Toxicogenomics directory of rat hepatotoxicants in vivo and in cultivated hepatocytes

Marianna Grinberg¹ · Regina M. Stöber² · Wiebke Albrecht² · Karolina Edlund² · Markus Schug² · Patricio Godoy² · Cristina Cadenas² · Rosemarie Marchan² · Alfonso Lampen³ · Albert Braeuning³ · Thorsten Buhrke³ · Marcel Leist⁴ · Axel Oberemm³ · Birte Hellwig¹ · Hennicke Kamp⁵ · Iain Gardner⁶ · Sylvia Escher⁷ · Olivier Taboureau^{8,9} · Alejandro Aguayo-Orozco⁸ · Agapios Sachinidis¹⁰ · Heidrun Ellinger-Ziegelbauer¹¹ · Jörg Rahnenführer¹ · Jan G. Hengstler²

Abstract

Transcriptomics is developing into an invaluable tool in toxicology. The aim of this study was, using a transcriptomics approach, to identify genes that respond similar to many different chemicals (including drugs and industrial compounds) in both rat liver in vivo and in cultivated hepatocytes. For this purpose, we analyzed Affymetrix microarray expression data from 162 compounds that were previously tested in a concentration-dependent manner in rat livers in vivo and in rat hepatocytes cultivated in sandwich culture. These data were obtained from the Japanese Toxicogenomics Project (TGP) and North Rhine-Westphalian (NRW) data sets, which represent 138 and 29 compounds, respectively, and have only 5 compounds in common between them. The in vitro gene expression data from the NRW data set were generated in the present study, while TGP is publicly available. For each of the data sets, the overlap between up- or down-regulated genes in vitro and in vivo was identified, and named in vitro–in vivo consensus genes. Interestingly, the in vivo–in vitro consensus genes overlapped to a remarkable extent between both data sets, and were 21-times (upregulated genes) or 12-times (down-regulated genes) enriched compared to random expectation. Finally, the genes in the TGP and NRW overlap were used to identify the upregulated genes with the highest compound coverage, resulting in a seven-gene set of *Cyp1a1*, *Ugt2b1*, *Cdkn1a*, *Mdm2*, *Aldh1a1*, *Cyp4a3*, and *Ehhadh*. This seven-gene set was then successfully tested with structural analogues of valproic acid that are not present in the TGP and NRW data sets. In conclusion, the seven-gene set identified in the present study responds similarly in vitro and in vivo to a wide range of different chemicals. Despite these promising results with the seven-gene set, transcriptomics with cultivated rat hepatocytes remains a challenge, because in general many genes are up- or downregulated by in vitro culture per se, respond differently to test compounds in vitro and in vivo, and/or show higher variability in the in vitro system compared to the corresponding in vivo data.

Keywords Rat liver · Cultivated hepatocytes · Hepatotoxicity · Ranking analysis · Batch effect correction · Bioinformatics

Introduction

Transcriptomics is a frequently applied tool in toxicology (Godoy et al. 2013; Ellinger-Ziegelbauer et al. 2011; Krug et al. 2013; Lohr et al. 2015; Reif 2015), and has been used to elucidate specific gene expression patterns that can then be associated with mechanisms of action and histopathological changes (Ippolito et al. 2016; Stoeber 2015; Rempel et al. 2015; De Abrew et al. 2015; Stemmer et al. 2007). Recently, studies have been performed to understand the principles governing transcriptome changes from safe to hazardous drug concentrations (Waldmann et al. 2014).

Marianna Grinberg and Regina M. Stöber shared first authorship.

Jörg Rahnenführer and Jan G. Hengstler shared senior authorship.

✉ Jan G. Hengstler
hengstler@ifado.de

Concentration-dependent analyses in human embryonic stem cells differentiating to neuroectoderm have shown that at least three concentration ranges should be differentiated: (1) a range of tolerance, where no gene expression changes occur; (2) a range of expression deregulation, where genes are up- or downregulated but no cytotoxic effects occur; and (3) a cytotoxic concentration range where in addition to gene expression alterations, cell death events are also induced. Often, concentrations that cause expression deregulation are associated with adverse effects. For example, valproic acid has been shown to cause gene expression deregulation without cytotoxicity at 150–550 μ M in cultivated neuronal precursor cells, a similar concentration range that leads to an increased risk of teratogenicity measured in an in vitro embryonic stem cell differentiation test (Waldmann et al. 2014).

To differentiate between safe and hazardous drug concentrations, it is important to identify the concentration range that causes deregulation of gene expression. This is a precondition to study whether deregulated genes are responsible for adverse effects. However, it is relatively expensive to perform genome-wide expression analyses for this task. Therefore, considering that gene expression alterations usually occur in clusters of correlated genes (Godoy et al. 2015, 2016), a small set of genes may be sufficient to identify the concentration dependence of gene expression deregulations. These genes may then be analyzed by qRT-PCR or reporter constructs instead of performing more expensive RNAseq or microarray experiments.

Recently, a genome-wide expression study on cultivated human hepatocytes exposed to 143 chemicals has reported a stereotypical stress response of gene expression, indicating that the same genes were up- or downregulated by many different compounds (Grinberg et al. 2014). Many of these stereotypical stress response genes also overlap with genes deregulated in human liver diseases, such as cirrhosis, hepatocellular cancer, and non-alcoholic fatty liver disease (Grinberg et al. 2014).

Such stereotypical gene expression responses have so far been studied in cultivated human hepatocytes only. Currently, there are no studies available that have investigated similar gene expression responses in cultivated rat hepatocytes. One challenge working with rat hepatocytes is that the genes deregulated after exposure of cultivated rat hepatocytes may show relatively large expression differences compared to rats treated in vivo (Schug et al. 2013). Not only the magnitude of expression deregulation may differ but also completely different genes may be affected. Therefore, the aim of this study was to identify a consensus signature of genes that are deregulated by many compounds. Moreover, these ‘consensus genes’ should respond similarly in vivo and in hepatocytes in vitro which could support translation between in vivo and in vitro data. For this purpose, we

generated Affymetrix microarray data from rat hepatocytes treated with 29 compounds that were previously tested in rats in vivo (Ellinger-Ziegelbauer et al. 2008). These compounds comprise genotoxic liver carcinogens, non-genotoxic liver carcinogens and non-carcinogenic substances. The in vivo and in vitro data of the 29 compounds have been collected as the NRW (North Rhine-Westphalian) dataset (Fig. 1). This allowed us to identify genes overlapping in vitro and in vivo (in vitro/in vivo overlap) (Fig. 1). The in vitro/in vivo overlap was also obtained for a different set of compounds for which public Affymetrix data were available, the Japanese Toxicogenomics Project (TGP) dataset (Fig. 1). Interestingly, a relatively high degree of overlap (compared to the randomly expected overlap) was obtained between the two databases although only five chemicals were present in common. These results demonstrate that as observed in human hepatocytes, a stereotypical expression response can also be induced in rat hepatocytes to a wide range of chemicals with different mechanisms of action. Based on the consensus genes of the NRW and the TGP data sets, a seven-gene signature was established and further compounds (not present in NRW or TGP) were tested. All seven additional compounds showed the expected deregulations of the seven-gene signature.

Materials and methods

Statistical analysis

The download and preprocessing of TGP data set was performed as previously described (Grinberg et al. 2014). Data preprocessing and all subsequent analyses were performed using the statistical programming language R, version 3.0.1 (R Development Core Team 2013). Gene microarray data of the already published in vivo data of the NRW data set (GSE68110); (Ellinger-Ziegelbauer et al. 2008) and the here-generated gene microarray data in vitro (GSE119933; see below) were processed by the same pipeline. For 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 on each microarray (Irizarry et al. 2003). 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. 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

Fig. 1 Study design. The study is based on two genome-wide data sets (NRW and TGP). In both data sets, test compounds were studied after oral administration to rats and analysis of liver tissue (in vivo), as well as after exposure of cultivated rat hepatocytes (in vitro). First, genes up- or down-regulated by the individual compounds in vitro and in vivo (in vitro–in vivo overlap) were identified in each data set. Next, the overlap between the two data sets was determined. The latter overlap was used for a ranking approach to identify gene sets that

cover the highest possible number of compounds, also named ‘top-ranking genes’. Finally, cultivated rat hepatocytes were incubated concentration dependently with independent compounds (validation compounds) not included in the previous two data sets (NRW and TGP) and the expression of seven ‘top-ranking genes’ was analyzed. The corresponding gene lists are available in the supplement (Tab. S1–12)

(Grinberg et al. 2014). PCA was performed on the basis of the 100 top-ranking genes with the highest fold change (absolute values) across all compounds. Numbers of significantly up- or down-regulated genes for the individual test compounds were obtained after adjustment by the method of Benjamini and Hochberg at a p value ≤ 0.01 (adj.p. Benjamini–Hochberg (BH) ≤ 0.01). In the corresponding plots, genes with adj.p.BH ≤ 0.01 with fold changes (FC) ≥ 1.5 , 2.0 and 3.0 are differentiated using color-coded bar plots. The selection value (SV) used in this study is a method to differentiate between stereotypic and compound-specific gene expression responses (Grinberg et al. 2014). Genes that are deregulated by many compounds reflect a stereotypic

response, in contrast to a 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 a significant expression change. Selection value 3 (SV3) then yields a list of genes deregulated in the same direction by at least three compounds. Euclidean distances were calculated to quantify the similarity of replicates as well as pairs of controls and compound-induced samples. For this purpose, Euclidean distances between all pairs of triplicates (duplicates) and between all pairs of compound-exposed samples and controls were calculated and compared with each other. Exclusivity analysis according to a method recently described by Grinberg

et al. (2014) was applied to first determine the 100 strongest upregulated (or downregulated) genes across all compounds. Next, these genes were assigned to the compound with the most extreme fold change. Finally, the plots indicate the numbers of such ‘match winning events’ for each compound. Analysis of overlap ratios is a recently established method to determine whether the overlap of two sets of genes is higher than randomly expected (Shinde et al. 2017). For instance, an overlap ratio of three means that two sets of genes have threefold more genes in common than randomly expected. Genes and gene combinations with the highest coverage were identified by the following method: first, the gene was identified that is deregulated by most of the test compounds; subsequently, the next ranking gene was added so that both are deregulated by as many compounds as possible, and so on.

Establishment of the NRW data set

The NRW data set consists of gene microarray data from 29 compounds that were tested in rat liver *in vivo* after the exposure of male Wistar rats, and *in vitro* using cultivated hepatocytes from male Wistar rats (*in vitro* data generated in the present study: GSE119933). The matrix of the test compounds is given in Table S1 (*in vivo*) and Table S2 (*in vitro*), including information on the three concentrations used *in vitro*, the number of samples obtained from different rats for each concentration, and the class of the compounds. Incubation of the cultivated rat hepatocytes with the test compounds was performed for 24 h. Rat hepatocytes were cultivated as sandwich cultures as described below. Cultivation was performed using six-well dishes where three wells were incubated with the same concentration of the test compound and the remaining three wells were used for solvent controls. Moreover, each compound was analyzed following treatment of hepatocytes from three different rats, isolated on different experimental days (three biological replicates). Affymetrix gene chip analysis of *in vitro*-cultivated hepatocytes was performed as described in Krug et al. (2013) with modifications. Briefly, samples from the six-well dishes were collected and stored in RNeasy Protect reagent from Qiagen until isolation of RNA. The RNA was quantified using a NanoDrop N-1000 spectrophotometer, and the integrity of RNA was confirmed with a standard sense automated gel electrophoresis system. Samples were used for transcriptional profiling only when their RNA quality indicator (RQI) number was > 8. First-strand cDNA was synthesized from 100 ng total RNA using an oligo-dT primer with an attached T7 promoter sequence, followed by the complementary second strand. The double-stranded cDNA molecule was used for *in vitro* transcription (IVT, standard Affymetrix procedure) using Genechip 3' IVT Express Kit. During synthesis of the aRNA (amplified RNA, also commonly referred to

as cRNA), a biotinylated nucleotide analogue was incorporated, which serves as a label for the message. After amplification, aRNA was purified with magnetic beads and 15 µg of aRNA was fragmented with the fragmentation buffer as per the manufacturer's instructions. Then 12.5 µg fragmented aRNA was hybridized with Affymetrix Rat Genome 230 2.0 Arrays as per the manufacturer's instructions. The chips were placed in a GeneChip Hybridization Oven-645 for 16 h at 60 rpm and 45 °C. For staining and washing, Affymetrix HWS kits were used on a Genechip Fluidics Station-450. For scanning, the Affymetrix GeneChip Scanner-3000-7G was used, and the image and quality control assessments were performed with Affymetrix GCOS software. All reagents and instruments were acquired from Affymetrix (Affymetrix, Santa Clara, CA, USA). The generated CEL files were used for further statistical analysis. The authors declare that microarray data were produced according to MIAME guidelines and will be deposited in ArrayExpress upon acceptance of the manuscript. Doses and exposure periods *in vivo* are available in Table S1 and the corresponding gene microarray data have already been published in Ellinger-Ziegelbauer et al. (2008).

Further experimental procedures

Isolation and cultivation of primary rat hepatocytes was performed according to a published standard operation procedure (Godoy et al. 2013; Schug et al. 2013). Male Wistar rats with a body weight of 220–300 g were purchased from Charles River (Sulzfeld, Germany). The animals had free access to food (sniff, Soest, Germany) and water and were kept under controlled temperature (18–26 °C), humidity (30–70%) and lighting (12 h light/dark circle). Prior to any experimental procedure, the animals were acclimated for a minimum of 6 days. This study was approved by the local committee for the welfare of experimental animals and was performed in accordance with national legislation. RNA was extracted from cultivated primary hepatocytes as described by Heise et al. (2012). Quantitative real-time PCR (qRT-PCR) was performed as described in Grinberg et al. (2014). Briefly, the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Darmstadt, Germany) was used to reversely transcribe RNA into cDNA. qRT-PCR with TaqMan probes was performed on the ABI 7500 Fast Real-Time PCR system (Applied Biosystems) to determine gene expression levels and the following primers were used: Cytochrome P450 family 1 subfamily a polypeptide 1, Cyp1a1: Rn00487218_m1, UDP Glucuronosyltransferase 2 family polypeptide B1, Ugt2b1: Rn00756519_m1, Mouse Double Minute 2 Homolog Protooncogene E3 ubiquitin protein ligase, Mdm2: Rn01502814_m1, Aldehyde dehydrogenase 1 family member A1, Aldh1a1: Rn00755484_m1, Cytochrome P450 family 4 subfamily a, polypeptide 3,

Cyp4a3: Rn04224033_u1, Cyclin-dependent kinase inhibitor 1A, Cdkn1a: Rn00589996_m1, Enoyl-CoA Hydratase And 3-Hydroxyacyl CoA Dehydrogenase, Ehhadh: Rn00592368_m1, Cyclin-Dependent Kinase Inhibitor 1C, Cdkn1c: Rn01502044_g1, and Peptidylprolyl isomerase A, PPIA: Rn00690933_m1. Twenty-five nanogram 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 that were cultivated on the same six-well plate as the treated cells were used as controls. Three biological replicates (hepatocytes isolated from three different rats; each isolation was performed at a different day) were analyzed.

Results

Data structure of the NRW data set

The NRW data set is comprised of 30 compounds that were previously tested in rats in vivo (Ellinger-Ziegelbauer et al. 2008; Table S1). Of these 30 compounds, 29 were tested in vitro in the present study using sandwich cultures of rat hepatocytes (Table S2). Male Wistar rats were used for both the in vivo and in vitro experiments. Exposure periods in vivo ranged between 24 h and 14 days with the exception of acetaminophen that was additionally tested after 6 and 12 h (Table S1). For the in vitro experiments, the 24-h time point was analyzed using three concentrations, with the highest representing the EC20 (Table S2). All samples from the in vitro and in vivo series were analyzed using Affymetrix microarrays. An overview of the gene expression for the individual replicates of all samples was obtained by principal component analysis (PCA) (Fig. 2). The gene expression changes obtained from the in vitro (green) and in vivo (orange) models clustered to different regions of the PCA plot (Fig. 2). At the beginning of the incubation period (green triangles), the in vitro samples clustered separately from the samples after 24-h compound exposure (green circles and squares). This agrees with the well-known phenomenon of spontaneous expression changes when hepatocytes are isolated and cultivated (Godoy et al. 2015, 2016; Zellmer et al. 2010). It also underlines the importance of time- and batch (experiment)-matched controls. Moreover, the in vitro samples (ctr and exp) appeared in two clusters separated along PC2. In vivo, controls and compound-exposed samples did not cluster separately when all samples were included in the PCA (Fig. 2).

To visualize the influence of the different test compounds, the in vitro (Suppl. Fig. S1) and in vivo (Suppl. Fig. S2) data were further analyzed separately. The degree of variability was illustrated by connecting lines between replicates

Fig. 2 Principal component analysis of gene expression data obtained from cultivated primary rat hepatocytes (in vitro) and from rat liver samples (in vivo) after incubation with 29 (in vitro) and 30 (in vivo) chemicals. The PCA plot is based on the 100 probe sets with highest variance and was generated to display the transcriptome data structure across in vivo and in vitro replicates. Each point represents one experiment (data from one microarray). Cultivated hepatocytes were harvested at the beginning of the exposure period (T0) and after 24-h exposure to test compounds (Exp) and solvent (Ctrl). In vivo, rat livers were analyzed 24 h after administration of test compounds (Expr) or solvent (controls)

(Suppl. Figs. S1B and S2B). In vivo, the distances between replicates (different rats) were relatively small and similar in all pairs of replicates. In contrast, the distances between replicates in vitro (hepatocytes from different rats cultivated on different days) showed a much higher variability. Analysis of Euclidean distances between all pairs of replicates illustrates the higher variability in vitro, which ranges from very small to very large distances; whereas a much narrower range was obtained in vivo (Fig. 3). Next, to illustrate the influence of the test compounds, mean values of the replicates were calculated (Suppl. Figs. S1C and S2C) and connecting lines between compound-exposed samples and the corresponding controls were introduced (Suppl. Figs. S1D and S2D). The Euclidean distances between all pairs of compound-exposed samples and corresponding controls in vitro were larger for the high concentrations compared to the middle and low concentrations (Fig. 4). Finally, controls were subtracted from the corresponding compound-exposed samples (Suppl. Figs. S1E and S2E). In vitro, the controls from the same six-well dishes (on each dish three wells were compound exposed while three further wells were solvent controls) were exclusively used as a reference. Subtraction of the six-well dish-matched controls resulted in only a single cluster

Fig. 3 Reproducibility between replicates. Boxplots of the pairwise Euclidean distances between replicates for all test conditions. The gray and the green points illustrate the distances of the controls and exposed samples, respectively. The grey lines at the right side of the boxes represent means and standard deviations of both controls and exposed samples. (Color figure online)

Fig. 4 Reproducibility between control and compound-exposed samples. Boxplots of the Euclidean distance between all pairs of triplicates for all test conditions. ‘Low’, ‘middle’, and ‘high’ indicate the distances between controls and exposed samples in vitro. 1–14 days indicate the exposure periods in vivo

(Suppl. Fig. S1E; left panels) compared to the scenario with several clusters before subtraction (Suppl. Fig. S1A), which probably resulted from experimental batch effects.

The in vivo experiments included various exposure periods (6 h–14 days). Because of the large number of rats, treatment of animals was performed in different experimental series. To decide how to subtract controls, a variance analysis was performed, which identified the parameters ‘experimental series’ and ‘exposure period’ as strong factors of influence. Therefore, ‘experimental series’ and ‘exposure period’-matched controls were subtracted for analysis of the in vivo data (Suppl. Fig. S2E, right panels).

Differentially expressed genes per compound

Based on the above-described assignment of controls (six-well dish matched for in vitro data; matched for exposure period and experimental series in vivo), the number of differentially expressed genes per test compound was identified after false discovery adjustment for multiple testing. The number of up- or downregulated genes varied widely for the individual test compounds (Fig. 5a). Compounds with less than 30 up- or down-regulated genes were named ‘low-profile compounds’ (eight compounds); correspondingly, compounds deregulating 30 or more genes are defined as ‘high-profile compounds’ (Table S13). A similar result was obtained in the exclusivity analysis (Fig. 6a). This analysis identifies the 100 strongest up- or downregulated genes across all compounds (most extreme fold changes), followed by the assignment of each of the 100 top genes to the compound with the highest fold change (compounds are ‘match-winners’ per gene). With the low-profile compounds only one gene was observed for few compounds in the exclusivity analysis (Fig. 6a) (except for AFB1, ETH and PRAZ—all low-profile compounds scored with zero), which means that the compounds that deregulate few genes also cause smaller expression changes (in relation to that caused by compounds that deregulate more genes) for those (few) genes. For the in vivo data, no clear relationship between exposure period and the number of up- or down-regulated genes was observed (Fig. 5b). Approximately 13% of the substances tested in vivo were low-profile compounds (Fig. 5b), which scored with one (CFX, MDA) to zero genes in the exclusivity analysis (Fig. 6b). Combining the in vitro and in vivo low-profile compounds resulted in 11 individual compounds.

Concentration dependency of gene deregulation

Concentration dependency of the in vitro data was illustrated using Venn diagrams (representative examples in Fig. 7). CFX and 2-NF are examples of two compounds with a plausible concentration progression, where most of the genes up- or downregulated at a lower concentration were also

Fig. 5 Number of significantly up- or down-regulated genes for the individual test compounds. **a** In vitro; **b** in vivo. The x-axis lists all chemicals that were tested at the indicated concentration (illustrated by green color) or exposure period (illustrated by orange color).

The y-axis gives the number of up- or downregulated genes with at least 1.5-, 2.0-, and 3.0-fold change and adjusted p value ≤ 0.01 corresponding to the light, middle and dark colors, respectively. (Color figure online)

Fig. 6 Exclusivity analysis of the deregulated genes. This analysis first determines the 100 strongest up- or downregulated genes across all compounds. Next, these genes were assigned to the compound with the most extreme fold change (only genes with $\text{adj.p.BH} \leq 0.01$ are considered in the ranking). **a** In vitro (dark green-colored bar plot). **b** In vivo (dark orange-colored bar plot); the corresponding light-colored bar plots indicate whether the top-ranking genes meet the following criteria: > 1.5 -fold with $\text{adj.p.BH} \leq 0.01$. The x-axis lists only those compounds that contributed at least once to the strongest genes in the ranking of all probe sets (**a** 31,099 and **b** 15,923 probe sets). (Color figure online)

Fig. 7 Concentration dependency of the in vitro data shown for representative examples. Orange-colored Venn diagrams show the overlap between genes that are upregulated at the low, middle and high

concentrations (> 1.5 -fold with $\text{adj.p.BH} \leq 0.01$); green-colored Venn diagrams summarize the downregulated genes (> 1.5 -fold downregulation with $\text{adj.p.BH} \leq 0.01$). (Color figure online)

deregulated at higher concentrations (Fig. 7). A frequently observed constellation was that only the highest tested concentration resulted in strong gene expression responses, as illustrated for AIAI (Fig. 7). However, it is important to consider that implausible concentration progressions were also observed. An extreme example is DEHA, where a large number of genes were deregulated upon treatment with the middle but not the highest concentration (Fig. 7). Experimental errors or sample mix-up may represent possible explanations. Because of these limitations in the data quality, it is important to apply an evaluation strategy that searches for commonalities within data sets and generates conclusions that do not depend on individual data subsets of specific compounds. For this purpose, the selection value concept was recently introduced (Grinberg et al. 2014).

A gene with a selection value of n indicates that at least n of the tested compounds up- or downregulate this gene at the indicated test condition. Therefore, working only with, for example, selection value 3 (SV3) genes improves the robustness of the analysis. A prerequisite for the success of the selection value strategy is that several compounds act by similar mechanisms, which leads to deregulation of overlapping gene sets. If this is the case, the selection value helps to exclude random expression changes due to, e.g., technical issues or sample mix-up and to focus on common changes induced by many compounds. An alternative would

be to simply exclude problematic data; however, it may become very difficult to define adequate exclusion criteria. Therefore, using the selection value strategy is an easy to handle technique to study if data sets are of low quality but nevertheless contain robust expression changes. Selection value analysis of all compounds tested in vitro and in vivo showed that the number of SV3 genes was still higher than 100 for up- or downregulated genes at the highest tested in vitro concentration (Fig. 8) and was, therefore, used for further analysis.

In vitro–in vivo overlap in the data set

In subsequent analyses, we studied whether the overlap between in vitro- and in vivo-deregulated SV3 genes is higher than randomly expected. For this purpose, the recently established overlap ratio (Shinde et al. 2017) was calculated, which considers the number of genes in both evaluated groups (here, genes deregulated in vivo and in vitro) and the total number of analyzed genes. An overlap ratio of 1.0 corresponds to a randomly expected result, while a ratio of, for example, 2.0 indicates that genes were two times enriched compared to random expectation. The significance of overrepresentation was calculated by Fisher's exact test. The overlap ratio of the upregulated in vitro and in vivo genes was 5.9 ($p < 0.001$) and 2.8 ($p < 0.01$)

Fig. 8 Selection values of the up- and downregulated genes (>11.5 -fold down- or upregulation and $\text{adj.p.BH} \leq 0.01$). A selection value of, e.g., three, means that at least three compounds upregulate the indicated genes at the specific test condition. **a** In vitro (green-colored bar plots); **b** in vivo (orange-colored bar plots). (Color figure online)

for the up- and downregulated genes, respectively (Fig. 9). Although the in vitro–in vivo overlap was clearly more than random, it should be considered that the majority of genes were nevertheless in the non-overlapping region, indicating substantial differences between the in vitro and in vivo situation (Fig. 9). All genes that were identified in the overlap of the in vitro and in vivo experiments are listed in Table S11, together with the information on the number of compounds that deregulated each individual gene.

Data structure of the TGP dataset

TGP (TG-GATES) is a publicly available genome-wide data set, which includes data from experiments that were conducted in rats in vivo, and in cultivated hepatocytes testing the same compounds. An overview of the tested compounds and test conditions is given in Table S3 (in vivo) and Table S4 (in vitro). A similar analysis pipeline that was

done for the NRW data set was also performed for the TGP data (Suppl. Figs. S3–S6). In addition, the selection value concept was applied as similarly done for the NRW data set (Suppl. Fig. S6), and genes deregulated by at least three compounds were included into the in vitro–in vivo overlap analysis (Fig. 10a). Although the in vitro–in vivo overlap ratios were smaller compared to the NRW data, they still were highly significant with p values smaller than 0.001. The identified in vitro–in vivo overlap genes are listed in Table S10.

Identification of a consensus gene set

Both the NRW and the TGP datasets had significantly overlapping SV3 genes that were influenced by the test compounds in vitro and in vivo, which provided the opportunity to compare the genes obtained from both datasets (strategy in Fig. 1). Twenty-three of the 77 upregulated and 22 of the

Fig. 9 Overlap of genes deregulated by the same compounds *in vitro* and *in vivo*. **a** Principle of the overlap ratio. **b** Venn diagrams showing the overlap between the *in vitro* and *in vivo* consensus SV3 genes of the NRW dataset. The consensus SV3 NRW genes were obtained by the association of the SV3 gene lists of the individual test conditions, i.e., the *in vitro* SV3 consensus list summarizes the genes that are up- or downregulated by at least three of the compounds at the

low, middle or high concentration; the *in vivo* SV3 consensus list summarizes the genes that are up- or downregulated by at least three of the compounds for at least one of the exposure periods (d1, d3, d7 or d14). The overlap ratio was calculated by the formula given in the rightmost column in the table in **a**; 77 (98) SV3 genes are up(down)-regulated in both the rat liver (*in vivo*) and primary rat hepatocytes (*in vitro*) (*in vitro/in vivo* consensus SV3 NRW genes)

98 downregulated *in vitro*–*in vivo* overlap genes identified in the NRW dataset were confirmed in the TGP data (Fig. 10b). This degree of overlap is remarkable considering that different test compounds were studied in both datasets, with only five common substances. The results suggest the existence of a stereotypical or consensus gene expression response in rat hepatocytes *in vitro* and *in vivo* (Fig. 10b).

The subsequent goal was to identify the smallest possible set of genes from the NRW–TGP overlap which, in combination, would ‘cover’ as many compounds as possible (‘high-coverage gene set’). A compound is considered ‘covered’ by a gene set, if it up- or downregulates at least one of the genes of the set (here, one deregulated gene is sufficient, because only genes with selection values of 3 or higher are used). In this gene coverage analysis, only the high-profile compounds were considered. To identify the ‘high-coverage gene set’, the individual gene that is deregulated by most of

the test compounds is initially identified. Subsequently, the next ranking gene is added so that both are deregulated by as many compounds as possible, and so on. These ‘high-coverage genes’ differed between the two data sets. In the NRW data set, only five upregulated genes (Aldh1a1, Gdf15, Cdkn1a, Cyp1a1, and Plk2) were sufficient to cover 16 of the 21 high-profile compounds within this data set (Fig. 11, left panel). Two genes (Aldh1a1 and Gdf15) were sufficient to identify 13 compounds, and the addition of each further gene only allowed for the identification of one additional compound each (Fig. 11, left panel). A similar scenario was obtained for the TGP data set, where only one gene (Cyp1a1) overlapped with the top-ranking genes in the NRW data set (Fig. 11, middle panel). The number of high-profile compounds covered by the top genes in TGP ($n=7$) was lower compared to NRW. Finally, the combined NRW–TGP data sets were analyzed to identify the top-ranking high-coverage

Fig. 10 Overlap analysis of the in vitro and in vivo consensus SV3 genes in TGP and identification of mutual genes in the NRW and TGP data sets. **a** Overlap of the in vitro- and in vivo-consensus SV3 genes for the TGP data set. The same technique was used as for the NRW data set in Fig. 9. **b** Overlap between the consensus SV3 TGP genes (140; 186) and the consensus SV3 NRW genes (77; 98). Orange- and green-colored Venn diagrams show the overlap between

the up- and downregulated genes (SV3 TGP genes: > threefold in vitro and with $\text{adj.p.BH} \leq 0.01$ in vivo vs. consensus SV3 NRW genes: > 1.5-fold and $\text{adj.p.BH} \leq 0.01$); 23 (22) TGP consensus SV3 genes out of 77 (98) NRW consensus SV3 genes are up(down)-regulated in both the rat liver (in vivo) and primary rat hepatocytes (in vitro) and in both datasets (consensus SV3 genes). (Color figure online)

genes (Fig. 11, right panel; Table S8). A mutual feature of all three analyses in Fig. 11 is that the top-ranking upregulated genes in the TGP allowed coverage of a higher number of compounds than the top-ranking downregulated genes.

Analysis of the seven-gene set with independent compounds

The ranking approach of the previous paragraph (Fig. 11) was performed to identify genes with the highest coverage of compounds in the NRW and TGP data sets. Therefore, it is of interest to determine whether expression of the identified

high-coverage genes is also altered by compounds not represented in the NRW and TGP data sets. To address this question, the seven top-ranking upregulated genes of the combined data set were chosen (*Cyp11a1*, *Ugt2b1*, *Cdkn1a*, *Mdm2*, *Aldh1a1*, *Cyp4a3*, *Ehhadh*) for follow-up analysis of further compounds using qRT-PCR. It is arbitrary that exactly seven genes were selected; yet increasing the gene number would only moderately improve coverage but increase the workload. Six structural analogues of valproic acid (VPA) (2-ethylhexanoic acid, 2-ethylbutyric acid, 2-ethyl-1-hexanol, 4-ene-valproic acid, 4-pentenoic acid and pivalic acid) were used to treat cultivated rat hepatocytes in

Fig. 11 ‘High-coverage genes’ and the number compounds. First, the gene is identified that is deregulated by most of the test compounds (indicated by ‘1’ on the *x*-axis); subsequently, the next ranking gene is added so that both genes are deregulated by as many compounds

as possible (indicated by ‘2’ on the *x*-axis), and so on. Orange color indicates the top-ranking up- and green the top-ranking downregulated genes. The analysis was performed for the NRW (left), TGP (middle) and the combined (left) data sets. (Color figure online)

addition to the parent compound. While VPA was included into the TGP data set, the six analogues were not present in either the TGP or the NRW data sets. All tested VPA analogues caused a clear increase in the expression of at least one of the seven genes, although the spectrum and intensity of induction differed among the compounds (Fig. 12; raw data: Table S14).

Discussion

In the present study, a set of genes was identified which show similar expression responses (up- or downregulation) *in vitro* in cultivated rat hepatocytes and in the liver of rats after exposure *in vivo*. This analysis is based on a relatively large number of compounds ($n = 162$) that have been studied *in vivo* and *in vitro* in genome-wide expression analyses stored in two databases, representing the NRW and the TGP data sets. Genome-wide expression data have been generated in the present study—the NRW dataset (GSE119933)—from cultivated hepatocytes treated with 29 chemicals, while the corresponding *in vivo* data were already available (Ellinger-Ziegelbauer et al. 2008, GSE68110). TGP represents a public data set which contains *in vitro* data from cultivated hepatocytes and livers from rats that were exposed to the same compounds. Only five compounds from both data sets (NRW and TGP) overlap. Analysis of data structure and quality was performed similar to a previous publication (Grinberg et al. 2014).

One of the striking features observed upon comparing *in vitro* and *in vivo* data is that Euclidean distances (as a measure of the extent of overall gene expression alteration) between exposed samples and controls are larger in rat livers *in vivo* compared to *in vitro*-cultivated hepatocytes.

Moreover, the variability (defined by the distance between replicates) of the *in vitro* data was more heterogeneous compared to the *in vivo* experiments. This suggests that the homeostatic control *in vivo* is more efficient compared to what is possible in the applied *in vitro* sandwich culture system. Profiling of concentration progression as previously described by Grinberg et al. (2014) indicated suboptimal data quality in parts of both data sets. To nevertheless allow for the identification of the most reliable genes that fulfil the criteria (1) similar response *in vivo* and *in vitro*, and (2) coverage of as many test compounds as possible, the selection value (SV) concept was used (Grinberg et al. 2014). Here, analysis was based on SV3 genes, which means that only those genes that were up- or downregulated by three or more compounds were considered for further analysis. The number of genes in the overlap of *in vitro*- and *in vivo*-deregulated genes was 5.9 times (up) and 2.8 times (down) enriched compared to random expectation, based on the calculation of overlap ratios as recently established (Shinde et al. 2017). Nevertheless, comparing genes deregulated *in vitro* and *in vivo* showed that the larger fraction of genes does not fall within the overlap. This observation confirms previous studies that identified major differences in gene expression of hepatocytes due to the isolation and cultivation process and due to the absence of non-parenchymal cells of the liver in the *in vitro* system (Zellmer et al. 2010; Heise et al. 2012; Ghallab 2015) and different responses to external stimuli (Godoy et al. 2009, 2010). *In vitro* cultivation induces upregulation of an inflammation/RNA processing and migration/cell cycle-associated gene cluster and downregulation of genes corresponding to mature liver functions (Godoy et al. 2016). However, the more than random overlap of compound-induced genes demonstrates that there are some genes showing similar treatment-induced gene

Fig. 12 Follow-up analysis of seven top-ranking genes with compounds not represented in the NRW and TGP data sets. Valproic acid (represented in TGP) was included as the positive control. Cultivated rat hepatocytes from three different animals were analyzed; mean val-

ues, as well as standard deviations are shown. The seven genes were chosen, because they ranked top among the upregulated genes from the combined NRW and TGP data sets

Fig. 12 (continued)

expression deregulations in vitro and in vivo. This leads to the difficult situation that some genes respond similarly in vitro and in vivo, while others respond differently. Interpretation of in vitro data from cultivated rat hepatocytes should, therefore, consider to which of the two groups (in vivo similar versus in vivo dissimilar) a gene up- or down-regulated by a test compound belongs. It is apparent that genes with clear in vivo–in vitro differences in response to chemicals should be avoided.

A relevant result of this study is the very large similarity obtained in the two independent data sets, NRW and TGP. In both data sets, in vitro–in vivo consensus genes could be identified as described in the previous paragraph. Notably, the in vivo–in vitro consensus genes of both data bases showed an overlap 21.4 times (upregulated genes) and 12.1

times (downregulated genes) higher than randomly expected. This is remarkable since only five test compounds were common to both data sets. The results suggest that a set of consensus or stereotypical response genes exists, which (1) show similar responses (in term of up- or downregulation after test compound exposure) in hepatocytes in vitro and liver in vivo, and (2) show a high coverage, meaning that the same genes are up- or downregulated by a large number of different chemicals.

For practical purposes, it is important to know the fraction of compounds that are covered by the aforementioned consensus or stereotypical response genes. A cumulative analysis demonstrates that the seven top genes cover approximately 50% of the compounds in the NRW and approximately 19% in the TGP data set. However, it should

be considered that the TGP data set in particular contains a relatively high fraction of 'low-profile compounds', which deregulate no or only small numbers of genes. This may be explained by experimental errors in one (or more) of the replicates, which cause outliers and consequently lead to a lack of statistical significance. Moreover, the selected doses or test compound concentrations may have been too low for these compounds, which may also explain the lack of a clear gene expression response. Therefore, the coverage by the seven-gene set may be higher for compounds and exposure conditions that cause strong gene expression changes.

An important question is whether the expression of the identified top-ranking genes (Fig. 11) is also influenced by compounds not included in the NRW and TGP data sets. To study this question, six structural analogues of valproic acid were chosen. Valproic acid itself had been included in the TGP, but not the six analogues. The seven high-coverage genes of the combined NRW and TGP data sets were analyzed by RT-PCR. For this analysis, we focused on the upregulated genes because it is technically easier and more specific to demonstrate increased expression, whereas decreased RNA levels may be due to initial cytotoxic effects. Interestingly, all analyzed analogues of valproic acid clearly induced at least one of the seven top-ranking genes. Currently, read-across techniques, where data-poor compounds of interest are compared to structurally related, data-rich compounds, represent an intensive field of research in regulatory toxicology (Leist et al. 2017). One core component of a read across is to not only study whether the compounds assessed together are structurally and pharmacokinetically similar, but also to show a toxicological/biological similarity. The case study of valproic acid presented here suggests that the identified top-ranking genes seem to be applicable for the comparison of VPA analogues. One conclusion of the read-across case study in Fig. 12 is that none of the studied analogues is inactive with respect to the induction of at least one gene within the seven-gene set; however, pivalic acid and 4-pentenoic acid are less active than valproic acid, while the effects of, e.g., 2-ethylhexanoic acid and 4-ene-VPA are similar to those of the parent compound. It remains to be studied, whether these differences in expression changes correspond to adverse effects of the VPA analogues.

In conclusion, the genes of the seven-gene set identified here are induced by a wide range of different chemicals and respond similarly in vitro and in vivo. However, genome-wide expression analysis in rat hepatocyte sandwich cultures remains a complex challenge since many genes are deregulated by the hepatocyte isolation and cultivation process, respond differently in vivo and in vitro, and expression data in vitro show a relatively high variability compared to in vivo experiments with the same compounds. These limitations need to be taken into account when gene signatures like the one extracted here

from in vivo and in vitro databases are used for a comparative toxicological characterization of compounds in vitro.

Acknowledgements The read-across and bioinformatic parts of the study were supported by the European Union funded project EUToxRisk (EU, no. 681002). Additional parts were supported by the projects StemCellNet (BMBF, 01EK1604A), Liver Simulator (BMBF, 031A355A), DILI (BMBF, 031L0074F), LiSyM (BMBF, 031L0045), LivSysTransfer (BMBF, 031L0119), DEEP (BMBF, 01KU1216) and test compounds of InnoSysTox (BMBF/EU, 031L0021A) were used.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- De Abrew KN, Overmann GJ, Adams RL et al (2015) A novel transcriptomics based in vitro method to compare and predict hepatotoxicity based on mode of action. *Toxicology* 328:29–39. <https://doi.org/10.1016/j.tox.2014.11.008>
- 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
- Ellinger-Ziegelbauer H, Adler M, Amberg A et al (2011) The enhanced value of combining conventional and “omics” analyses in early assessment of drug-induced hepatobiliary injury. *Toxicol Appl Pharmacol* 252(2):97–111. <https://doi.org/10.1016/j.taap.2010.09.022>
- Ghallab A (2015) Highlight report: blueprint for stem cell differentiation into liver cells. *EXCLI J* 14:1017–1019
- Godoy P, Hengstler JG, Ilkavets I et al (2009) Extracellular matrix modulates sensitivity of hepatocytes to fibroblastoid dedifferentiation and transforming growth factor beta-induced apoptosis. *Hepatology* 49(6):2031–2043. <https://doi.org/10.1002/hep.22880>
- Godoy P, Schug M, Bauer A, Hengstler JG (2010) Reversible manipulation of apoptosis sensitivity in cultured hepatocytes by matrix-mediated manipulation of signaling activities. *Methods Mol Biol* 640:139–155. https://doi.org/10.1007/978-1-60761-688-7_7
- Godoy P, Hewitt NJ, Albrecht U et al (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(8):1315–1530
- Godoy P, Schmidt-Heck W, Natarajan K et al (2015) Gene networks and transcription factor motifs defining the differentiation of stem cells into hepatocyte-like cells. *J Hepatol* 63(4):934–942. <https://doi.org/10.1016/j.jhep.2015.05.013> (Erratum in: *J Hepatol* 64(2):525–6)
- Godoy P, Widera A, Schmidt-Heck W et al (2016) Gene network activity in cultivated primary hepatocytes is highly similar to diseased mammalian liver tissue. *Arch Toxicol* 90(10):2513–2529
- Grinberg M, Stöber RM, Edlund K et al (2014) Toxicogenomics directory of chemically exposed human hepatocytes. *Arch Toxicol* 88(12):2261–2287. <https://doi.org/10.1007/s00204-014-1400-x>
- Heise T, Schug M, Storm D et al (2012) In vitro–in vivo correlation of gene expression alterations induced by liver carcinogens. *Curr Med Chem* 19(11):1721–1730

- Ippolito DL, AbdulHameed MD, Tawa GJ et al (2016) Gene expression patterns associated with histopathology in toxic liver fibrosis. *Toxicol Sci* 149(1):67–88. <https://doi.org/10.1093/toxsci/kfv214>
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4(2):249–264
- Krug AK, Kolde R, Gaspar JA et al (2013) Human embryonic stem cell-derived test systems for developmental neurotoxicity: a transcriptomics approach. *Arch Toxicol* 87(1):123–143. <https://doi.org/10.1007/s00204-012-0967-3>
- Leist M, Ghallab A, Graepel R et al (2017) Adverse outcome pathways: opportunities, limitations and open questions. *Arch Toxicol* 91(11):3477–3505
- Lohr M, Hellwig B, Edlund K et al (2015) Identification of sample annotation errors in gene expression datasets. *Arch Toxicol* 89(12):2265–2272. <https://doi.org/10.1007/s00204-015-1632-4>
- Reif R (2015) Expression classifiers for developmental toxicants. *EXCLI Journal* 14:1291–1293
- Rempel E, Hoelting L, Waldmann T et al (2015) A transcriptome-based classifier to identify developmental toxicants by stem cell testing: design, validation and optimization for histone deacetylase inhibitors. *Arch Toxicol* 89(9):1599–1618. <https://doi.org/10.1007/s00204-015-1573-y>
- Schug M, Stöber R, Heise T et al (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. <https://doi.org/10.1007/s00204-012-0999-8>
- Shinde V, Hoelting L, Srinivasan SP et al (2017) Definition of transcriptome-based indices for quantitative characterization of chemically disturbed stem cell development: introduction of the STOP-Tox_{ukn} and STOP-Tox_{ukk} tests. *Arch Toxicol* 91(2):839–864
- Stemmer K, Ellinger-Ziegelbauer H, Ahr HJ, Dietrich DR (2007) Carcinogen-specific gene expression profiles in short-term treated Eker and wild-type rats indicative of pathways involved in renal tumorigenesis. *Cancer Res* 67(9):4052–4068
- Stoebner R (2015) Transcriptomic signature for drug-induced steatosis. *EXCLI J* 14:1259–1260
- Waldmann T, Rempel E, Balmer NV et al (2014) Design principles of concentration-dependent transcriptome deviations in drug-exposed differentiating stem cells. *Chem Res Toxicol* 27(3):408–420. <https://doi.org/10.1021/tx400402j>
- Zellmer S, Schmidt-Heck W, Godoy P et al (2010) Transcription factors ETF, E2F, and SP-1 are involved in cytokine-independent proliferation of murine hepatocytes. *Hepatology* 52(6):2127–2136. <https://doi.org/10.1002/hep.23930>

Affiliations

Marianna Grinberg¹ · Regina M. Stöber² · Wiebke Albrecht² · Karolina Edlund² · Markus Schug² · Patricio Godoy² · Cristina Cadenas² · Rosemarie Marchan² · Alfonso Lampen³ · Albert Braeuning³ · Thorsten Buhrke³ · Marcel Leist⁴ · Axel Oberemm³ · Birte Hellwig¹ · Hennicke Kamp⁵ · Iain Gardner⁶ · Sylvia Escher⁷ · Olivier Taboureau^{8,9} · Alejandro Aguayo-Orozco⁸ · Agapios Sachinidis¹⁰ · Heidrun Ellinger-Ziegelbauer¹¹ · Jörg Rahnenführer¹ · Jan G. Hengstler²

¹ Department of Statistics, TU Dortmund University, Dortmund, Germany

² Leibniz Research Centre for Working Environment and Human Factors, Technical University of Dortmund (IfADo), Ardeystrasse 67, 44139 Dortmund, Germany

³ Department of Food Safety, German Federal Institute for Risk Assessment, Max-Dohrn-Str. 8-10, 10589 Berlin, Germany

⁴ Fachbereich Biologie, University of Konstanz, Universitätsstrasse 1, 78457 Konstanz, Germany

⁵ BASF SE, Experimental Toxicology and Ecology, Ludwigshafen am Rhein, Germany

⁶ Simcyp (A Certara Company), Sheffield, UK

⁷ Fraunhofer Institute of Toxicology and Experimental Medicine (ITEM), Hannover, Germany

⁸ Center for Protein Research, Novo Nordisk Foundation Center for Protein Research (NNFCPR), University of Copenhagen (UCPH), Copenhagen, Denmark

⁹ UMRS 973, INSERM, Sorbonne Paris Cité, University of Paris, Diderot, France

¹⁰ Institute of Neurophysiology and Centre for Molecular Medicine Cologne (CMMC), University of Cologne (UKK), Robert-Koch-Str. 39, 50931 Cologne, Germany

¹¹ Investigational Toxicology, Drug Discovery, Pharmaceuticals, Bayer AG, 42096 Wuppertal, Germany