Application of High-Throughput Transcriptomics for Mechanism-Based Biological Read-Across of Short-Chain Carboxylic Acid Analogues of Valproic Acid

Nanette G. Vrijenhoek, Matthias M. Wehr, Steven J. Kunnen, Lukas S. Wijaya, Giulia Callegaro, Martijn J. Moné, Sylvia E. Escher and Bob van de Water

1Leiden Academic Centre of Drug Research, Leiden University, Leiden, The Netherlands; 2ITEM, Fraunhofer Institute, Hannover, Germany

Abstract
Chemical read-across is commonly evaluated without specific knowledge of the biological mechanisms leading to observed adverse outcomes in vivo. Integrating data that indicate shared modes of action in humans will strengthen read-across cases. Here we studied transcriptomic responses of primary human hepatocytes (PHH) to a large panel of carboxylic acids to include detailed mode-of-action data as a proof-of-concept for read-across in risk assessment. In rodents, some carboxylic acids, including valproic acid (VPA), are known to cause hepatic steatosis, whereas others do not. We investigated transcriptomics responses of PHHs exposed for 24 h to 18 structurally different VPA analogues in a concentration range to determine biological similarity in relation to in vivo steatotic potential. Using a targeted high-throughput screening assay, we assessed the differential expression of ~3,000 genes covering relevant biological pathways. Differentially expressed gene analysis revealed differences in potency of carboxylic acids, and expression patterns were highly similar for structurally similar compounds. Strong clustering occurred for steatosis-positive versus steatosis-negative carboxylic acids. To quantitatively define biological read-across, we combined pathway analysis and weighted gene co-expression network analysis. Active carboxylic acids displayed high similarity in gene network modulation. Importantly, free fatty acid synthesis modulation and stress pathway responses are affected by active carboxylic acids, providing coherent mechanistic underpinning for our findings. Our work shows that transcriptomic analysis of cultured human hepatocytes can reinforce the prediction of liver injury outcome based on quantitative and mechanistic biological data and support its application in read-across.

1 Introduction
There is a need for new approach methodologies (NAMs) for chemical safety assessment that will complement existing and future requirements for non-animal based testing for next generation risk assessment (Kavlock et al., 2018; Pistollato et al., 2021). Therefore, read-across assessment and integrated approaches to testing and assessment (IATAs) in the context of adverse outcome pathways (AOPs) are implemented in toxicity testing strategies (Escher et al., 2019; Rovida et al., 2021). Read-across among compounds is commonly used for “gap-filling” of target compounds by studying effects of structurally similar source compounds (Schultz et al., 2015). In routine chemical read-across, however, knowledge of biological mechanisms is lacking. In contrast, AOPs are frameworks to organize mechanistic knowledge and information regarding biological processes that lead to an adverse event, but it is time-consuming and data-demanding to build AOPs. The application of AOPs in toxicity testing is supporting IATAs in regulatory decision-making (Patlewicz et al., 2014; Tollefsen et al., 2014; Vinken, 2015). Furthermore, toxicity test systems are being developed for high-throughput screening, allowing testing of numerous compounds at multiple concentrations. Such biological assays aim to unravel the mode-of-action (MoA) underlying adverse events, contributing to AOP development and assessing similarities in MoA among compounds. Transcriptomic
techniques are continuously developed to be more cost-effective and less time-consuming, allowing high-throughput gene expression profiling of multiple compounds at different concentrations in parallel (Mav et al., 2018; Gwinn et al., 2020; Harrill et al., 2021). The combination of these techniques has already proven a robust method to quantify concentration-dependent chemical-induced transcriptomic responses for a panel of different liver toxicants in HepaRG cells (Ramaiahgari et al., 2019).

Here, we combined a biological similarity study with physicochemical data, transcriptomic data and information on the in vivo adverse outcome using 18 structurally similar carboxylic acids. It was our premise that unravelling the cellular MoA could eventually strengthen a read-across assessment. Hereto, we performed transcriptomic analysis on the entire panel of 18 carboxylic acids. One widely studied carboxylic acid is valproic acid (VPA), an anticonvulsant with a potential for liver toxicity. Since the MoA of VPA is largely known and an AOP of drug-induced liver steatosis has been established (van Breda et al., 2018), we were able to systematically assess whether other carboxylic acids induce mechanistically similar pathway responses. Furthermore, a selection of these short-chain fatty acids (SCFA) has previously been tested in animals and also caused liver steatosis. Therefore, we anticipated that different carboxylic acids might demonstrate a similar transcriptomic response, albeit with different potency. We used primary human hepatocytes (PHH) as the test system and deployed the Toxicogenomics-MAPr tool (TXG-MAPr; Callegaro et al., 2021) to derive toxicological fingerprints induced by carboxylic acids. This tool allows a rapid quantitative mechanistic understanding of transcriptomic data based on weighted gene co-expression network analysis (WGCNA) of the TG-GATEs database. In short, the TXG-MAPr consists of modules, which are gene co-expression networks based on the gene expression data of PHHs exposed to more than 150 compounds and available in the TG-GATEs database. These modules are also annotated with functional information including pathway annotation and transcription factor enrichment. The eigengene (EG) score summarizes the log2 fold change (log2FC) of the genes within a module, and this is also reflected in the size and color of a module circle in the TXG-MAPr. The gene that has the highest correlation between the log2FC and the module EG score is called a hub gene, and this gene is the most representative of the entire module. The TXG-MAPr tool has a feature for uploading new transcriptomic data, which is used to calculate gene sub-network (module) perturbations captured in a module EG score. We compared our data with VPA expression profiles from the TG-GATEs database and with full-transcriptome panels generated in HepG2 cell lines. Finally, stress pathway responses, involved in many toxicological responses, were investigated.

2 Materials and methods

Cell culture

PHHs (LiverPool™ 10-donor mixed gender pooled cryoplatable human hepatocytes, Cat No. X008001-P) were acquired from BioReclamationIVT. Upon thawing, cells were diluted in warm thawing medium (SEKISUI XenoTech OptiThaw Hepatocyte Media, Cat No. K8000) and centrifuged for 10 min at 100 g. Cells were resuspended in InVitroGro CP medium (BioIVT, Cat no. Z99029) supplemented with Torpedo™ antibiotics mix (BioIVT, Cat no. Z99000) and plated in clear 96-well plates precoated with collagen (Corning, Cat no. 10469602) at 70,000 cells per well. 8 hours after plating, the medium was replaced with InVitroGro HI medium (BioIVT, Cat no. Z99009) supplemented with Torpedo™ mix. One day after plating, cells were chemically exposed to carboxylic acids in supplemented InVitroGRO™ HI medium.

HepG2 cells were acquired from ATCC (clone HB8065). They were maintained and exposed in high-glucose DMEM supplemented with 10% (v/v) FBS, penicillin (25 U/mL) and streptomycin (25 μg/mL). Cells were used between passage 12 and 15 and plated at 50,000 cells per well in clear 96-well plates. One day after plating, cells were exposed to carboxylic acids.

Reagents and exposures

All carboxylic acids were acquired from Sigma Aldrich except for 2-propylheptanoic acid (Endeavour Speciality Chemicals Ltd) and 2-ethylpentanoic acid (Santa Cruz) through the EU-ToxRisk consortium. These compounds and their characteristics are summarized in Table 1. All treatments were freshly prepared on the day of exposure and were accompanied with DMSO in a final treatment concentration of 0.1%. The concentration range of the exposures for all compounds and in all cell types, except samples measured for whole transcriptome, was 0.2 mM and a log2 scale of 6 concentrations (0.5 to 16 mM). The 0.2 mM concentration was included in the range of all compounds since a concentration around the cmax was desired (cmax: maximum concentration of VPA detected in plasma in patients (Huppelschoten, 2017)). To compare gene expression between TempO-seq EU-ToxRisk 2.2 gene set and whole transcriptome gene set, in a separate experiment HepG2 cells were exposed to VPA in a concentration range of 0.12, 0.6, 1.2, 2.4, 4.9, 7.3 mM. Exposure time for all treatments for EU-ToxRisk 2.2. gene set was 24 h; but for the HepG2 whole transcriptome gene set samples were obtained at both 8 h and 24 h. All data is from 3 independent biological replicates. For the negative control (0.1% DMSO) condition, 3 technical replicates per biological replicate were pooled and sequenced as 1 sample.

Abbreviations

AOP, adverse outcome pathway; CR, concentration response; DEGs, differentially expressed genes; EG, eigengene; FC, fold change; HDAC, histone deacetylase; GO, gene ontology; IATA, integrated approaches to testing and assessment; MACCS, Molecular ACCess System; MCFA, medium-chain fatty acids; MoA, mode-of-action; NAMs, new approach methodologies; PHH, primary human hepatocyte; PS, potency score; SCFA, short-chain fatty acids; TG-GATEs, Toxicogenomics Project-Genomics Assisted Toxicity Evaluation System; TXG-MAPr, Toxicogenomics-MAPr; VPA, valproic acid; WGCNA, weighted gene co-expression network analysis
Tab. 1: Carboxylic acids

<table>
<thead>
<tr>
<th>Name</th>
<th>CAS</th>
<th>Abbreviation</th>
<th>Structure (SMILES)</th>
<th>Structural similarity relative to VPA</th>
<th>Chain length at position 2</th>
<th>Branched</th>
<th>Steatosis in vivo data</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valproic acid</td>
<td>99-66-1</td>
<td>VPA</td>
<td>CCCCC(CCC)C(=O)O</td>
<td>1</td>
<td>3.3.0</td>
<td>branched</td>
<td>positive</td>
<td>Espandiari et al., 2008; Tong et al., 2005; Sugimoto et al., 1987;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Löscher, 1992; Abdel-Dayem, 2014; Knapp, 2008; Ibrahim, 2012; Acosta, 2012</td>
</tr>
<tr>
<td>2-Propyl-heptanoic acid</td>
<td>31080-39-4</td>
<td>2PHP</td>
<td>CCCCC(CCC)C(=O)O</td>
<td>0.97</td>
<td>5.3.0</td>
<td>branched</td>
<td>unknown</td>
<td></td>
</tr>
<tr>
<td>2-Ethyl-heptanoic acid</td>
<td>3274-29-1</td>
<td>2EHPA</td>
<td>CCCCC(CCC)C(=O)O</td>
<td>0.97</td>
<td>5.2.0</td>
<td>branched</td>
<td>unknown</td>
<td></td>
</tr>
<tr>
<td>2-Propyl-hexanoic acid</td>
<td>3274-28-0</td>
<td>2PHA</td>
<td>CCCCC(CCC)C(=O)O</td>
<td>0.97</td>
<td>4.3.0</td>
<td>branched</td>
<td>unknown</td>
<td></td>
</tr>
<tr>
<td>2-Ethyl-hexanoic acid</td>
<td>149-57-5</td>
<td>EHXA</td>
<td>CCCCC(CC)=O</td>
<td>0.97</td>
<td>4.2.0</td>
<td>branched</td>
<td>positive</td>
<td>Juberg et al., 1998; Szilagyi, 2012</td>
</tr>
<tr>
<td>2-Ethyl-pentanoic acid</td>
<td>20225-24-5</td>
<td>EPA</td>
<td>CCCCC(CCC)C(=O)O</td>
<td>0.94</td>
<td>3.2.0</td>
<td>branched</td>
<td>unknown</td>
<td></td>
</tr>
<tr>
<td>2-Methyl-hexanoic acid</td>
<td>4536-23-6</td>
<td>2-MHXA</td>
<td>CCCCC(CCC)C(=O)O</td>
<td>0.91</td>
<td>4.1.0</td>
<td>branched</td>
<td>unknown</td>
<td></td>
</tr>
<tr>
<td>2-Propyl-4-pentenoic acid (4-ene VPA)</td>
<td>1575720</td>
<td>2EVPA</td>
<td>CCCCC(CCC)=O</td>
<td>0.89</td>
<td>3.3.0</td>
<td>branched</td>
<td>positive</td>
<td>Tang, 1995; Bailie, 1992</td>
</tr>
<tr>
<td>Octanoic acid (caprylic acid)</td>
<td>124072</td>
<td>OCT</td>
<td>CCCCCCCCCC(=O)O</td>
<td>0.89</td>
<td>6.0.0</td>
<td>unbranched</td>
<td>negative</td>
<td>Szilagyi, 2012</td>
</tr>
<tr>
<td>2-Methyl-pentanoic acid</td>
<td>97-61-0</td>
<td>MPA</td>
<td>CCCCC(CCC)=O</td>
<td>0.88</td>
<td>3.1.0</td>
<td>branched</td>
<td>unknown</td>
<td>Szilagyi, 2012</td>
</tr>
<tr>
<td>Hexanoic acid (caproic acid)</td>
<td>142-62-1</td>
<td>HEX</td>
<td>CCCCCC(=O)O</td>
<td>0.86</td>
<td>4.0.0</td>
<td>unbranched</td>
<td>negative</td>
<td>Szilagyi, 2012</td>
</tr>
<tr>
<td>2-Ethyl-1-hexanol</td>
<td>104767</td>
<td>2E1H</td>
<td>CCCCC(CC)=O</td>
<td>0.81</td>
<td>4.2.0</td>
<td>branched</td>
<td>positive</td>
<td>Semino, 1998</td>
</tr>
<tr>
<td>2-Methyl-butyrice acid</td>
<td>1730-91-2</td>
<td>MBUT</td>
<td>CC[C@H]C(=O)O</td>
<td>0.8</td>
<td>2.1.0</td>
<td>branched</td>
<td>unknown</td>
<td></td>
</tr>
<tr>
<td>2,2-Dimethyl-pentanoic acid</td>
<td>1185393</td>
<td>2.2MPA</td>
<td>CCCC(CC)=O</td>
<td>0.79</td>
<td>3.1.1</td>
<td>branched</td>
<td>unknown</td>
<td></td>
</tr>
<tr>
<td>4-Pentenoic acid</td>
<td>591800</td>
<td>4PA</td>
<td>C=CCCC(=O)O</td>
<td>0.78</td>
<td>3.0.0</td>
<td>branched</td>
<td>positive</td>
<td>Thayer, 1984; Yuge, 1990; Tang et al., 1995; Bailie, 1992</td>
</tr>
<tr>
<td>2-Ethyl-butyrice acid</td>
<td>88-09-5</td>
<td>2EB</td>
<td>CCCC(CC)=O</td>
<td>0.73</td>
<td>2.2.0</td>
<td>branched</td>
<td>negative</td>
<td>Szilagyi, 2012</td>
</tr>
<tr>
<td>Propionic acid</td>
<td>79094</td>
<td>PROP</td>
<td>CCC(=O)O</td>
<td>0.62</td>
<td>2.0.0</td>
<td>unbranched</td>
<td>negative</td>
<td>Szilagyi, 2012</td>
</tr>
<tr>
<td>Pivalic acid</td>
<td>75-98-9</td>
<td>PIV</td>
<td>CC(C)(C)=O</td>
<td>0.53</td>
<td>1.1.1</td>
<td>branched</td>
<td>negative</td>
<td>IUCLID, Shell Netherlands, 1990</td>
</tr>
</tbody>
</table>
Annotation of carboxylic acids

Carboxylic acids were annotated by the following parameters: steatosis induction, chain length, branching, and VPA-similarity. Steatosis induction was determined by literature research and previously reported (Escher et al., 2021). Chain length was determined by counting Cs from the 2nd C from the acid group. Branched carboxylic acids have 2 or more chains from the 2nd C. VPA similarity scores of all analogues relative to the target compound were calculated using the MACCS (Molecular ACCess System) fingerprint from RDKit and Dice algorithm, where 1 is VPA and 0.53 is most dissimilar to VPA in our compound set (Tab. 1).

Transcriptomics

After 24 h of compound exposure, cells were washed with PBS and lysed with 50 µL 1x BNN lysis buffer (BioClavis, Glasgow, Scotland) diluted in PBS. After 15 min at RT, the lysate was frozen at -80°C and shipped to BioClavis, where lysates were analyzed using TempO-Seq technology, deploying either the SI500+ gene set (Mav et al., 2018) plus additional genes, together called EU-ToxRisk 2.2 (Tab. S1), for the PHH and HepG2 samples, referred to as SI500+ set in the manuscript for the PHH and HepG2 samples, or the whole transcriptome gene set (manifest supplied in Tab. S1). Probe alignment for EU-ToxRisk 2.2 gene set and whole transcriptome gene set was performed by BioClavis (Yeakley et al., 2017; Limonciel et al., 2018; Yang et al., 2020). Briefly, FASTQ files were aligned using Bowtie, allowing for up to 2 mismatches in the target sequence. This bioinformatics pipeline contained several quality controls with mapped/unmapped reads, replicate clustering, and sample clustering (Yeakley et al., 2017). When samples passed these tests, result count tables were ready for analysis.

Statistical analysis

Read counts of all probes were analyzed by an in-house pipeline using R (Fig. S1). Data wrangling was done with packages reshape (Wickham, 2007), tidyr2, and dplyr3. First, we performed a quality control on the transcriptomic data. Hereafter, we analyzed the library size and count distribution of the dataset. Samples with a library size below 500k were discarded, since they were visibly clear outliers from the library size boxplots and may not represent a complete transcriptomic response (Fig. S2A).

Variance of the probes followed an unequal bimodal histogram, suggesting that part of the probe set was not activated by any of the compounds (Fig. S2B). Probes with overall variance below 1 were discarded since no variance indicated no changes in those genes in the whole dataset. They were excluded from DEG calculations to strengthen the significance of genes with higher variability in their expression, leaving ~2,700 genes.

For DEG calculation of targeted transcriptomics, we used the same approach as for RNA-seq data, which can be analyzed by the DESeq2 package, which is based on a negative binomial regression model (Love et al., 2014). Therefore, we checked and confirmed a negative binomial distribution in our targeted transcriptome data by plotting the pooled variance with package edgeR (Robinson et al., 2010) (Fig. S1C).

In the analysis of samples sequenced with the whole transcriptome, no probes were discarded. Counts were normalized to counts per million. All chemical-treated samples were compared to the DMSO control treated samples to calculate the log2FC and p-values with the DESeq2 package. DEGs had a p-adjusted value of < 0.05 and |log2FC| > 1.5 unless otherwise stated (Fig. S1). FCs and adjusted p-values were uploaded to the TXG-MAPr tool4 (Callegaro et al., 2021), whereupon EG scores of the modules were extracted. Stress pathway target genes of the TempO-Seq data were extracted from the DoRothEA database, containing transcription factor-target interactions (Garcia-Alonso et al., 2019). Interactions are assigned levels ranging from A-E, where A is highest and E is lowest confidence. Levels ABC were used to extract target genes of corresponding stress pathway transcription factors.

All figures were made with either the R packages ggplot5, UpSetR6 or pheatmap7, or the TXG-MAPr tool (Callegaro et al., 2021) and assembled in Adobe Illustrator V25.2.

Fig. 1 (see the next page): Differentially expressed genes in response to carboxylic acid exposure

A) Vulcano plots displaying DEGs for VPA-treated PHHs.
B) Dose response graphs for the number of DEGs in response to all carboxylic acids. Red = significantly upregulated, blue = significantly downregulated. Similarity indicates structural similarity to VPA; steatosis indicates in vivo potential.
C) PS = potency score, which is the sum of all DEGs in a concentration response (CR).
D) Heatmap showing the top 50 probes upregulated for VPA at 4 mM for all carboxylic acids. Darker blue in the similarity annotation means more similar to VPA, light-dark green indicates concentration range. Purple colors reflect the potency score of B.

The genes with probe number are displayed vertically.

---

1 doi:10.14573/altext.2107261s

2 tidyr: Tidy Messy Data. https://cran.r-project.org/package=tidy

3 dplyr: A Grammar of Data Manipulation. https://cran.r-project.org/package=dplyr

4 https://bg-mapr.eu/WGGNA_PHH/TGGATEs_PHH/


7 pheatmap: Pretty Heatmaps. https://cran.r-project.org/package=pheatmap
3 Results

3.1 Differentially expressed genes induced by carboxylic acids in PHH can be used to distinguish potent from non-potent responders

To investigate the MoA in response to carboxylic acids in PHH, we performed targeted RNAseq for all compounds at 24 h after exposure in a concentration response (CR) manner. VPA displayed a clear CR in DEGs (Fig. 1A,B) (p-adj < 0.05 & |log2FC| > 1.5). At the highest concentrations, up to 412 genes were differentially expressed, whereas at the lowest concentration only 3 DEGs were observed. The sum of all DEGs in the VPA CR equals 1117, which we named the potency score (PS). By quantifying DEGs for each compound CR, we found that some carboxylic acids showed a CR, whereas others did not. Less potent responders included pivalic acid (PS equals 25 DEGs across the whole CR) and 2-ethylbutyric acid (PS = 441), which are both short-chain carboxylic acids and classified as steatosis-negative. Strong inducers of DEGs included 2-propyl-4-pentenoic acid (PS = 1000), a reactive metabolite of VPA that is known to induce steatosis, and 2-ethyl-hexanoic acid (PS = 928), another well-known steatosis-inducing compound (Escher et al., 2021). Moreover, all three unbranched carboxylic acids, namely propionic, hexanoic and octanoic acid (Tab. 1), caused a CR in the DEGs in PHH (Fig. 1B).

In support of a biological similarity, we anticipated a strong overlap of DEGs in response to different carboxylic acids. When focusing on the middle concentration of 4 mM, we identified some overlapping genes for active responders, but only 6 genes were found to overlap for 14 compounds (Fig. 1C, letter D). Furthermore, 9 genes were unique for 13 compounds (Fig. 1C, letter B). Since most genes were solely found as DEG in single compounds, no obvious gene signature was detected (Fig. 1C).

We hypothesized that PHHs may have the same genes responding to carboxylic acids, but that this response was initiated for a subset of carboxylic acids at a lower concentration. Therefore, we looked at the entire concentration response to determine if overlapping genes are detected at different concentrations. We analyzed the top-50 highest upregulated genes of the 4 mM VPA condition (based on FC with p-adj < 0.05) and looked at the FC of these genes for all compounds across the concentration range. We ranked the compounds by their potency score, expecting to see a similar response in the more potent compounds. Indeed, besides a clear dose response in the top hits for VPA, a very similar response was observed for highly potent carboxylic acids (Fig. 1D).

These genes included CYP enzymes, signal transduction proteins (Wnt5a), growth factors (BDNF), apoptosis regulators (CRO1A), and fatty acid metabolism-related proteins (FYN, TUBB2B). Less potent responders, like pivalic acid and ethylbutyric acid, showed an upregulation of these same genes as well, but only at higher concentrations, indicating similarity in MoA despite reduced potency. Finally, we looked at the overlapping genes at the highest concentration (data not shown). No genes were found to overlap for all compounds. However, all analogues, excluding the less potent pivalic acid and 2-ethyl-1-hexanol, had 16 overlapping genes, some of which are mainly involved in cellular stress response pathways (ATF3, CHAC1, CCL2, HAVCR1, DNAJB1, RRM2, HSPA6, ANXA1), lipid metabolism (UGT2A3, NR1D1) or ion channel activity (KCNK1).

3.2 Carboxylic acid exposures provide transcriptomic fingerprints in which potent compounds cluster together

We found highly similar dose responses in the topmost DEGs among potent carboxylic acids, although there was no clear cross-compound gene signature upon 4 mM exposures that could be linked to the AOP for steatosis (Fig. 1C). Also, at the high concentration, a few genes could be linked to a general stress response, but only 2 genes were clearly linked to lipid metabolism. Therefore, an interactive PHH TXG-MAPr tool based on WGCNA, (Callegaro et al., 2021) was used to further explore the transcriptomic datasets beyond the top 50 genes, allowing to derive concentration responses on gene network activation using all (~2,700) genes for 18 compounds. Some gene co-expression modules responded to VPA and were perturbed proportionally to concentration (Fig. 2A). Upregulated modules included PHH:31, PHH:280, PHH:126 and PHH:217, whereas PHH:325 was downregulated by VPA (see Fig. S5 for more details about these modules). Next, we compared the highest concentration of potent and less potent carboxylic acid by looking at the Pearson correlation of all module EG scores in the TXG-MAPr (Fig. 2B). The module activation pattern of VPA correlated well with 2-propyl-4-pentenoic acid (Pearson R = 0.91) and significantly less with non-responder pivalic acid (Pearson R = 0.11). By correlating all compounds and all concentrations, potent carboxylic acids correlated well with each other, whereas less potent compounds clustered much less (Fig. 2C, Fig. S3). However, carboxylic acids with a less potent response, like 2-ethylbutyric acid, did cluster with more potent compounds like 2-propylheptanoic acid, suggesting a similar yet weaker biological response in PHHs (Fig. S3, cluster 3). Furthermore, even though all unbranched carboxylic acids display a dose response in DEG counts, the higher concentrations do not cluster together based on module EG scores. This finding suggests that compounds with different lengths of unbranched carboxylic acids display a different biological MoA at these high concentrations.

Next, we determined which modules contributed to the similarity of the fingerprint. We focused on the concentration response of modules with |EG score| > 2 and found a similar module activity signature for active compounds. Again, the same modules mentioned before exhibited a dose response (Fig. S4). Even the unbranched carboxylic acids had a higher EG score for our hit modules. So, these modules suggested a fingerprint for carboxylic acid exposure in PHHs and strongly contributed to cluster potent compounds together. This indicated that the MoA of potent compounds was similar but did not yet indicate which underlying biological pathways were affected.

3.3 Module 31 plays a key role in a VPA-like fingerprint and reflects a gene network that represents key events in steatosis

To unravel the biological pathways involved in carboxylic acid exposure, we had a closer look at our hit modules. Module 31 (PHH:31) showed the highest EG scores for VPA as well as other active carboxylic acids with strong module correlation with
Vrijenhoek et al.
ALTEX 39(2), 2022

VPA. Therefore, we further evaluated the biological representation of this gene network. PHH:31 displayed a concentration-response activation for carboxylic acids that also had a concentration response in DEGs (Fig. 3A). There are 31 genes involved in PHH:31, seven of which are measured in the S1500+ gene set (Fig. 3B), namely: KIF5C (major role in organelle transport, also regulates SIRT6, which is essential in lipid regulation), GUCY1B3 (main receptor for nitric oxide, which can be induced by VPA), FYN (regulator of fatty acid oxidation), EEF1A2 (promotes GTP-dependent binding of aminoacyl-tRNA to ribosomes), TUBB2B (axon guidance and neuronal migration), SLC27A1 (regulating long chain fatty acid substrates during beta-oxidation), and NR1I3 (regulates response elements of CYP enzymes). Interestingly, carboxylic acids that were structurally more similar to VPA had similar EG scores at 1 mM, indicating similar sensitivity (Fig. 3C, left). Annotation based on in vivo steatotic potential also demonstrated active carboxylic acids clustering (Fig. 3C, right). When inspecting PHH:31 biological interpretation, gene ontology (GO) terms associated with PHH:31 included fatty acid metabolism and li-

Fig. 2: Toxicogenomics-MAPr correlations for carboxylic acids
A) Application of the TXG-MAPr tool for VPA. Low (0.2 mM), middle (8 mM) and high (16 mM) concentrations are shown. Modules are represented by circles, whose size (small to big, increasing degree of perturbation) and color (blue to red, repression to activation) are proportional to the EG score. Modules that visibly show a dose response are annotated by arrows in the last map. B) Correlation plot of the EG scores of all modules for highly correlated compounds VPA and 2-propyl-4-pentenoic acid (both 16 mM) (Pearson R = 0.91), and for non-correlating compounds pivalic acid and valproic acid (both 16 mM) (Pearson R = 0.11). C) Heatmap showing correlation for all carboxylic acids at 16 mM.
ed set of genes could pose an uncertainty in using PHH:31 for a read-across assessment. Therefore, we investigated whether a similar response is observed when using the full transcriptome. For a cross-systems comparison, we also evaluated PHH data for VPA from TG-GATEs, which is measured with a microarray platform. Moreover, we determined whether a similar transcriptomics response is observed for HepG2, either using the S1500+ or whole-transcriptome TempO-Seq gene panel. We found that previously identified VPA-responding modules show similar activation patterns compared to TG-GATEs data, gase activities (Fig. 3D). Moreover, PPAR signaling was found to be involved, which is part of the AOP for steatosis. Overall, gene annotation of PHH:31 revealed expected biological terms linked to steatosis, an adverse outcome of VPA exposure.

3.4 Comparison of the whole transcriptome with the S1500+ set on VPA-induced module activation
PHH:31 was critical for carboxylic acid responses, yet for the S1500+ gene panel the overall coverage of PHH:31 was limited to 7 out of 31 genes, including the hub gene. This limited set of genes could pose an uncertainty in using PHH:31 for a read-across assessment. Therefore, we investigated whether a similar response is observed when using the full transcriptome. For a cross-systems comparison, we also evaluated PHH data for VPA from TG-GATEs, which is measured with a microarray platform. Moreover, we determined whether a similar transcriptomics response is observed for HepG2, either using the S1500+ or whole-transcriptome TempO-Seq gene panel. We found that previously identified VPA-responding modules show similar activation patterns compared to TG-GATEs data,
**Fig. 4: Comparison of transcriptomic data of liver cells stimulated with VPA**

A) TXG-MAP of 24-h VPA-stimulated PHHs of the TG-GATEs, PHHs with TempO-Seq S1500+ EU-ToxRisk panel 2.2, HepG2 cells with TempO-Seq S1500+ EU-ToxRisk panel 2.2, and HepG2 with TempO-Seq whole transcriptome panel. B) Correlation of the EG scores of the highest concentration in the TG-GATEs opposite scores for 4 mM PHHs with the TempO-Seq S1500+ panel, HepG2 with S1500+ panel, and HepG2 with the TempO-Seq whole transcriptome panel. TG-GATES vs TempO-seq S1500: R = 0.47, TG-GATES vs HepG2 TempO-seq: R = 0.3, TG-GATES vs HepG2 Whole Tr.: R = 0.55. C) Module correlation plot of all concentrations and all panels in a concentration range. Black boxes represent the correlations in B. Transcriptome measurements cluster together, especially at the higher concentrations. D) Fold changes of PHH:31 in HepG2 stimulated with VPA in a dose range. Underlined genes are also present in the S1500+ panel. E) Selection of PHH:31 genes that are in the S1500+ panel set, measured for all TempO-Seq experiments.
regardless of the difference in gene set measured and the used transcriptomics platform (Fig. 4A). Naturally, the EG score was lower for the S1500+ panel due to lower coverage of PHH:31 since the module EG score is determined by the log2FC of all genes. The HepG2 cell response to VPA exposure was much stronger as can be appreciated by larger circles representing multiple modules, including PHH:31. When measured at the whole-transcriptome level, the response was even stronger, being the result of the measurement of more genes. Examining the correlation scores, we clearly observed PHH:31 to be strongly activated in all correlation plots, suggesting that this module is a key driver of the VPA regulated response (Fig. 4B). Interestingly, the strongest correlation was observed for 5 mM VPA data in TG-GATEs with whole transcriptome HepG2 data, indicating a better comparison when the whole transcriptome is used. Finally, when correlating all transcriptomic datasets, data from the same cell type and probe set (S1500+ or whole transcriptome) clearly clustered together, indicating the same transcriptomic platform and gene set should be used for cross-compound comparison (Fig. 4C). Next, we observed that the EG score in module 31 for the whole transcriptome for HepG2 was dominated by a subset of genes. Eight genes showed a dose response, of which 4 genes are in the S1500+ panel (Fig. 4D). Since the genes that exhibited a dose response were also present in the S1500+ set, these genes clearly contributed most to the EG score of PHH:31 in the S1500+ panel (Fig. 4D,E). Furthermore, PHHs had even more genes contributing to this module with a dose response for FYN and GUCY1B3, genes that do not contribute in HepG2 (Fig. 4E). So, although the S1500+ set did not cover the whole transcriptome, we were still able to get a reliable fingerprint of VPA by leveraging the PHH TXG-MAPr, allowing reliable MoA determination and comparison with other carboxylic acids for read-across.

### 3.5 Modest stress response activation in PHH by carboxylic acids

Next, we analyzed the modules that, in addition to PHH:31, were distinctly up- or downregulated upon VPA exposure and treatment with carboxylic acids with similar MoA, including PHH:217, PHH:280, PHH:126 and PHH:325 (Fig. S5). PHH:217 is associated with ion transport, like sodium channel transport and potassium channel regulator activities. Indeed, the MoA of VPA includes blocking sodium channels and T-type calcium channels, which may contribute to its antiepileptic effect (Bourin, 2020). PHH:325 activation, associated with the NRF2 pathway, suggested that the oxidative stress pathway was altered upon VPA exposure. The other two modules had no clear GO terms. However, when we zoomed in on genes involved in PHH:280, we noticed that DDIT3 and HSPA13 were upregulated, both of which are well-known to be involved in the unfolded protein response. PHH:126 displayed regulation of more extracellular matrix functions, like protein binding (NEFL), catenin binding (DSC3), and polysaccharide binding (SBSPON). Given that we found two modules indicating activation of different stress pathways and that the steatosis AOP also contains stress pathway activation, we focused on the stress pathway target gene responses of PHHs to carboxylic acids. We used the DoRothEA database to select genes that are modulated by key stress pathway transcription factors (Fig. 5). Again, we measured similarity in the responses of the most potent carboxylic acids. There was a relatively weak response for the oxidative stress response. For example, SRXN1, a well-known target of NRF2 (Tonelli et al., 2018), was not significantly regulated (data not shown). For other pathways, there was more prominent up- or downregulation of target genes. For instance, for the unfolded protein response, both DDIT3 and PPARC1A were upregulated. The latter is known to induce PPARC transcription activity as well as PPARD, a known regulator of peroxisomal beta-oxidation of fatty acids. Overall, gene expression in various stress response pathways was perturbed at high concentrations and more strongly by compounds similar to VPA. However, no particular stress pathway was strongly or solely activated, which was also highlighted by the most substantial gene alterations for genes having functions in multiple stress response pathways or cellular processes.

### 4 Discussion

In this study, we explored the transcriptomic response of PHH cells upon exposure to carboxylic acids that are structurally related to VPA. Some carboxylic acids, including VPA itself, altered transcriptional activity already at low concentrations, whereas others did not – or only minimally – affect gene expression, even at high concentrations (Fig. 1A,B). Overlapping genes at a single concentration did not reveal a clear pathway-related response (Fig. 1C). However, when looking at the entire concentration response, similarities in most susceptible genes among the potent compounds indicated similar MoA (Fig. 1D). This notion was also reflected in the overlapping co-expressed gene modules, including modules 31, 325, 126, 217 and 280 (Fig. 2), all of which are associated with process known to be modulated by VPA, like fatty acid metabolism, PPAR signaling, ion channel activity, and stress responses (Fig. 3, 5; Fig. S5). Even though our targeted transcriptome panel did not cover all the modules completely (Fig. 4; Fig. S5), our analysis demonstrated that transcriptomic fingerprints can be used to demarcate similarity in MoA of structurally similar chemicals and can, therefore, support read-across approaches in the context of toxicity assessment.

Chemical read-across commonly lacks information on the biological MoA, yet still relies on the premise that structurally similar compounds yield similar toxicological profiles. Nowadays, toxicity testing aims to employ the richness of omics data integration to establish an understanding of mechanisms underlying toxicity as an adverse outcome, and read-across studies are performed with toxicogenomics datasets such as TG-GATEs, DrugMatrix, Connectivity Map, and LINCS1000 (Serra et al., 2020). Here, we performed a systematic transcriptomics-driven biological similarity analysis using a panel of 18 structurally similar VPA analogues. By combining both physicochemical (i.e., structural similarity) and biological (i.e., transcriptomics) data, we found that indeed VPA-like compounds with a high similarity have a similar transcriptomic re-
Fig. 5: Stress response activation by carboxylic acids
A) Chosen transcription factors for stress pathways and number of genes filtered. Annotation of heatmap. B) Heatmap of stress response genes. Left bar indicates to which pathway a probe belongs. Euclidean clustering is applied per pathway.
spontaneous in overall gene expression as well as potency. In contrast, pivalic acid, with the lowest similarity score, had no response at all at the highest concentration studied, indicating far lower potency and/or different MoA relative to VPA. Compounds with a higher similarity score than pivalic acid but lower than octanoic acid and 2-propyl-4-pentenoic acid showed a varying potency in concentration responses on both gene FC levels as well as in the number of DEGs. Thus, there is no overall linear relationship between potency and structural similarity. We cannot exclude that these discrepancies are related to differences in toxicokinetic properties of the carboxylic acid analogues in PHHs. Regardless, despite the differences in potency, all carboxylic acids with a transcriptomics potency score > 628, based on the sum of all the DEGs across the concentration range, had a similar gene expression profile when comparing to top-50 genes upregulated by VPA.

Additionally, we established a transcriptomic gene-network fingerprint and found clear correlations among potent carboxylic acids. The highest gene module EG score was found for PHH:31, associated with PPAR signaling and acyl-CoA metabolism, both well-known processes in the mechanisms of VPA toxicity and steatogenesis (Silva et al., 2008; Szalowska et al., 2014). Indeed, PPAR signaling is a key event in the AOP for chemical-induced liver steatosis, just like inhibition of beta-oxidation, an essential pathway in acyl-CoA metabolism (Landesmann et al., 2012; van Breda et al., 2018). Integration of the S1500+ TempO-Seq panel with the TXG-MAPr turned out to be a high-throughput and cost-effective approach to derive quantitative biological read-across information on relevant toxicological responses. Indeed, the S1500+ gene set is a good alternative for whole transcriptome analysis (Callegaro et al., 2021). To corroborate this further, we compared this panel with whole transcriptome targeted sequencing and the TG-GATEs data and found that key genes responsible for module activation were indeed present in the S1500+ panel. Low correlation scores of the transcriptomic fingerprint of VPA measured by different numbers of genes (S1500+ vs whole transcriptome) can be explained by the lower EG scores of modules due to lower module gene coverage, since the module EG score is a combination score of log2FC for all genes that are present. Nevertheless, the module fingerprint for VPA and carboxylic acids was clearly determined by PHH:31, which also represented the expected biological interpretation. The user-friendly interface of the TXG-MAPr tool and the visualization of DEG data were ideal for revealing a shared mode-of-action, which contributes to hazard identification. Tools like the TXG-MAPr have a great potential to visualize biological similarities at the gene level and could be an integral part of similarity assessment. Here we showed that changes at the expressed genome level can help evaluate a shared MoA in a qualitative way and point toward differences in potency within structurally similar carboxylic acids. Nonetheless, for a complete read-across, which includes the derivation of a threshold value, one still requires additional readouts closer to the adverse outcome, as not all early events might progress to late key events (Escher et al., 2021). In this case, the accumulation of triglycerides in hepatocytes is, for example, a suitable late KE close to the adverse outcome liver steatosis.

In summary, we found that structurally related carboxylic acids had similar toxicological profiles, especially for compounds that show the highest structural similarity. Transcriptomics is one essential building block in biological similarity assessment and contributes to hazard identification. Tools like the TXG-MAPr have a great potential to visualize biological similarities at the gene level and could be an integral part of similarity assessment. When modules are more completely mapped to pathological outcomes, we may be able to better predict the adverse outcomes based on transcriptomics. Since safety sciences are moving away from animal studies in risk assessment, targeted transcriptomics is likely to become a significant tool for underpinning in vivo toxicity data. Mapping transcriptomics onto AOPs delivers a mechanistic understanding that is key to IATA.
References


Tollefsen, K. E., Scholz, S., Cronin, M. T. et al. (2014). Applying adverse outcome pathways (AOPs) to support integrated approaches to testing and assessment (IATA). Regul Toxicol Pharmacol 70, 629-640. doi:10.1016/j.yrtph.2014.09.009


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

Acknowledgements
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 681002 (EU-ToxRisk), the Innovative Medicine Initiative TransQST project (grant No. 116030), eTRANSAFE project (grant number 777365), and the Ontology project supported by the Cosmetics Europe and Cefic research programmes.