase 2	204641_at	Mitotic kinase; controls centrosome separation and bipolar spindle formation in mitotic cells

Table 1 continued

Symbol	Gene	Probe set	Function of the gene product
SPC25	Kinetochore protein Spc25	209891_at	Nuclear protein; chromosome segregation, spindle checkpoint activity
TPX2	Targeting protein for Xklp2	210052_s_at	Spindle-associated protein; spindle-assembly factor, colocalises with apoptotic microtubules
ZWINT	ZW 10 interactor	204026_s_at	Nuclear protein of the MIS12 complex; kinetochore formation and spindle checkpoint activity
GINS1	GINS complex subunit 1	206102_at	Nuclear protein; initiation of DNA replication and progression of DNA replication forks
GINS2	GINS complex subunit 2	221521_s_at	
KIAA0101	PCNA-associated factor	202503_s_at	Cytosolic and nuclear protein; regulates DNA repair during DNA replication
LMNB1	Lamin-B1	203276_at	Membrane protein in the nuclear laminar; DNA replication, stress response and development
MCM10	Minichromosome maintenance 10	220651_s_at	Nuclear protein; functions as replication initiation factor
TOP2A	Topoisomerase II alpha	201291_s_at 201292_at	Nucleoplasm enzyme; transiently breaks and reunites double strand DNA during replication
KIF23	Kinesin-like protein KIF23	204709_s_at	Protein required for the myosin contractile ring formation during cytokinesis
KIF4A	Chromosome-associated kinesin KIF4A	218355_at	Microtubulus motor protein; spindle formation
NUSAP1	Nucleolar and spindle-associated protein 1	218039_at 219978_s_at	Cytosolic and nuclear protein; stabilization of microtubules
ANLN	Anillin	222608_s_at 1552619_a_at	Nuclear, actin binding protein; role in cytokinesis, deregulated in many cancers
CEP55	Centrosomal protein 55 kDa	218542_at	Cytosolic and nuclear protein; mitotic exit and cytokinesis
PRC1	Protein regulator of cytokinesis 1	218009_s_at	Cytosolic and nuclear protein; regulator of cytokinesis, cross-links antiparallel microtubules
PBK	PDZ binding kinase	219148_at	Mitotic kinase; phosphorylates MAP kinase p38, only active during mitosis
PRR11	Proline rich 11	228273_at	Cytosolic protein; expression increases from G1 to G2/M phase
E2F8	Transcription factor E2F8	219990_at	Transcription factor; regulates transcription of other transcription factors with role in cell cycle
MYBL1	Myeloblastosis-related protein A	213906_at	Transcriptional activator; master regulator of meiotic genes, role in proliferation and differentiation
HELLS	Lymphoid-specific helicase	223556_at 227350_at	Helicase with role in normal development and survival, chromatin remodeling and DNA methylation
Category: DNA synthesis, recombination and repair			
DHFR[1,2]	Dihydrofolate reductase	202533_s_at	Enzyme in folate metabolism; synthesis of purines — DNA synthesis
RRM2	Ribonucleotide reductase M2 polypeptide	201890_at 209773_s_at	Cytosolic enzyme; biosynthesis of deoxyribonucleotides, inhibits Wnt signaling
TK1	Thymidine kinase	202338_at 1554408_a_at	Cytosolic kinase for DNA synthesis; phosphorylation of thymidine to deoxythymidine monophosphate
TYMS	Thymidylate synthase	202589_at 1554696_s_at	Enzyme which contributes to the de novo mitochondrial thymidylate biosynthesis pathway
FANCI	Fanconi-associated nuclease 1	213007_at	Nuclease; required for maintenance of chromosomal stability, key role in DNA repair — DNA repair/recombination
RAD51AP1	RAD51-associated protein 1	204146_at	Nuclear, DNA binding protein; DNA damage response
Category: Immune response and inflammation			
CCL2[1]	C–C motif chemokine 2	216598_s_at	Secreted protein; chemotactic factor attracting monocytes and basophils
CXCL6[1]	Chemokine (C–X–C motif) ligand 6	206336_at	Secreted protein with chemotactic activity for neutrophils; inflammation and development

Table 1 continued

Symbol	Gene	Probe set	Function of the gene product
FSTL1[1]	Follistatin-related protein 1	208782_at	Secreted glycoprotein; inflammatory protein, enhancing synthesis of pro-inflammatory cytokines and chemokines by immune cells
HAVCR1	Hepatitis A virus cellular receptor 1	207052_at	Membrane protein receptor; role in T-helper cell development, cell surface receptor for the virus
Symbol	Gene	Probe set	Function of the Gene Product
Category: Cytoskeleton and trafficking			
ARL14	ADP-ribosylation factor-like 14	220468_at	Cytoplasmic vesicle GTPase; controls transport of vesicles along microtubules — Transport/trafficking
KIF20A	Kinesin-like protein KIF20A	218755_at	Cytosolic protein; controls transport along microtubules
KRT7[1]	Keratin type II cytoskeletal 7	209016_s_at	Cytoplasmic intermediate filament protein; responsible for structural cell integrity, stimulates DNA synthesis — Cytoskeleton
PALMD	Palmdelphin	218736_s_at 222725_s_at	Cytosolic protein; role in the cell shape control and cell dynamics
Category: Metabolism			
ALDH8A1[1]	Aldehyde dehydrogenase fam. 8 member A1	220148_at	Cytosolic enzyme; converts 9-cis-retinal to 9-cis retinoic acid
CPS1[1,2]	Carbamoyl-phosphate synthase	204920_at 217564_s_at	Mitochondrial protein; role in the urea cycle, removes excess ammonia from the cell
Category: Other			
ABCA8[1,2]	ATP-binding cassette sub-family A, member 8	204719_at	Membrane located, ATP-dependent lipophilic drug transporter
BCL2A1[1,2]	Bcl-2-related protein A1	205681_at	Cytoplasmic protein involved in apoptosis regulation
DEPDC1	DEP domain-containing protein 1A	222958_s_at 232278_s_at	Nuclear protein; involved in transcriptional regulation as a transcriptional corepressor
SHCBP1	SHC SH2 domain-binding protein 1	219493_at	Cytosolic protein; signaling pathways in proliferation, cell growth and differentiation
UBE2C	Ubiquitin-conjugating enzyme E2 C	202954_at	Enzyme involved in protein ubiquitinylation and degradation
UHRF1	Ubiquitin-like PHD and RING finger domain-containing protein 1	225655_at	Nuclear epigenetic regulator enzyme; bridges DNA methylation and chromatin modification
WDR72[1,2]	WD repeat-containing protein 72	227174_at	Mutations in this gene have been associated with amelogenesis imperfecta hypomaturation type 2A3

[3]204962_s_at, 225834_at, 229490_s_at, 230554_at, 232325_at[1], 244567_at

[1] Gene deregulated in liver disease (NASH, cirrhosis and/or HCC). [2]Unstable baseline gene. [3]Not annotated, functionally unclear probe set

hepatocytes incubated with 158 chemicals, further sets of gene array data from human donors that were generated in this study, and publicly available genome-wide datasets of human liver tissue from patients with NASH, cirrhosis, and HCC. This resulted in a toxicotranscriptomics directory that is now publicly available under http://wiki.toxbank. net/toxicogenomics-map/. The directory can be downloaded in the form of an EXCEL table in which each gene can be identified together with answers to the following questions: (1) is the gene deregulated by chemicals and, if yes, by how many and which compounds; (2) is the change in gene expression the stereotypical response that is typically observed when hepatocytes are exposed to high concentrations of chemicals; (3) is the gene also deregulated in human liver disease; (4) does the gene belong to a group of 'unstable baseline genes' that are up or downregulated due

to the culture conditions without exposure to chemicals? In the following paragraph, the basis of these questions and their relevance to the identification of adequate genes in biomarker identification programs will be discussed.

Stereotypical versus specific responses

One basic principle observed in this study is that complex but stereotypical expression responses are caused by close to cytotoxic concentrations of numerous compounds; this stereotypical response should be differentiated from compound-specific influences. The principle of stereotypical- and compound-specific responses was first observed by unsupervised clustering, followed by a systematic study using the selection value (SV) concept. Unsupervised clustering of the genes deregulated by chemicals in vitro

Table 2 A Overrepresented GO groups for SV20 genes (unadjusted *p* value ≤0.001, in total 13 upregulated, here are all listed, in total 88 down-regulated, here only the top 15 are listed). B Overrepresented TFBS (unadjusted *p* value ≤0.01)

A: Overrepresented GO groups

Up

Xenobiotic metabolic process (GO:0006805), alkaloid catabolic process (GO:0009822), monoterpenoid metabolic process (GO:0016098), vitamin D metabolic process (GO:0042359), exogenous drug catabolic process (GO:0042738), sulfation (GO:0051923), 3′-phosphoadenosine 5′-phosphosulfate metabolic process (GO:0050427), bile acid catabolic process (GO:0030573), drug catabolic process (GO:0042737), positive regulation of S phase of mitotic cell cycle (GO:0045750), steroid catabolic process (GO:0006706), dibenzo-p-dioxin catabolic process (GO:0019341)

Down

Cell division (GO:0051301), mitotic prometaphase (GO:0000236), mitosis (GO:0007067), regulation of transcription involved in G1/S phase of mitotic cell cycle (GO:0000083), DNA replication (GO:0006260), protein localization to kinetochore (GO:0034501), regulation of cyclin-dependent protein kinase activity (GO:0000079), anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process (GO:0031145), DNA replication initiation (GO:0006270), DNA strand elongation involved in DNA replication (GO:0006271), S phase of mitotic cell cycle (GO:0000084), mitotic sister chromatid segregation (GO:0000070), organ regeneration (GO:0031100), CENP-A containing nucleosome assembly at centromere (GO:0034080), M/G1 transition of mitotic cell cycle (GO:0000216)

B: Overrepresented TFBS

Up

HNF4 (M01032), TCF-4 (M00671), Nkx2-5 (M00240), GATA-1 (M00128), FOXO4 (M00472), GATA-X (M00203), LEF1 (M00805), Gfi-1 (M00250)

Down

E2F (M01032), E2F-1 (M00671), AFP1 (M00240), Nrf-1 (M00128), IRF-7 (M00472), E2F-1 (M00203), ATF (M00805), BRCA1 (M00250), NF-Y (M01032), PXR,_CAR,_LXR,_FXR (M00671), RORalpha1 (M00240)
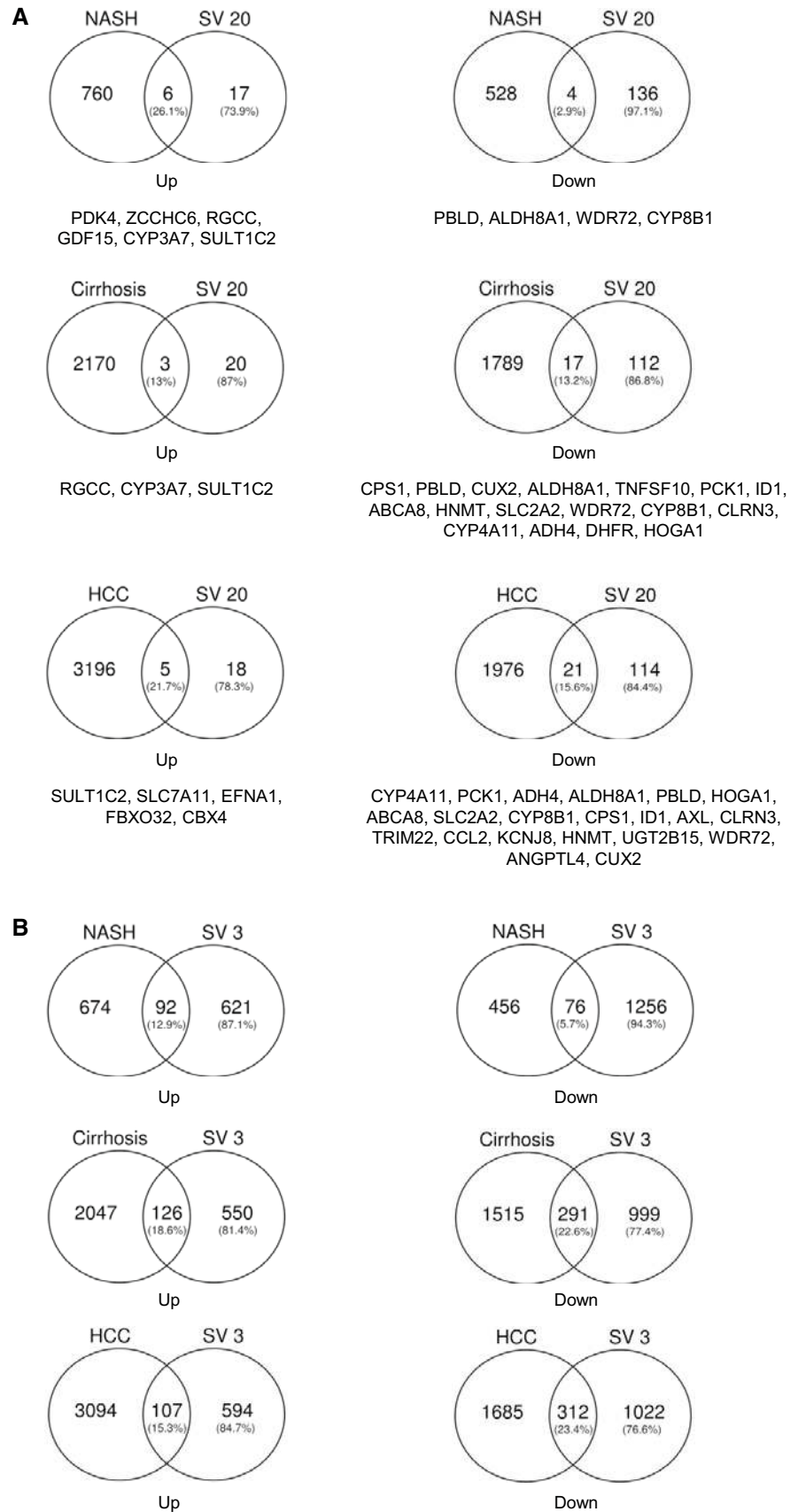
illustrated that some clusters were deregulated by relatively large numbers of chemicals. To further analyze these stereotypical responses, selection value 20 genes (SV20) were defined as genes that were at least threefold up- or downregulated by at least 20 test compounds. Twenty compounds represent a relatively large fraction, because only 32 of the studied compounds contributed to the 100 most deregulated genes. Of the 31 up- and 179 downregulated 'consensus' or SV20 genes identified, GO group analysis identified xenobiotic metabolism as the most significantly overrepresented motif of the upregulated and cell cycle progression of the downregulated genes.

The knowledge whether a gene belongs to the stereotypical response of chemically exposed hepatocytes is relevant when selecting candidate genes for biomarker evaluation or other studies. On the one hand, deregulation of 'stereotypical stress response genes' reliably indicates that the hepatocytes are stressed by exposure to the test compound. On the other hand, this deregulation is unlikely to represent a specific molecular mechanism of toxicity. While the SV20 genes indicate a stereotypical stress response, the SV3 genes may be helpful when attempting to identify specific mechanisms of toxicity. The SV3 genes represent a broad spectrum of biological functions, including energy and lipid metabolism, inflammation, differentiation, protein modification and degradation, endogenous and xenobiotic metabolism, cytoskeletal organization, immune response and several factors involved in transcriptional regulation, thereby representing candidates for further studies aimed at identifying the responsible mechanisms of hepatotoxicity.

Human disease genes

The toxicotranscriptomics directory (http://wiki.toxbank.net/toxicogenomics-map/) indicates whether a gene is up- or downregulated in NASH, cirrhosis, or hepatocellular cancer (HCC). Also included are not only the genes that are up- or downregulated by chemicals in vitro, but also those deregulated in the same direction (either up or down) in human liver disease. This may be of interest for the selection of candidates for biomarker evaluation programs in toxicology, because human disease genes reflect mechanisms triggered by a disturbed liver microenvironment in vivo. One of these genes is the phase II metabolizing enzyme SULT1C2, which is upregulated in NASH, cirrhosis, and HCC and also by at least 20 of the analyzed chemicals. Similarly, CYP3A7 is a SV20 chemical consensus gene that is also upregulated in NASH and cirrhosis. CYP3A7 is the predominant cytochrome P450 in human fetal liver, while CYP3A4 becomes most abundant after birth (Pang et al. 2012). Therefore, chemicals upregulating this cytochrome P450 isoform reactivate a fetal expression pattern. The present study shows that a similar response is induced by hepatotoxic compounds and in liver disease. Thus, induction of CYP3A7 seems to represent a stereotypical response of stressed hepatocytes in vitro and in vivo. A further gene where alteration in expression overlaps in both chemical stereotypical (SV20) and disease genes is RGCC. Relatively little is known about this gene. It was previously reported to cause epithelial to mesenchymal transition in kidney cells (Huang et al. 2009), and to represent a p53

Fig. 10 Overlap between genes altered by the test compounds (**a** SV20 genes, **b** SV3 genes) and genes altered by the human liver diseases non-alcoholic steatohepatitis (NASH), liver cirrhosis, and hepatocellular cancer (HCC). The genes differentially expressed in each of the three liver diseases are in Table S17. **a** The genes in the overlap are listed below the corresponding Venn diagrams. **b** The genes in the overlap are listed in Table S18

A

NASH | SV 20
760 | 6 (26.1%) | 17 (73.9%)
Up

PDK4, ZCCHC6, RGCC, GDF15, CYP3A7, SULT1C2

NASH | SV 20
528 | 4 (2.9%) | 136 (97.1%)
Down

PBLD, ALDH8A1, WDR72, CYP8B1

Cirrhosis | SV 20
2170 | 3 (13%) | 20 (87%)
Up

RGCC, CYP3A7, SULT1C2

Cirrhosis | SV 20
1789 | 17 (13.2%) | 112 (86.8%)
Down

CPS1, PBLD, CUX2, ALDH8A1, TNFSF10, PCK1, ID1, ABCA8, HNMT, SLC2A2, WDR72, CYP8B1, CLRN3, CYP4A11, ADH4, DHFR, HOGA1

HCC | SV 20
3196 | 5 (21.7%) | 18 (78.3%)
Up

SULT1C2, SLC7A11, EFNA1, FBXO32, CBX4

HCC | SV 20
1976 | 21 (15.6%) | 114 (84.4%)
Down

CYP4A11, PCK1, ADH4, ALDH8A1, PBLD, HOGA1, ABCA8, SLC2A2, CYP8B1, CPS1, ID1, AXL, CLRN3, TRIM22, CCL2, KCNJ8, HNMT, UGT2B15, WDR72, ANGPTL4, CUX2

B

NASH | SV 3
674 | 92 (12.9%) | 621 (87.1%)
Up

NASH | SV 3
456 | 76 (5.7%) | 1256 (94.3%)
Down

Cirrhosis | SV 3
2047 | 126 (18.6%) | 550 (81.4%)
Up

Cirrhosis | SV 3
1515 | 291 (22.6%) | 999 (77.4%)
Down

HCC | SV 3
3094 | 107 (15.3%) | 594 (84.7%)
Up

HCC | SV 3
1685 | 312 (23.4%) | 1022 (76.6%)
Down

inducible gene involved in cell cycle arrest (Saigusa et al. 2007). ALDH8A1 is one example of a gene that belongs to the downregulated chemical consensus (SV20) genes which is also decreased in NASH, cirrhosis, and HCC. This may be of pathophysiological relevance, since ALDH8A1 converts 9-cis-retinal into 9-cis retinoic acid (Lin and Napoli 2000). 9-cis retinoic acid is a ligand of the retinoid X receptor (RXR) in hepatocytes, one of the master regulators of gene expression.

Further genes that overlapped between SV20 and liver disease are involved in normal metabolic liver functions. Downregulation of these genes may represent a state where the hepatocytes shift their balance from metabolism to regeneration. Examples are CPS1, the first enzyme of the urea cycle and key factor of ammonia detoxification (Simmer et al. 1990; Schliess et al. 2014); PCK1 as a main control enzyme of gluconeogenesis (Pilz et al. 1992); SLC2A2, also known as the glucose carrier GLUT2 (Froguel et al. 1991); CYP8B1, a key enzyme in bile acid metabolism (Gåfvels et al. 1999); CYP4A11, the major fatty acid omega-hydroxylase, which is involved in controlling the balance of lipids (Antoun et al. 2006); ABCA8, one of the liver's ABC transporters (Tsuruoka et al. 2002); and ADH4, an aldehyde dehydrogenase that metabolizes numerous substrates, including retinol, hydroxysteroids, and also ethanol (Kimura et al. 2009). Such complex but stereotypical patterns of gene deregulation induced by a chemical can also be interpreted as a situation of disturbed hepatocyte physiology and could be the result of different insults to the liver. However, if a gene induced by chemicals in vitro is also induced by the micro-environment of a diseased liver, this at least demonstrates that the involved mechanism is not a pure in vitro artifact. In vivo validation is of high relevance, since parts of the signaling network of cultivated hepatocytes are altered compared to hepatocytes in an intact liver, for example enhanced Akt activity mediating antiapoptotic mechanisms or increased MAK kinase signaling that causes features of epithelial-to-mesenchymal transdifferentiation (Godoy et al. 2009; 2010); therefore, many responses observed in cultivated cells represent in vitro artifacts and should not be used for evaluation of chemicals. Only approximately 20 % of the chemically influenced genes in hepatocytes in vitro overlap with the genes altered in disease. This of course does not mean that the remaining 80 % are irrelevant, but rather that these genes may lead to specific mechanisms of chemical toxicity that are not induced in NASH, cirrhosis, and hepatocellular cancer. However, further studies are required to evaluate their in vivo relevance.

Unstable baseline genes

The toxicotranscriptomics directory (http://wiki.toxbank. net/toxicogenomics-map/) also contains the information of which genes are up- or downregulated as a consequence of isolation stress or cultivation conditions of the hepatocytes. Such unstable baseline genes can nevertheless represent useful biomarkers. However, a careful comparison with time-matched controls should be performed to avoid false-positive results. If equivalent alternatives are available, it may be advisable to avoid 'unstable baseline genes.'

Limitations and technical aspects

One limitation of the present dataset of cultivated human hepatocytes in Open TG-GATEs is that only two replicates are available. Since this low number limits the validity of statistical tests, genes were considered as up- or down-regulated when the mean difference to controls was at least threefold. It may be worthwhile to determine whether smaller thresholds would also be useful, because gene expression changes smaller than threefold can also be of toxicological relevance. However, the number of false-positive genes due to multiple testing will increase with decreasing thresholds. Since the goal of this study was to identify general principles of the toxicotranscriptome in human hepatocytes, we preferred a relatively high threshold to decrease the probability of false-positive results, even with the disadvantage that the lists of differential genes remain incomplete. Moreover, replications of gene expression analyses in hepatocytes from different donors using a small number of compounds showed that the majority, but certainly not all, gene deregulations observed in the Open TG-GATEs dataset are reproducible. As mentioned above, we recommend the use of SV3 genes as candidates when searching for compound-specific mechanisms, because higher data stability can be expected. Because only two replicates are available, SV1 genes (meaning a threefold up- or downregulation for at least one compound) have a high probability of containing false positives. This risk substantially decreases if similar deregulations are observed for at least three compounds (SV3). The problems associated with the fact that only two replicates were tested per compound demonstrate how important it will be in the future to analyze each compound with more than two biological replicates.

The PCA and heatmap analyses presented in this study were based on the 100 most deregulated probe sets across all analyzed compounds. In a similar fashion, previous studies were based on the 100 genes with highest variability (Krug et al. 2013; Waldmann et al. 2014). We have seen that this technique offers optimal conditions to obtain an overview of compound effects. However, the reported results did not depend on the choice of a certain number of probe sets. Similar principles were seen if the PCA or heatmap analyses were based on 50, 500, or even all probe sets.

Two of the challenges working with large datasets are batch effects, as well as experimental errors in a subset

of samples. This is difficult to avoid, especially when large datasets are generated by consortia with several partners. An example of an unexpected result in the Open TG-GATEs dataset is the expression response induced by carbon tetrachloride in cultivated human hepatocytes which caused only very few gene expression alterations, although carbon tetrachloride is known to represent a strong hepatotoxic compound (Hoehme et al. 2007, 2010). An experimental explanation may be that it is technically challenging to suspend the highly lipophilic carbon tetrachloride in a way that all cultivated cells are exposed homogeneously (Bauer et al. 2009). This unexpected subset of data was identified by a tool developed in this study, the 'progression profile error indicator,' which identifies non-monotonous concentration progression of deregulated genes. Some compounds, including carbon tetrachloride, showed an unusual concentration progression with a high fraction of genes deregulated at a lower but not at higher concentration. Here, we decided to exclude such unusual datasets identified by a number of curation steps. This seems to be justified for the present study, which aims at identification of the underlying key principles of global gene expression alterations. The compounds with unusual concentration progression will be reanalyzed in a follow-up study.

Perspectives

The concentrations of the present study were chosen based on the cytotoxicity. Gene arrays were performed with close to cytotoxic compound concentrations. In a preliminary analysis, we already tested whether the gene expression data differentiate between hepatotoxic and non-hepatotoxic compounds. Based on the literature data, it was differentiated whether the tested chemicals were previously reported to be hepatotoxic in humans or not. However, unsupervised clustering did not differentiate between the hepatotoxic and non-hepatotoxic chemicals (Fig. 7). This is not surprising, since the tested concentrations were not selected to represent an in vivo relevant range but are based on the cytotoxicity. Future studies with in vivo relevant concentrations have to show whether selected biomarkers can discriminate between hepatotoxic and non-hepatotoxic substances. In conclusion, the presented toxicogenomics directory offers a basis for a rationale choice of genes for such studies.

References

Alexa A, Rahnenführer J (2010) topGO: topGO: enrichment analysis for gene ontology. R package version 2.12.0

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11(10):R106

Antoun J, Amet Y, Simon B, Dréano Y, Corlu A, Corcos L, Salaun JP, Plée-Gautier E (2006) CYP4A11 is repressed by retinoic acid in human liver cells. FEBS Lett 580(14):3361–3367

Balmer NV, Klima S, Rempel E, Ivanova VN, Kolde R, Weng MK, Meganathan K, Henry M, Sachinidis A, Berthold MR, Hengstler JG, Rahnenführer J, Waldmann T, Leist M (2014) From transient transcriptome responses to disturbed neurodevelopment: role of histone acetylation and methylation as epigenetic switch between reversible and irreversible drug effects. Arch Toxicol 88(7):1451–1468

Bauer A, Schumann A, Gilbert M, Wilhelm C, Hengstler JG, Schiller J, Fuchs B (2009) Evaluation of carbon tetrachloride-induced stress on rat hepatocytes by 31P NMR and MALDI-TOF mass spectrometry: lysophosphatidylcholine generation from unsaturated phosphatidylcholines. Chem Phys Lipids 159(1):21–29

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc B 57:289–300

Campos G, Schmidt-Heck W, Ghallab A, Rochlitz K, Pütter L, Medinas DB, Hetz C, Widera A, Cadenas C, Begher-Tibbe B, Reif R, Günther G, Sachinidis A, Hengstler JG, Godoy P (2014) The transcription factor CHOP, a central component of the transcriptional regulatory network induced upon CCl4 intoxication in mouse liver, is not a critical mediator of hepatotoxicity. Arch Toxicol 88(6):1267–1280

Elkon R, Linhart C, Sharan R, Shamir R, Shiloh Y (2003) Genome-wide in silico identification of transcriptional regulators controlling the cell cycle in human cells. Genome Res 13:773–780

Ellinger-Ziegelbauer H, Gmuender H, Bandenburg A, Ahr HJ (2008) Prediction of a carcinogenic potential of rat hepatocarcinogens using toxicogenomics analysis of short-term in vivo studies. Mutat Res 637(1–2):23–39

Fijten RR, Jennen DG, van Delft JH (2013) Pathways for ligand activated nuclear receptors to unravel the genomic responses induced by hepatotoxicants. Curr Drug Metab 14(10):1022–1028

Froguel P, Zouali H, Sun F, Velho G, Fukumoto H, Passa P, Cohen D (1991) CA repeat polymorphism in the glucose transporter GLUT 2 gene. Nucleic Acids Res 19(13):3754

Gåfvels M, Olin M, Chowdhary BP, Raudsepp T, Andersson U, Persson B, Jansson M, Björkhem I, Eggertsen G (1999) Structure and chromosomal assignment of the sterol 12alpha-hydroxylase gene (CYP8B1) in human and mouse: eukaryotic cytochrome P-450 gene devoid of introns. Genomics 56(2):184–196

Godoy P (2013) Recent advances in 2D and 3D in vitro systems using primary hepatocytes, alternative hepatocyte sources and non-parenchymal liver cells and their use in investigating mechanisms of hepatotoxicity, cell signaling and ADME. Arch Toxicol 87:1315–1530

Godoy P, Hengstler JG, Ilkavets I, Meyer C, Bachmann A, Müller A, Tuschl G, Mueller SO, Dooley S (2009) Extracellular matrix modulates sensitivity of hepatocytes to fibroblastoid dedifferentiation and transforming growth factor beta-induced apoptosis. Hepatology 49(6):2031–2043

Godoy P, Lakkapamu S, Schug M, Bauer A, Stewart JD, Bedawi E, Hammad S, Amin J, Marchan R, Schormann W, Maccoux L, von Recklinghausen I, Reif R, Hengstler JG (2010a) Dexamethasone-dependent versus -independent markers of epithelial to mesenchymal transition in primary hepatocytes. Biol Chem 391(1):73–83

Godoy P, Schug M, Bauer A, Hengstler JG (2010b) Reversible manipulation of apoptosis sensitivity in cultured hepatocytes by matrix-mediated manipulation of signaling activities. Methods Mol Biol 640:139–155

Harbron C, Chang KM, South MC (2007) RefPlus: an R package extending the RMA Algorithm. Bioinformatics 23(18):2493–2494 Epub 2007 Jul 10

Heise T, Schug M, Storm D, Ellinger-Ziegelbauer H, Ahr HJ, Hellwig B, Rahnenfuhrer J, Ghallab A, Guenther G, Sisnaiske J, Reif R, Godoy P, Mielke H, Gundert-Remy U, Lampen A, Oberemm A, Hengstler JG (2012) In vitro–in vivo correlation of gene expression alterations induced by liver carcinogens. Curr Med Chem 19(11):1721–1730

Hewitt NJ, Lechón MJ, Houston JB, Hallifax D, Brown HS, Maurel P, Kenna JG, Gustavsson L, Lohmann C, Skonberg C, Guillouzo A, Tuschl G, Li AP, LeCluyse E, Groothuis GM, Hengstler JG (2007) Primary hepatocytes: current understanding of the regulation of metabolic enzymes and transporter proteins, and pharmaceutical practice for the use of hepatocytes in metabolism, enzyme induction, transporter, clearance, and hepatotoxicity studies. Drug Metab Rev 39(1):159–234

Hirode M, Omura K, Kiyosawa N, Uehara T, Shimuzu T, Ono A, Miyagishima T, Nagao T, Ohno Y, Urushidani T (2009) Gene expression profiling in rat liver treated with various hepatotoxic-compounds inducing coagulopathy. J Toxicol Sci 34(3):281–293

Hoehme S, Hengstler JG, Brulport M, Schäfer M, Bauer A, Gebhardt R, Drasdo D (2007) Mathematical modelling of liver regeneration after intoxication with CCl(4). Chem Biol Interact 168(1):74–93

Hoehme S, Brulport M, Bauer A, Bedawy E, Schormann W, Hermes M, Puppe V, Gebhardt R, Zellmer S, Schwarz M, Bockamp E, Timmel T, Hengstler JG, Drasdo D (2010) Prediction and validation of cell alignment along microvessels as order principle to restore tissue architecture in liver regeneration. Proc Natl Acad Sci USA 107(23):10371–10376

Huang WY, Li ZG, Rus H, Wang X, Jose PA, Chen SY (2009) RGC-32 mediates transforming growth factor-beta-induced epithelial-mesenchymal transition in human renal proximal tubular cells. J Biol Chem 284(14):9426–9432

Kamiya A, Inoue Y, Gonzalez FJ (2003) Role of the hepatocyte nuclear factor 4alpha in control of the pregnane X receptor during fetal liver development. Hepatology 37(6):1375–1384

Kimura Y, Nishimura FT, Abe S, Fukunaga T, Tanii H, Saijoh K (2009) Polymorphisms in the promoter region of the human class II alcohol dehydrogenase (ADH4) gene affect both transcriptional activity and ethanol metabolism in Japanese subjects. J Toxicol Sci 34(1):89–97

Kiyosawa N, Ando Y, Watanabe K, Niino N, Manabe S, Yamoto T (2009) Scoring multiple toxicological endpoints using a toxicogenomic database. Toxicol Lett 188(2):91–97

Krug AK, Kolde R, Gaspar JA, Rempel E, Balmer NV, Meganathan K, Vojnits K, Baquié M, Waldmann T, Ensenat-Waser R, Jagtap S, Evans RM, Julien S, Peterson H, Zagoura D, Kadereit S, Gerhard D, Sotiriadou I, Heke M, Natarajan K, Henry M, Winkler J, Marchan R, Stoppini L, Bosgra S, Westerhout J, Verwei M, Vilo J, Kortenkamp A, Hescheler J, Hothorn L, Bremer S, van Thriel C, Krause KH, Hengstler JG, Rahnenführer J, Leist M, Sachinidis A (2013) Human embryonic stem cell-derived test systems for developmental neurotoxicity: a transcriptomics approach. Arch Toxicol 87(1):123–143

Lake AD, Novak P, Fisher CD, Jackson JP, Hardwick RN, Billheimer DD, Klimecki WT, Cherrington NJ (2011) Analysis of global and absorption, distribution, metabolism, and elimination gene expression in the progressive stages of human nonalcoholic fatty liver disease. Drug Metab Dispos 39:1954–1960

Lin M, Napoli JL (2000) cDNA cloning and expression of a human aldehyde dehydrogenase (ALDH) active with 9-cis-retinal and identification of a rat ortholog, ALDH12. J Biol Chem 275(51):40106–40112

NIBIO (2013) National Institute of Biomedical Innovation. http://toxico.nibio.go.jp. Stand: 13.08.2013

Pang XY, Cheng J, Kim JH, Matsubara T, Krausz KW, Gonzalez FJ (2012) Expression and regulation of human fetal-specific CYP3A7 in mice. Endocrinology 153(3):1453–1463

Pilz AJ, Willer E, Povey S, Abbott CM (1992) The genes coding for phosphoenolpyruvate carboxykinase-1 (PCK1) and neuronal nicotinic acetylcholine receptor alpha 4 subunit (CHRNA4) map to human chromosome 20, extending the known region of homology with mouse chromosome 2. Ann Hum Genet 56(Pt 4):289–293

Saigusa K, Imoto I, Tanikawa C, Aoyagi M, Ohno K, Nakamura Y, Inazawa J (2007) RGC32, a novel p53-inducible gene, is located on centrosomes during mitosis and results in G2/M arrest. Oncogene 26(8):1110–1121

Schliess F, Hoehme S, Henkel SG, Ghallab A, Driesch D, Böttger J, Guthke R, Pfaff M, Hengstler JG, Gebhardt R, Häussinger D, Drasdo D, Zellmer S (2014) Integrated metabolic spatial-temporal model for the prediction of ammonia detoxification during liver damage and regeneration. Hepatology. doi:10.1002/hep.27136 [Epub ahead of print]

Schug M, Stöber R, Heise T, Mielke H, Gundert-Remy U, Godoy P, Reif R, Blaszkewicz M, Ellinger-Ziegelbauer H, Ahr HJ, Selinski S, Günther G, Marchan R, Blaszkewicz M, Sachinidis A, Nüssler A, Oberemm A, Hengstler JG (2013) Pharmacokinetics explain in vivo/in vitro discrepancies of carcinogen-induced gene expression alterations in rat liver and cultivated hepatocytes. Arch Toxicol 87(2):337–345

Seglen PO (1976) Preparation of isolated rat liver cells. Methods Cell Biol 13:29–83

Simmer JP, Kelly RE, Rinker AG Jr, Scully JL, Evans DR (1990) Mammalian carbamyl phosphate synthetase (CPS). DNA sequence and evolution of the CPS domain of the Syrian hamster multifunctional protein CAD. J Biol Chem 265(18):10395–10402

Tsuruoka S, Ishibashi K, Yamamoto H, Wakaumi M, Suzuki M, Schwartz GJ, Imai M, Fujimura A (2002) Functional analysis of ABCA8, a new drug transporter. Biochem Biophys Res Commun 298(1):41–45

Tung EK, Mak CK, Fatima S, Lo RC, Zhao H, Zhang C, Dai H, Poon RT, Yuen MF, Lai CL et al (2011) Clinicopathological and prognostic significance of serum and tissue Dickkopf-1 levels in human hepatocellular carcinoma. Liver Int 31:1494–1504

Uehara T, Minowa Y, Morikawa Y, Kondo C, Maruyama T, Kato I, Nakatsu N, Igarashi Y, Ono A, Hayashi H, Mitsumori K, Yamada H, Ohno Y, Urushidani T (2011) Prediction model of potential hepatocarcinogenicity of rat hepatocarcinogens using a large-scale toxicogenomics database. Toxicol Appl Pharmacol 255(3):297–306

Ulitsky I, Maron-Katz A, Shavit S, Sagir D, Linhart C, Elkon R, Tanay A, Sharan R, Shiloh Y, Shamir R (2010) Expander: from expression microarrays to networks and functions. Nat Protoc 5:303–322

Väremo L, Nielsen J, Nookaew I (2013) Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. Nucleic Acids Res 4:4378–4391

Waldmann T, Rempel E, Balmer NV, König A, Kolde R, Gaspar JA, Henry M, Hescheler J, Sachinidis A, Rahnenführer J, Hengstler JG, Leist M (2014) Design principles of concentration-dependent transcriptome deviations in drug-exposed differentiating stem cells. Chem Res Toxicol 27(3):408–420

Watt AJ, Garrison WD, Duncan SA (2003) HNF4: a central regulator of hepatocyte differentiation and function. Hepatology 37(6):1249–1253

Weng MK, Natarajan K, Scholz D, Ivanova VN, Sachinidis A, Hengstler JG, Waldmann T, Leist M (2014) Lineage-specific regulation of epigenetic modifier genes in human liver and brain. PLoS ONE 9(7):e102035

Wu J, Wang C, Li S, Li S, Wang W, Li J, Chi Y, Yang H, Kong X, Zhou Y, Dong C, Wang F, Xu G, Yang J, Gustafsson JÅ, Guan Y (2013) Thyroid hormone-responsive SPOT 14 homolog promotes hepatic lipogenesis, and its expression is regulated by liver X receptor α through a sterol regulatory element-binding protein 1c-dependent mechanism in mice. Hepatology 58(2):617–628

Xing L, Wu L, Liu Y, Ai N, Lu X, Fan X (2014) LTMap: a web server for assessing the potential liver toxicity by genome-wide transcriptional expression data. J Appl Toxicol 34(7):805–809

Zellmer S, Schmidt-Heck W, Godoy P, Weng H, Meyer C, Lehmann T, Sparna T, Schormann W, Hammad S, Kreutz C, Timmer J, von Weizsäcker F, Thürmann PA, Merfort I, Guthke R, Dooley S, Hengstler JG, Gebhardt R (2010) Transcription factors ETF, E2F, and SP-1 are involved in cytokine-independent proliferation of murine hepatocytes. Hepatology 52(6):2127–2136

Zhang JD, Berntenis N, Roth A, Ebeling M (2014) Data mining reveals a network of early-response genes as a consensus signature of drug-induced in vitro and in vivo toxicity. Pharmacogenomics J 14(3):208–216