Grouping of UVCB Substances with Dose-Response Transcriptomics Data from Human Cell-Based Assays

John S. House[1,8], **Fabian A. Grimm**[2], **William D. Klaren**[2,9], **Abigail Dalzell**[3], **Srikeerthana Kuchi**[4,10], **Shu-Dong Zhang**[4], **Klaus Lenz**[5], **Peter J. Boogaard**[6], **Hans B. Ketelslegers**[7], **Timothy W. Gant**[3], **Ivan Rusyn**[2], **Fred A. Wright**[1]

[1]Bioinformatics Research Center, North Carolina State University, Raleigh, NC, USA

[2]Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX, USA

[3]Public Health England, Centre for Radiation, Chemical and Environmental Hazards, Harwell Science Campus, Oxon, UK

[4]Northern Ireland Centre for Stratified Medicine, Ulster University, L/Derry, Northern Ireland, UK

[5]SYNCOM Forschungs und Entwicklungsberatung GmbH, Ganderkesee, Germany

[6]SHELL International BV, The Hague, The Netherlands

[7]Concawe, Brussels, Belgium

[8]current address: Biostatistics & Computational Biology Branch, National Institute of Environmental Health Sciences, RTP, NC, USA

[9]current address: ToxStrategies, Inc., Asheville, NC, USA

[10]current address: MRC-University of Glasgow Centre for Virus Research, Glasgow, Scotland, UK

Abstract

The application of *in vitro* biological assays as new approach methodologies (NAMs) to support grouping of UVCB (unknown or variable composition, complex reaction products, and biological materials) substances has recently been demonstrated. In addition to cell-based phenotyping as NAMs, *in vitro* transcriptomic profiling is used to gain deeper mechanistic understanding of biological responses to chemicals and to support grouping and read-across. However, the value of gene expression profiling for characterizing complex substances like UVCBs has not been explored. Using 141 petroleum substance extracts, we performed dose-response transcriptomic profiling in human induced pluripotent stem cell (iPSC)-derived hepatocytes, cardiomyocytes,

Correspondence: Fred A. Wright, PhD 1 Lampe Drive, CB#7566 Raleigh, NC, 27695, USA (fred_wright@ncsu.edu).

neurons, and endothelial cells, as well as cell lines MCF7 and A375. The goal was to determine whether transcriptomic data can be used to group these UVCBs and to further characterize the molecular basis for *in vitro* biological responses. We found distinct transcriptional responses for petroleum substances by manufacturing class. Pathway enrichment informed interpretation of effects of substances and UVCB petroleum-class. Transcriptional activity was strongly correlated with concentration of polycyclic aromatic compounds (PAC), especially in iPSC-derived hepatocytes. Supervised analysis using transcriptomics, alone or in combination with bioactivity data collected on these same substances/cells, suggest that transcriptomics data provide useful mechanistic information, but only modest additional value for grouping. Overall, these results further demonstrate the value of NAMs for grouping of UVCBs, identify informative cell lines, and provide data that could be used for justifying selection of substances for further testing that may be required for registration.

1 Introduction

Substances classified as UVCBs (unknown or variable composition, complex reaction products, and biological materials) comprise over 20% of chemical registrations in Europe and present difficult challenges for hazard and risk evaluations (ECHA, 2017). Petroleum substances are UVCBs with a complexity that arises primarily from the presence of very large – hundreds to as many as millions – numbers of isomeric chemical constituents. The physicochemical processes during oil refining are complex, and varying sources of crude oil are used at different times in manufacturing facilities. For this reason, petroleum substance groupings and CAS numbers are typically based on physicochemical properties and performance characteristics rather than chemical characterization of the constituents (Salvito et al., 2020), albeit petroleum substances are made up of a few classes of hydrocarbons (alkanes, iso-alkanes, cyclo-alkanes and (poly)aromatics), and the actual chemical variation is highly determined by the physicochemical properties. Current practice to harmonize the identification of potential hazards among petroleum UVCBs is based on broad product categories (CONCAWE, 2020) that are largely informed by the product performance criteria, manufacturing processes, and the presence of polycyclic aromatic compounds (PAC) and other potentially hazardous constituents (Clark et al., 2013; McKee et al., 2015). However, regulatory agencies question the application of groupings and read-across for UVCB product categories due to insufficient justification for considerations of chemical or biological sameness of the products in each category (ECHA, 2020). Therefore, alternative data streams have been proposed as potentially relevant for supporting grouping of petroleum UVCBs to improve chemical (Grimm et al., 2017; Roman-Hubers et al., 2021; Onel et al., 2019) and hazard characterization (House et al., 2021), and to ultimately reduce and refine the need for new animal testing for registration of these products.

Several previous studies tested the hypothesis that grouping of complex substances and environmental mixtures can be achieved using data from *in vitro* assays in induced pluripotent stem cell (iPSC)-derived and other cell types (Grimm et al., 2016, 2019; Chen et al., 2020). This previous work demonstrated that despite the inherent complexity and the variability of samples, *in vitro* data can be used to discern informative biological patterns corresponding to chemical composition or manufacturing categories. In addition,

these studies suggested that the diversity and physiological relevance of the data from studies in only a few iPSC-derived cell types such as hepatocytes (Grimm et al., 2015) and cardiomyocytes (Burnett et al., 2019) can yield sensitive multi-dimensional information to aid in grouping, providing a strong rationale and basis for future read-across efforts and prioritization of substances within manufacturing categories.

More recently, House et al. (2021) combined all these aspects in a comprehensive effort to investigate 141 substances, a compendium of samples comprising the majority of petroleum-based UVCBs registered under the Regulation on Registration, Evaluation and Authorisation of Chemicals (REACH) in the European Union (CONCAWE, 2019). That study generated *in vitro* bioactivity data from 15 human cell types as new approach methodologies (NAM) data to support substance grouping into 16 major categories of petroleum-based UVCBs. Extensive quality control was used to determine assays, including those specific to cell type, that were most informative and provided discernible dose-response relationships. The outcomes of this study showed that overall summaries of bioactivity yielded substance rankings concordant with their chemical composition and expected hazard potential as obtained from physical and analytical chemistry data. Moreover, unsupervised and supervised analyses suggested that the bioassay data provided important additional information relevant to the substance categorization; bioassay data alone appeared as informative to this categorization as traditional physicochemical data.

These observations were critical in identifying the most informative cell types and bioassays, providing potential cost savings in future studies. It is important to recognize that the ability of *in vitro* bioassays to provide relevant information may have only a partial relationship to *in vivo* relevance and health risk (ECHA, 2020). For example, data from iPSC-derived cardiomyocytes were among the most sensitive with respect to their ability to provide concordance with a manufacturing category (House et al., 2021). When aggregating across several *in vitro* data types, iPSC-derived hepatocytes showed a strong concordance between overall activity and PAC content, a known *in vivo* indicator of potential human health hazard (McKee et al., 2015). These observations further support the use of *in vitro* bioactivity assays as providing a potential framework for prioritization of substances within manufacturing categories but also highlight the need for additional mechanistic, such as gene expression, evidence of the effects to enable *in vivo* translation.

Overall, the previous results have strongly supported the utility of *in vitro* NAM for interrogating and grouping complex substances, including petroleum-based UVCBs. The eventual purpose of these groupings is to support read-across, and mechanistic evidence for the nature of perturbations underlying the bioassay phenotype responses to chemicals is gaining prominence in decision-making (Samet et al., 2020). In this context, we expect these data to inform read-across hypotheses and prioritize substances for *in vivo* testing. A refined understanding of the chemical effects at the intracellular level could potentially provide more informative *in vitro* models to characterize UVCBs, reducing the number of cell types and assays even further while retaining the ability to prioritize UVCBs for further testing within categories and to serve as a basis for read-across. The use of *in vitro* transcriptomic profiling of UVCBs is a natural step to provide this biological context. Indeed, the use of transcriptomics in toxicology is now well-established (Joseph, 2017), and this data stream

is an increasingly popular NAM (Harrill et al., 2019). High-throughput transcriptomics has been used to interrogate biological effects of a large number of chemicals and perform transcriptomics dose-response analyses (Harrill et al., 2021; House et al., 2017). Previous high-throughput transcriptomic data of some petroleum-derived UVCBs showed that *in vitro* gene expression changes were specific to broad categories (e.g., heavy fuel oils vs. straight run gas oils) (Grimm et al., 2016).

In this study, we tested the hypothesis that transcriptomic profiles can be used to support grouping of petroleum substances and provide informative mechanistic data for existing groupings based on manufacturing class. For this, 141 petroleum substances, previously analyzed using extensive bioassays across a multitude of human cell types, were used as representative UVCBs. These substances were interrogated using transcriptomic profiling in six cell types in a dose-response design.

2   Materials and methods

Chemicals

All chemicals used in these studies, except for petroleum substances, were obtained from Sigma-Aldrich (St. Louis, MO), unless otherwise noted. Samples of petroleum substances were supplied by Concawe (Brussels, Belgium). To enable *in vitro* studies of petroleum substances, extraction of petroleum substances into dimethyl sulfoxide (DMSO) was performed using American Society for Testing and Materials standard procedures (ASTM International, 2014). The DMSO extraction used herein was designed to concentrate the "biologically active" fraction (i.e., mostly 3–7 ring PAC, but also other polar constituents) of each petroleum substance; the extracts obtained using this method are used routinely for safety testing (e.g., mutagenicity) and chemical characterization of the refinery streams (CONCAWE, 1994). Briefly, 4 g of each tested petroleum substance (Tab. 1) was first dissolved in 10 mL of cyclohexane; 10 mL of DMSO (Fisher Scientific, Waltham, MA) was added, and the mixture was vigorously shaken for several minutes. The DMSO layer was removed using a glass pipette, and the cyclohexane was re-extracted with an additional 10 mL of DMSO. Both PAC-enriched DMSO layers were combined and diluted 2:1 with two volumes of 4% (w/v) sodium chloride solution. Following subsequent extraction with 20 mL and 10 mL cyclohexane to isolate the PAC fraction, the organic layers were washed twice with distilled water and filtered through anhydrous sodium sulfate. The procedure of extraction was also performed without addition of the petroleum substances, and the resulting fraction was designated as "vehicle" (method blank) to be used as a reference for comparisons. Petroleum substance extracts were further diluted to enable concentration-response testing.

Study design

Overall, this study conducted concentration-response over 4 points (3 serial 1-$\log_{10}$ dilutions of each extract performed in duplicate, and ~45 vehicle controls for each extract). All samples were aliquoted into 384-well "master" plates (Masterblock 384-well, V bottom, Deepwell polypropylene plate; Cat. No. 781271; Greiner Bio-One North America, Monroe, NC) as detailed elsewhere (House et al., 2021). Plates were sealed with aluminum film and

stored at −80°C until use. Copies of each master plate were prepared for use in all *in vitro* experiments. The final concentration of DMSO in assay wells following addition of test substances was 0.25–0.5% (v/v), depending on the cell type, see House et al. (2021) for details.

Cell types

A total of 6 human cell types were used in these experiments (Tab. S1[1]). Cell type and vendor selections were based on the following considerations. Cells were chosen to be of human origin and to represent diverse organs/tissues. We used both iPSC-derived cells as well as established cell lines. These *in vitro* models had to be reproducible (i.e., a particular cell/donor can be obtained from a commercial source) and suitable for evaluation of both "functional" and "cytotoxicity" endpoints so that the specificity of the effects of test compounds could be assessed. Four of these cell types (hepatocytes, endothelial cells, neurons, and cardiomyocytes) were human iPSC-derived (FujiFilm-CDI, Madison, WI). Two cell types (A375 malignant melanoma cells and MCF7 breast cancer cells) were from ATCC (Manassas, VA). All cells were cultured as detailed elsewhere (House et al., 2021), and additional cell culturing information is given in the supplementary information[1]. Cells were plated in 384-well plates in densities recommended by the supplier, using optimized media supplied by the same company or optimized for density by experimentation for each cell line. Cells were cultured without treatment for a period of time required to achieve functional capacity. Plating density, cell culture conditions and duration are detailed elsewhere (House et al., 2021). Cells were treated with test substances in a series of dilutions to evaluate concentration-response as described above.

Transcriptomics – Quality control

Overall plate design was explained previously in detail (House et al., 2021). In brief, the transcriptomic experiments were treated similarly but with only 6 cell types (iCell hepatocytes, iCell cardiomyocytes, iCell neurons, iCell endothelial, MCF7, and A375) and without the highest dose (undiluted extracted UVCBs) that elicited cellular toxicity for some cell type/treatment combinations in phenotypic assays. Raw sequenced reads were aligned and counted using the pipeline developed by our group (House et al., 2017). Samples with < 100K counts were removed, as were probes not expressed across at least 5% of the sample space. For the few genes with more than 1 probe, counts were summed to the gene level. At the gene level, these steps resulted in ~2,500 genes per cell type assessed for differential gene expression (DEG) and concentration response (CRG) for each of 141 UVCBs. Each plate contained three types of controls: media alone – "media", media with DMSO – "DMSO", and method blanks – "vehicle". See House et al. (2021) for more detail. Within a cell type, these three types of controls were examined for correlation between each other (across the transcriptomic space), and controls whose correlation exceeded 3 standard deviations from mean correlation were removed as outliers. All three control types were examined with principal component analysis (PCA) of the top 500 expressed genes. PCA scatter plots (Fig. 1) revealed little difference between vehicle and DMSO, and thus vehicle controls were used as the most appropriate reference for all subsequent data analyses. For

---

[1]doi:10.14573/altex.2107051s

this publication, after quality control (QC), vehicle controls included 40–48 replicates (23–24 for iCell cardiomyocytes), providing a solid anchor for dose-response quantification followed by treatment dose response replication of the remaining serial dilutions. We highlight that this approach, applied over multiple doses, six cell lines, and ~2,500 genes provides considerable resolution for analyses of substances used in this study. In this study, the primary unit of observation was not a gene, but a substance. Each substance was interrogated over multiple doses and cell types, with nearly 90,000 expression data points, providing substantial replication for the observations of interest.

Transcriptomics – Differential gene expression

For each cell type, the complete normalized count matrix of 141 UVCBs and vehicle controls was calculated using DESeq2 (Love et al., 2014). The maximum dose was compared to vehicle controls for each cell type/treatment/gene combination, and log2-fold-change values and p-values were calculated with an equal variance assumption. False discovery q-values were calculated from the p-values of each vector using the R v4.1 p.adjust function (method="BH") of roughly 2,500 genes for a given UVCB treatment/cell type combination; and FDR 0.1 was considered as a cut-off to identify expression-responsive gene(s) for a given cell type/UVCB combination. Normalized counts from DESeq2 were subsequently used for examination of dose-response criteria.

Transcriptomics – Concentration response

We applied a set of statistical flags described in detail in House et al. (2017) to assess whether a concentration-response calculation was warranted for each cell type/gene/treatment combination. Linear modeling using DESeq2 (Love et al., 2014) was then conducted of $\log_2(\text{count} + 0.5)$ transformed counts against $\log_{10}(\text{concentration})$ for each retained cell type/treatment/gene. In all cases, the control concentration was converted to the lowest treatment concentration divided by 10. For example, if the three concentrations given were 1/10 stock, 1/100 stock, and 1/1,000 stock, the vehicle control concentration was considered as 1/10,000, and concentration was coded as a predictor vector of −4, −3, −2, and −1 after $\log_{10}$ transformation. For a given treatment and cell type, false discovery q-values for linear trend (positive or negative) for concentration were calculated across all assessed genes as described above for a given cell type/treatment, and genes with FDR 0.1 were considered as concentration responsive.

Pathway analysis

Pathway analysis and visualization was conducted in R with the xgr package (Fang et al., 2016) version 1.1.8 using Reactome Ontologies as gene sets. For both differentially expressed genes (DEGs) and concentration responsive genes (CRGs), an FDR of 5% was used to conduct pathway analysis. The gene set background was considered to consist of all TempOSeq-interrogated genes retained after low-count removal for a given cell type. Subsequently, pathway enrichment was conducted against the C2Reactome ontology sets.

Polycyclic aromatic compound (PAC) analysis

Weight percentages of PAC in all tested samples were determined by gas chromatography-coupled mass selective detection (GC/MSD) as detailed previously (Roy et al., 1988). Briefly, each substance was extracted as detailed above and dried. The amount of each extract was then determined using the weight difference of the empty flask and following solvent evaporation. The extract was then dissolved in cyclohexane to a final concentration of 50 mg/mL and used for analytical assays. Sample separation was achieved on a Zebron-5HT capillary column (30 m; 0.25 mm; 0.25 mm; Phenomenex, Torrance, CA). Quantitative integration of the chromatograms was achieved using standards of naphthalene, phenanthrene, 1,2-benzanthracene, benzo[a]pyrene, bebenzo[g,h,i]perylene, and coronene. The resulting PAC profiles consist of weight percentages by ring number and are listed elsewhere (House et al., 2021).

Establishing correspondence to bioactivity

The focus of this study was to characterize changes in gene expression in response to treatment, and comparison across UVCB categories, and not the bioactivity measures that were the subject of the earlier report (House et al., 2021). However, to establish the relevance of the results with the current design and the effect of sampling variation, we computed overall summed bioactivity across the 42 assays from the 15 cell types reported earlier (House et al., 2021). Summed bioactivity was the summed values over the 42 assays, where each bioactivity phenotype was standardized to a common unit variance, and large values correspond to high activity throughout. As the unit of study is each of the 141 UVCBs, we ran a cross-validated regression tree model for predicting bioactivity using gene responsiveness to treatment as a predictor for all ~2,500 TempOSeq genes, where the model was trained within each of the cell types. For regression trees, we used the *xgboost* R package (v 04.4) with default settings, and leave-one-out cross-validation to obtain predictions without overfitting. Prediction accuracy was recorded as the Pearson correlation r between true and predicted summed bioactivity. We interpret high correlations between predicted and actual bioactivity as indicative of high experimental reproducibility in this NAM system, even when the ground truth of biological effects is unknown.

Supervised category analysis

As detailed in House et al. (2021), we trained a machine-learning statistical model to predict the existing categories (CONCAWE, 2019, 2020) of petroleum-based UVCBs under REACH (Tab. 1). As iCell hepatocytes appeared to be the most sensitive cell type in overall gene expression changes in this experiment, we performed analysis using the Prediction Analysis of Microarrays (PAM) package in Rv3.6 (Tibshirani et al., 2002) using iCell hepatocyte expression data for the 141 UVCBs, along with the 8 PAC and 42 bioassay phenotypic measurements from House et al. (2021), using (a) expression data alone, (b) expression + PAC, (c) expression + bioassay measurements, and (d) expression + PAC + bioassay measurements. Leave-one-out cross-validation was performed due to the small minimum category sizes in some instances, and we used a PAM shrinkage threshold of $z = 1.28$ for construction of the classifier. For each instance, we computed two measures of classification accuracy: the exact matching accuracy (proportion of matches of cross-

validated category assignment vs. true assignment) from the associated confusion matrix, and the proportion correctly assigned to one of two major hazard groups. For the latter, we used an ordering of the categories in comparison to PAC to establish a group cutpoint (between bitumens and base oils), as shown below in Section 3. For the exact match criterion, the 95th percentile of null permutations was computed in House et al. (2021) as 0.163 and used for statistical significance testing.

3 Results

Previous studies of petroleum-derived UVCBs have shown that combinations of analytical data (e.g., PAC content) and cell-based bioassays can be used to group, categorize, and largely recapitulate manufacturing-based classifications (House et al., 2021; Grimm et al., 2016). This study further extends knowledge on the application of *in vitro* data for grouping of complex substances by including transcriptional response measures for a variety of cell types in a dose-escalation design. While it is well established that transcriptomics provides direct insight into underlying mechanisms of response (Harrill et al., 2021; Yauk et al., 2020), the sensitivity and specificity of using expression profiles for grouping of complex UVCBs, in comparison with previous bioactivity measures, has not been explored.

The overall schematic of the study design and data processing pipeline are shown in Figure 1. Six cell types (iCell hepatocytes, iCell cardiomyocytes, iCell neurons, iCell endothelial, MCF7, and A375) were subjected to the 4-point treatment (3 dilution concentrations and controls) with 141 UVCB extracts as described in Section 2, performed in duplicate for each dilution, and compared to ~45 vehicle controls. The TempO-Seq probe sets used herein interrogate the transcription of ~2,900 (~2,500 after collapsing to gene level and low-count removal) expressed transcripts, and the entire experiment provided ~28 million expression data points. The TempO-Seq preprocessing and analysis closely followed the pipeline of House et al. (2017), producing raw sequencing counts that were then subjected to quality control as described in Section 2, followed by normalization by DESeq2 (Love et al., 2014). Differential expression analysis via DESeq2 was performed using both logarithmic fold-change of maximum dose vs. control ("$\log_2$ FC"), as a simple robust contrast, and a concentration-response trend test using controls and all concentrations, which was expected to be more powerful for monotonic dose-response relationships (Leuraud and Benichou, 2001). Finally, data visualization and various summaries and pathway analyses were used to interpret the biological context.

PCAs of thousands of genes provide a rich empirical visualization environment to examine the overall gene expression pattern relationships in vehicle, media, and DMSO controls (Fig. 1B). In most of the six cell types, the control gene expression patterns are overlapping, suggesting little difference among the control types. This relatively even mixing occurred even when some variational patterns emerged, e.g., for A375s with a small portion of outlying controls, and in the elongated pattern emerging from endothelial controls. For iCell cardiomyocytes, the media controls appear to be somewhat different from vehicle and DMSO. However, our later conclusions are supported across multiple cell types, and we concluded that vehicle controls are appropriate for differential expression analysis. Table S1[1] shows the average number of reads per expressed transcript for each cell type and

treatment, including the three control types. The combinations were roughly comparable in average sequence counts, except that iCell neurons had relatively lower average counts (~300 reads per transcript/probe vs. 500–700 for other cell types).

Differential expression analysis was performed on the 141 substances for the six cell types. To judge overall cell type-specific transcriptional responses, we recorded the numbers of DEGs for the fold-change analysis (based on fold change and false discovery cut-offs). iCell hepatocytes showed the most DEGs (~2%–3% in each direction), followed by cardiomyocytes and endothelial cells (Fig. 2A). The number of significant Reactome pathways (across all substances) followed a similar pattern, with ~75 significant pathways for hepatocytes. Overall, as expected, a somewhat larger number of CRGs were significant ($q < 0.1$), and here endothelial cells showed many more CRGs (Fig. 2B), while cardiomyocytes showed more significant Reactome pathways. In aggregate, hepatocytes, cardiomyocytes, and endothelial cells appeared to be the most dose-responsive cell types across the substances. Hepatocytes, in particular, demonstrated consistent effects on gene expression, and we use this cell type as an exemplar for several of the main figures in this manuscript. Figure 2C shows individual genes in hepatocytes that had the largest number of instances of differential and concentration-responsive outcomes ($q < 0.1$) across the 141 substances. The top five DEGs were *Cyp1A1*, *UGT1A10*, *Cyp1B*, *CDH2*, and *UGT1AB* (increasing with dose in most substances). The effects on these genes were highly consistent, appearing as significant in ~50% of the substances. The same data shown in Figure 2C are shown for all cell types in Figures S1 and S2[1].

Next, we considered the number of DEGs and CRGs for each substance in each pre-defined manufacturing category. These values are displayed in Figure 3 after aggregating across all cell types (left panels), with a category ordering based on the mean DEG/CRG ranking across all cell types and the categories displaying the most transcriptomic perturbations at the top. Note that a few substances did not show DEG/CRGs after correction for multiple comparisons. The results for hepatocytes are shown in the right panels and are somewhat more variable, both within and across categories, as they are based on a single cell type. However, the correspondence with the bioactivity-based ordering is still apparent, with some exceptions. For example, bitumens, SRGO, and petrolatum appear to have a relatively greater number of DEGs in hepatocytes than might be expected from the all-cell types results. The overall results and ordering show that transcriptomic responsiveness is similar to prior knowledge of manufacturing category, based on analytic properties and previously measured bioactivity (House et al., 2021).

Pathway analyses for enrichment of DEGs in Reactome pathways was performed within each cell type and for each manufacturing category. The results for iCell hepatocytes are shown in Figure 4A for DEGs, expressed as the enrichment fold change by the xgr package. The most commonly perturbed pathways, across multiple categories, included "unfolded protein response," "metabolism of proteins," "diabetes pathways", and "biological oxidations." Another set of related pathways appeared, involving fatty acid metabolism, cholesterol biosynthesis, and PPARα activation. We note that our approach to pathway enrichment requires a contrast between significant genes and the remaining background interrogated set of genes, so that for example the RAE category shows few

significant pathways although the number of DEGs is large as previously shown. Similar patterns and results occurred for the CRGs (Fig. 4B).

The gene expression effects for individual substances, organized by category, are shown in Figure 5A (DEGs) and 5B (GRGs), displaying the number of significant genes per substance ($q < 0.1$). Again, among the cell types, hepatocytes, cardiomyocytes, and endothelial cells showed the greatest effects of UVCBs on gene expression. Direction of expression effects (increase/decrease with increasing concentration) were approximately balanced, except for endothelial cells, which displayed a preponderance of downregulated genes. The number of significant CRGs was generally higher than for DEGs, which we attribute to the greater power of the trend analyses in which all concentrations are used.

Heterogeneity of differential gene expression patterns can also be viewed by considering, for each gene, the number of substances within the manufacturing category in which the gene was significant. Figure 6A shows these results for DEGs (q 0.1) in iCell hepatocytes, with genes ordered according to decreasing number of times perturbed across all substances. The height of each bar represents the proportion of times that gene was perturbed for the given category or class. For manufacturing categories with large numbers of significant genes (such as HFOs), some groups of genes are differentially expressed in most substances within the category. For CRGs (Fig. 6B), the patterns are even clearer, with the plots appearing "denser" due to the larger number of significant genes. Due to the uniform ordering of genes across all test substances, an approximate concordance can be discerned among the categories of high bioactivity (lower on the plots), with the leftmost genes showing the greatest evidence of differential expression as a common feature across multiple categories. For categories of lower bioactivity, the genes showing the most evidence of differential expression are dispersed more evenly throughout, as exemplified by bitumens and the kerosene substances.

In a manner similar to results presented in House et al. (2021), we reasoned that substances with higher content of 3–7 ring PAC may elicit more prominent gene expression changes. For each cell type, the number of DEGs and CRGs was compared to the PAC 3–7 content across the 141 UVCBs. Among the cell types, iCell hepatocytes showed the highest correlation (Spearman $\rho = 0.77$) with DEGs (Fig. 7A) and CRGs (Fig. 7C). Subplots in Figures 7B and 7D show the results within each manufacturing category. It is notable that the positive relationship is discernable even within manufacturing category, provided the category spans a sufficient range of PAC content, as can be observed for HFOs and CGOs. The results of this analysis provide an anchoring to a known aspect of substance hazard for petroleum UVCBs (McKee et al., 2015, 2018). Moreover, the results provide a clear criterion for which cell type, in this case iCell hepatocytes, might be selected for future investigation over other cell types. PAC correlations (Spearman and Pearson) and p-values for each cell type and either DRGs or CRGs are summarized in Table 2.

Finally, we investigated the extent to which machine learning models can be trained to recognize features that are representative of a manufacturing category. Although bioactivity as reported in House et al. (2021) is not the primary subject of this report, the ability of these data to support machine learning analyses can be initially motivated by comparison

of expression to bioactivity. A quantitative summary of the 42 bioactivity predictors (House et al., 2021) was used as a response for a cross-validated regression tree model using gene expression within each cell type (see Methods). The resulting correlations for prediction vs. observed bioactivity were: A375 ($r = 0.81$), iCell cardiomyocytes ($rc = 0.61$), iCell endothelial ($r = 0.76$), iCell hepatocytes ($r = 0.84$), MCF7 ($r = 0.68$), and iCell neurons ($r = 0.35$).

These high correlations in these results support the potential informativeness of these data for categorization and further support iCell hepatocytes as the most informative cell type investigated for expression. Such a "supervised" analysis of categories can potentially provide information about the uniformity of substances in a manufacturing category, as well as highlight substances that are difficult to group within the category and are therefore a priority for future testing or to serve as prioritized group representatives. Such substances may be difficult to identify with unsupervised analyses, which use all the features available (e.g., thousands of genes) and where uninformative feature variation can overwhelm the inference. For this analysis, we used a software originally designed for gene expression class prediction (Tibshirani et al., 2002), but which can use any quantitative predictors. We used the expression of 2,388 expressed genes in various combinations with the 8 analytic (i.e., PAC) and 42 bioactivity predictors previously described in House et al. (2021) for these 141 substances. The results are shown in Figure 8 using hepatocyte expression. In cross-validated analyses, "exact" matches of the model predictions to actual manufacturing category were 31% for expression alone, 35% for bioactivity + expression, 40% for PAC + expression, and 40% for PAC + bioactivity + expression. Although these values are lower than 50%, they are highly significant in comparison to the null 95% permutation threshold of 16%. Closer examination of the results by manufacturing category ordered by bioactivity (Fig. 8) showed much higher classification accuracy (ranging from 87% to 89%) when grouping substances with similar hazard potential. Here, transcriptomic data appears to provide relatively modest improvement compared to analytic or bioactivity analyses (House et al., 2021), which we attribute to the difficulty in gene expression feature selection from among thousands of genes and to the fact that saturated hydrocarbon constituents that were enriched in the extracts used herein may not elicit substance-specific gene expression changes. Using only expression for prediction leads to more "clumping" of prediction into the large categories such as HFOs and BOs, while the combination of predictor types spreads these predictions across other categories, a phenomenon that can be seen when comparing the diagonals of the upper left and lower right panels.

4   Discussion

Assessing the potential human health hazard of UVCB substances and, more specifically, defining a targeted testing strategy that will assist in refining and reducing animal testing is a particularly challenging problem in regulatory decision-making. Hazard characterization based on individual components of complex substances is largely intractable, and limited animal testing data are available for risk characterization. However, ethical and economic considerations indicate a critical need to reduce animal testing (Herrmann et al., 2019); therefore, *in vitro* test-based NAMs are under active consideration as the future of risk assessment (Kavlock et al., 2018). Indeed, much work is being performed to collect and

catalogue *in vitro* test data on thousands of chemicals (Williams et al., 2017) as well as to demonstrate how they can be used in support of regulatory decision-making (Paul Friedman et al., 2020; Berggren et al., 2015; Chiu et al., 2018; Escher et al., 2019). Far less NAM data is available for complex substances, such as UVCB or environmental mixtures (Drakvik et al., 2020; Bopp et al., 2019).

Previously, we examined the utility of NAMs for grouping of complex UVCBs (House et al., 2021); 141 petroleum-based UVCB substances were grouped based on their biological responses from cell type-specific assays across 15 human cell types. Here, we assessed the informativeness and ability of the transcriptomic data across 6 of these human cell lines to add further mechanistic information to the grouping of these substances. Transcriptomics was among the first omics data types to be used for classification and prediction of hazards and risks of drugs and environmental chemicals (Ganter et al., 2005; Uehara et al., 2010; Waters et al., 2008). While some of the early, over-optimistic forecasts about the value of transcriptomic data for toxicity prediction did not materialize, it has been proposed that these data be routinely collected in toxicology studies and used in risk-based evaluations (Yauk et al., 2020; Liu et al., 2019). Among the most notable developments that support transcriptomics data as a "screening" NAM is the opportunity to conduct high-throughput experiments that interrogate multiple cell-based models and can test for concentration-response in gene expression (House et al., 2017; Phillips et al., 2019). This approach represents a path forward in decision-making, as compared to the traditional use of transcriptomic data to provide mechanistic evidence (Harrill et al., 2019, 2021). Transcriptomic data are high-dimensional and provide a comprehensive set of information on the state of the cells or tissues in both health and disease; this information has been exploited to not only classify individual chemicals with respect to their potential hazard (Ganter et al., 2005; Uehara et al., 2010), but also to group chemicals based on the similarity in their effects (Low et al., 2011; De Abrew et al., 2016, 2019), one of the justifications for grouping and read-across in the regulatory context (Schultz et al., 2015). Finally, we emphasize that transcriptomics provides biological context due to the very nature of genomic annotation that other NAMs (e.g., based on bioactivity as in House et al., 2021) do not so readily provide, and the high adoption rate and standard use of transcriptomics (Yauk et al., 2020; Liu et al., 2019) make it more attractive than the use of NAMs that require highly specialized methods.

In our study, probes from the TempO-seq s1500$^+$ gene set (n = 2,982 optimized for human pathway coverage and representative of the human transcriptome) were assessed in replicate across a 4-point concentration dose response. We examined both differential gene expression response at the highest treatment dose compared to controls as well as concentration response effects across all concentrations. The inclusion of gene expression profiling in this study, in addition to the biological response data from the high content *in vitro* screening generated earlier (House et al., 2021), provide (1) biological context to the challenge of grouping of UVCBs with cell-based data, and (2) additional clues as to the suitability of various cell types for screening and prioritization.

For the first challenge, we note that the imaging-based phenotypes on these UVCBs (House et al., 2021) were considered by regulators as insufficient for the purpose of supporting the

similarity argument in product registration (ECHA, 2020). It was noted that the relationships between *in vitro* results and *in vivo* effects of these substances needed to be clarified, especially if the registration argues that PAH constituents present the most bioactive, or worst case, fraction of the test substance. Indeed, assessment of DEGs and CRGs across the cell-line/UVCB treatment space revealed additional mechanistic information. For example, in agreement with the hypothesis that PAH are eliciting the majority of transcriptional changes, iPSC-derived hepatocytes were among the most responsive to UVCB-elicited transcriptomic alterations for both DEGs and CRGs across these 141 petroleum substances, with transcriptomic changes observed in active xenobiotic metabolizing genes. Genes most often up-regulated in response to petroleum UVCBs were downstream of nuclear receptor-activated transcription of xenobiotic metabolism genes (e.g., cytochrome P450s and UDP-glucuronosyltransferases) in response to hydrocarbons (Goedtke et al., 2020). Consistent with previously published results (Grimm et al., 2016), the wax and petrolatum categories elicited the fewest transcriptomic changes, while aromatic extracts elicited the strongest response. These results were apparent not only across a summary of all 6 cell types but were also largely replicated in hepatocytes alone. Thus, these data are informative with respect to the second challenge and the design of future *in vitro* experiments for testing petroleum UVCBs, because our results suggest that iPSC-derived hepatocytes may be an especially useful cell type for profiling complex substances that contain PAC 3–7 constituents. This effect was more apparent for DEGs (Spearman's $\rho = 0.77$ for the aggregate measure vs. PAH 3–7) than for CRGs (Spearman's $\rho = 0.57$). The strength of these relationships, using gene expression and iCell hepatocytes alone, is similar to that observed ($\rho = 0.81$) for a summary of 4 bioactivity assays in iCell hepatocytes (House et al., 2021).

Various aspects of within- and between-category gene expression changes, as well as variation across cell types, have been explored in this study. Many of these differences are apparent in the "experimental expression fingerprints" provided in Figure 5. More highly significant changes are evident in the concentration response analysis than the differential expression (max. dose vs. vehicle control) analysis, which we attribute to increased power in the use of all the concentration data. The data from iCell hepatocytes exhibited similar proportions and distributions of up- and downregulated DEGs/CRGs, increased expression changes in the fuel oil and aromatic extract categories, within-category response-heterogeneity, and nearly absent gene expression changes within the wax and petrolatum categories. The "gene fingerprints" exhibited in iCell hepatocytes in Figure 6 indicate that for many categories that elicit an overall modest transcriptomic response (e.g., base oils; BO), the genes that are differentially expressed across numerous within-category substances tend to be the same as those identified in categories eliciting a stronger response (e.g., HFOs). Collectively, for petroleum UVCBs, we argue that our data are supportive of hepatocytes as the most suitable cell type for screening if a single cell type is used. For other types of complex substances, we recommend that several cell types still be examined (Chen et al., 2020, 2021; Hsieh et al., 2021), but our general approach may serve as a useful model for such investigations.

Our results on supervised grouping of substances into petroleum substance-specific categories indicates that expression patterns can be used to confirm assignments of individual substances into categories with an accuracy that is much higher than chance

(for exact category matches, 38% vs. 16%), but that this classification is more effective if performed using a combination of expression with other bioactivity/analytic data that had been reported previously (House et al., 2021). As noted earlier, one difficulty with expression-based classification is posed by the large number of potential classifiers, which can produce overfitting and reduce cross-validated accuracy. It is also worth noting that existing categorization is based on manufacturing processes that may have an imperfect relationship with biological response. For example, while the correlation between bioactivity and expression vs. PAH 3–7 content is relatively high (House et al., 2021) (Fig. 7), the within-category range for PAC values is large, and categories are not monolithic. Thus, the accuracy of the transcriptomic and *in vitro* assays in general for supervised classification may have upper bounds that do not reflect on the assays themselves.

One additional consideration with respect to the study design and the potential use of these data in support of testing proposals and read-across hypothesis concerns the choice of the DMSO extraction (ASTM International, 2014) to enable *in vitro* testing of petroleum UVCBs. This method preferentially extracts 3–7 ring PAC, but the regulators have noted that "*testing DMSO extracts does not provide a basis for reliably predicting the properties of the [whole] substance*" (ECHA, 2020). Unfortunately, testing the material which is left after DMSO extraction presents numerous challenges because the solvents that would need to be used are incompatible with *in vitro* testing. Alternative delivery of the "whole substance" can be achieved through passive dosing (Hammershoj et al., 2020; Trac et al., 2021); however, the methods to deliver complex UVCBs in small volume *in vitro* testing conditions have not been established yet. Thus, additional work is needed to improve the relevance of *in vitro* test methods for use in regulatory decisions on UVCBs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Data availability

MIAME-compliant expression data and experimental metadata have been uploaded to the public repository Gene Expression Omnibus (GSE186121).

References

ASTM International (2014). Standard Test Method for Determining Carcinogenic Potential of Virgin Base Oils in Metalworking Fluids. E1687-10. ASTM International. doi:10.1520/E1687-10 (accessed 01.07.2019).

Berggren E, Amcoff P, Benigni R et al. (2015). Chemical safety assessment using read-across: Assessing the use of novel testing methods to strengthen the evidence base for decision making. Environ Health Perspect 123, 1232–1240. doi:10.1289/ehp.1409342 [PubMed: 25956009]

Bopp SK, Kienzler A, Richarz AN et al. (2019). Regulatory assessment and risk management of chemical mixtures: Challenges and ways forward. Crit Rev Toxicol 49, 174–189. doi:10.1080/10408444.2019.1579169 [PubMed: 30931677]

Burnett SD, Blanchette AD, Grimm FA et al. (2019). Population-based toxicity screening in human induced pluripotent stem cell-derived cardiomyocytes. Toxicol Appl Pharmacol 381, 114711. doi:10.1016/j.taap.2019.114711 [PubMed: 31425687]

Chen Z, Liu Y, Wright FA et al. (2020). Rapid hazard characterization of environmental chemicals using a compendium of human cell lines from different organs. ALTEX 37, 623–638. doi:10.14573/altex.2002291 [PubMed: 32521033]

Chen Z, Lloyd D, Zhou YH et al. (2021). Risk characterization of environmental samples using in vitro bioactivity and polycyclic aromatic hydrocarbon concentrations data. Toxicol Sci 179, 108–120. doi:10.1093/toxsci/kfaa166 [PubMed: 33165562]

Chiu WA, Guyton KZ, Martin MT et al. (2018). Use of high-throughput in vitro toxicity screening data in cancer hazard evaluations by IARC monograph working groups. ALTEX 35, 51–64. doi:10.14573/altex.1703231 [PubMed: 28738424]

Clark CR, McKee RH, Freeman JJ et al. (2013). A GHS-consistent approach to health hazard classification of petroleum substances, a class of UVCB substances. Regul Toxicol Pharmacol 67, 409–420. doi:10.1016/j.yrtph.2013.08.020 [PubMed: 24025648]

CONCAWE (1994). The use of the dimethyl sulphoxide (DMSO) extract by the IP 346 method as an indicator of the carcinogenicity of lubricant base oils and distillate aromatic extracts. European Petroleum Refiners Association – Concawe Division. https://www.concawe.eu/wp-content/uploads/2017/01/rpt9451ocr-2005-00417-01-e.pdf (accessed 01.05.2020).

CONCAWE (2019). REACH Roadmap for Petroleum Substances. European Petroleum Refiners Association – Concawe Division. https://www.concawe.eu/wp-content/uploads/REACH-Roadmap-for-Petroleum-Substances-2019-update.pdf (accessed 15.05.2020).

CONCAWE (2020). Hazard Classification and Labelling of Petroleum Substances in the European Economic Area – 2020. 22/20. C. C. a. L. T. F. (STF-23). https://www.concawe.eu/wp-content/uploads/Hazard-classification-and-labelling-of-petroleum-substances-in-the-European-Economic-Area-%E2%80%93-2020.pdf (accessed 25.06.2021).

De Abrew KN, Kainkaryam RM, Shan YK et al. (2016). Grouping 34 chemicals based on mode of action using connectivity mapping. Toxicol Sci 151, 447–461. doi:10.1093/toxsci/kfw058 [PubMed: 27026708]

De Abrew KN, Shan YK, Wang X et al. (2019). Use of connectivity mapping to support read across: A deeper dive using data from 186 chemicals, 19 cell lines and 2 case studies. Toxicology 423, 84–94. doi:10.1016/j.tox.2019.05.008 [PubMed: 31125584]

Drakvik E, Altenburger R, Aoki Y et al. (2020). Statement on advancing the assessment of chemical mixtures and their risks for human health and the environment. Environ Int 134, 105267. doi:10.1016/j.envint.2019.105267 [PubMed: 31704565]

ECHA – European Chemical Agency (2017). Read-Across Assessment Framework (RAAF). Considerations on multi-constituent substances and UVCBs. https://echa.europa.eu/documents/10162/13630/raaf_uvcb_report_en.pdf/3f79684d-07a5-e439-16c3-d2c8da96a316 (accessed 25.08.2020).

ECHA (2020). Testing Proposal Decision on Substance EC 295-332-8 "extracts (petroleum), deasphalted vacuum residue solvent". https://echa.europa.eu/documents/10162/6cda0e05-11af-541b-6dfa-b2101db95a5a (accessed 02.09.2020).

Escher SE, Kamp H, Bennekou SH et al. (2019). Towards grouping concepts based on new approach methodologies in chemical hazard assessment: The read-across approach of the EU-ToxRisk project. Arch Toxicol 93, 3643–3667. doi:10.1007/s00204-019-02591-7 [PubMed: 31781791]

Fang H, Knezevic B, Burnham KL et al. (2016). XGR software for enhanced interpretation of genomic summary data, illustrated by application to immunological traits. Genome Med 8, 129. doi:10.1186/s13073-016-0384-y [PubMed: 27964755]

Ganter B, Tugendreich S, Pearson CI et al. (2005). Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. J Biotechnol. 119, 219–244. doi:10.1016/j.jbiotec.2005.03.022 [PubMed: 16005536]

Goedtke L, Sprenger H, Hofmann U et al. (2020). Polycyclic aromatic hydrocarbons activate the aryl hydrocarbon receptor and the constitutive androstane receptor to regulate xenobiotic metabolism in human liver cells. Int J Mol Sci 22, 372. doi:10.3390/ijms22010372

Grimm FA, Iwata Y, Sirenko O et al. (2015). High-content assay multiplexing for toxicity screening in induced pluripotent stem cell-derived cardiomyocytes and hepatocytes. Assay Drug Dev Technol 13, 529–546. doi:10.1089/adt.2015.659 [PubMed: 26539751]

Grimm FA, Iwata Y, Sirenko O et al. (2016). A chemical-biological similarity-based grouping of complex substances as a prototype approach for evaluating chemical alternatives. Green Chem 18, 4407–4419. doi:10.1039/c6gc01147k [PubMed: 28035192]

Grimm FA, Russell WK, Luo YS et al. (2017). Grouping of petroleum substances as example UVCBs by ion mobility-mass spectrometry to enable chemical composition-based read-across. Environ Sci Technol 51, 7197–7207. doi:10.1021/acs.est.6b06413 [PubMed: 28502166]

Grimm FA, House JS, Wilson MR et al. (2019). Multidimensional in vitro bioactivity profiling for grouping of glycol ethers. Regul Toxicol Pharmacol 101, 91–102. doi:10.1016/j.yrtph.2018.11.011 [PubMed: 30471335]

Hammershoj R, Birch H, Sjoholm KK et al. (2020). Accelerated passive dosing of hydrophobic complex mixtures-controlling the level and composition in aquatic tests. Environ Sci Technol 54, 4974–4983. doi:10.1021/acs.est.9b06062 [PubMed: 32142613]

Harrill J, Shah I, Setzer RW et al. (2019). Considerations for strategic use of high-throughput transcriptomics chemical screening data in regulatory decisions. Curr Opin Toxicol 15, 64–75. doi:10.1016/j.cotox.2019.05.004 [PubMed: 31501805]

Harrill JA, Everett LJ, Haggard DE et al. (2021). High-throughput transcriptomics platform for screening environmental chemicals. Toxicol Sci 181, 68–89. doi:10.1093/toxsci/kfab009 [PubMed: 33538836]

Herrmann K, Pistollato F and Stephens ML (2019). Beyond the 3Rs: Expanding the use of human-relevant replacement methods in biomedical research. ALTEX 36, 343–352. doi:10.14573/altex.1907031 [PubMed: 31329258]

House JS, Grimm FA, Jima DD et al. (2017). A pipeline for high-throughput concentration response modeling of gene expression for toxicogenomics. Front Genet 8, 168. doi:10.3389/fgene.2017.00168 [PubMed: 29163636]

House JS, Grimm FA, Klaren WD et al. (2021). Grouping of UVCB substances with new approach methodologies (NAMs) data. ALTEX 38, 123–137. doi:10.14573/altex.2006262 [PubMed: 33086383]

Hsieh NH, Chen Z, Rusyn I et al. (2021). Risk characterization and probabilistic concentration-response modeling of complex environmental mixtures using new approach methodologies (NAMs) data from organotypic in vitro human stem cell assays. Environ Health Perspect 129, 17004. doi:10.1289/EHP7600 [PubMed: 33395322]

Joseph P (2017). Transcriptomics in toxicology. Food Chem Toxicol 109, 650–662. doi:10.1016/j.fct.2017.07.031 [PubMed: 28720289]

Kavlock RJ, Bahadori T, Barton-Maclaren TS et al. (2018). Accelerating the pace of chemical risk assessment. Chem Res Toxicol 31, 287–290. doi:10.1021/acs.chemrestox.7b00339 [PubMed: 29600706]

Leuraud K and Benichou J (2001). A comparison of several methods to test for the existence of a monotonic dose-response relationship in clinical and epidemiological studies. Stat Med 20, 3335–3351. doi:10.1002/sim.959 [PubMed: 11746322]

Liu Z, Huang R, Roberts R et al. (2019). Toxicogenomics: A 2020 vision. Trends Pharmacol Sci 40, 92–103. doi:10.1016/j.tips.2018.12.001 [PubMed: 30594306]

Love MI, Huber W and Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550. doi:10.1186/s13059-014-0550-8 [PubMed: 25516281]

Low Y, Uehara T, Minowa Y et al. (2011). Predicting drug-induced hepatotoxicity using QSAR and toxicogenomics approaches. Chem Res Toxicol 24, 1251–1262. doi:10.1021/tx200148a [PubMed: 21699217]

McKee RH, Adenuga MD and Carrillo JC (2015). Characterization of the toxicological hazards of hydrocarbon solvents. Crit Rev Toxicol 45, 273–365. doi:10.3109/10408444.2015.1016216 [PubMed: 25868376]

McKee RH, Tibaldi R, Adenuga MD et al. (2018). Assessment of the potential human health risks from exposure to complex substances in accordance with REACH requirements. "White spirit" as a case study. Regul Toxicol Pharmacol 92, 439–457. doi:10.1016/j.yrtph.2017.10.015 [PubMed: 29069582]

Onel M, Beykal B, Ferguson K et al. (2019). Grouping of complex substances using analytical chemistry data: A framework for quantitative evaluation and visualization. PLoS One 14, e0223517. doi:10.1371/journal.pone.0223517 [PubMed: 31600275]

Paul Friedman K, Gagne M, Loo LH et al. (2020). Utility of in vitro bioactivity as a lower bound estimate of in vivo adverse effect levels and in risk-based prioritization. Toxicol Sci 173, 202–225. doi:10.1093/toxsci/kfz201 [PubMed: 31532525]

Phillips JR, Svoboda DL, Tandon A et al. (2019). BMDExpress 2: Enhanced transcriptomic dose-response analysis workflow. Bioinformatics 35, 1780–1782. doi:10.1093/bioinformatics/bty878 [PubMed: 30329029]

Roman-Hubers AT, McDonald TJ, Baker ES et al. (2021). A comparative analysis of analytical techniques for rapid oil spill identification. Environ Toxicol Chem 40, 1034–1049. doi:10.1002/etc.4961 [PubMed: 33315271]

Roy TA, Johnson SW, Blackburn GR et al. (1988). Correlation of mutagenic and dermal carcinogenic activities of mineral oils with polycyclic aromatic compound content. Fundam Appl Toxicol 10, 466–476. doi:10.1016/0272-0590(88)90293-x [PubMed: 3286347]

Salvito D, Fernandez M, Jenner K et al. (2020). Improving the environmental risk assessment of substances of unknown or variable composition, complex reaction products, or biological materials. Environ Toxicol Chem 39, 2097–2108. doi:10.1002/etc.4846 [PubMed: 32780492]

Samet JM, Chiu WA, Cogliano V et al. (2020). The IARC monographs: Updated procedures for modern and transparent evidence synthesis in cancer hazard identification. J Natl Cancer Inst 112, 30–37. doi:10.1093/jnci/djz169 [PubMed: 31498409]

Schultz TW, Amcoff P, Berggren E et al. (2015). A strategy for structuring and reporting a read-across prediction of toxicity. Regul Toxicol Pharmacol 72, 586–601. doi:10.1016/.yrtph.2015.05.016 [PubMed: 26003513]

Tibshirani R, Hastie T, Narasimhan B et al. (2002). Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A 99, 6567–6572. doi:10.1073/pnas.082099299 [PubMed: 12011421]

Trac LN, Sjo Holm KK, Birch H et al. (2021). Passive dosing of petroleum and essential oil UVCBs-whole mixture toxicity testing at controlled exposure. Environ Sci Technol 55, 6150–6159. doi:10.1021/acs.est.1c00343 [PubMed: 33829772]

Uehara T, Ono A, Maruyama T et al. (2010). The Japanese toxicogenomics project: Application of toxicogenomics. Mol Nutr Food Res 54, 218–227. doi:10.1002/mnfr.200900169 [PubMed: 20041446]

Waters M, Stasiewicz S, Merrick BA et al. (2008). CEBS – Chemical effects in biological systems: A public data repository integrating study design and toxicity data with microarray and proteomics data. Nucleic Acids Res 36, D892–D900. doi:10.1093/nar/gkm755 [PubMed: 17962311]

Williams AJ, Grulke CM, Edwards J et al. (2017). The CompTox chemistry dashboard: A community data resource for environmental chemistry. J Cheminform 9, 61. doi:10.1186/s13321-017-0247-6 [PubMed: 29185060]
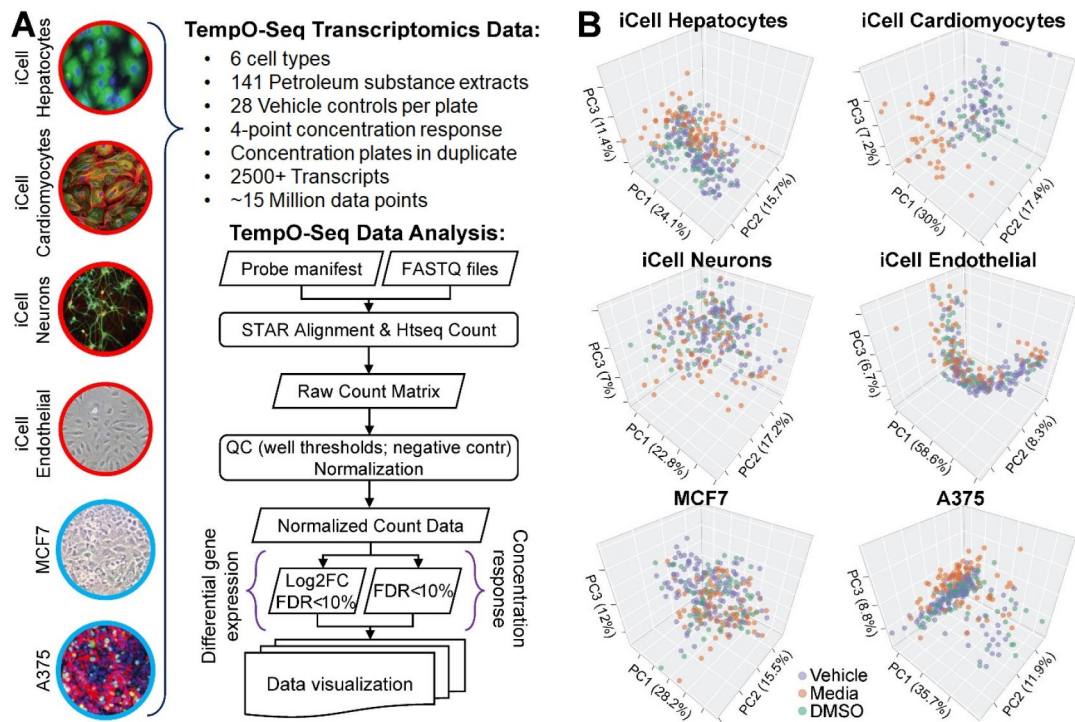
Yauk CL, Harrill AH, Ellinger-Ziegelbauer H et al. (2020). A cross-sector call to improve carcinogenicity risk assessment through use of genomic methodologies. Regul Toxicol Pharmacol 110, 104526. doi:10.1016/j.yrtph.2019.104526 [PubMed: 31726190]

Fig. 1: Overview of the multi-cell *in vitro* transcriptomic analysis of the effects of 141 petroleum substances

(A) Summary of the dataset and data processing pipeline. (B) Principal component analysis of negative control (vehicle (method blank), media, or DMSO, see Section 2 for description) gene expression signatures in each cell type.
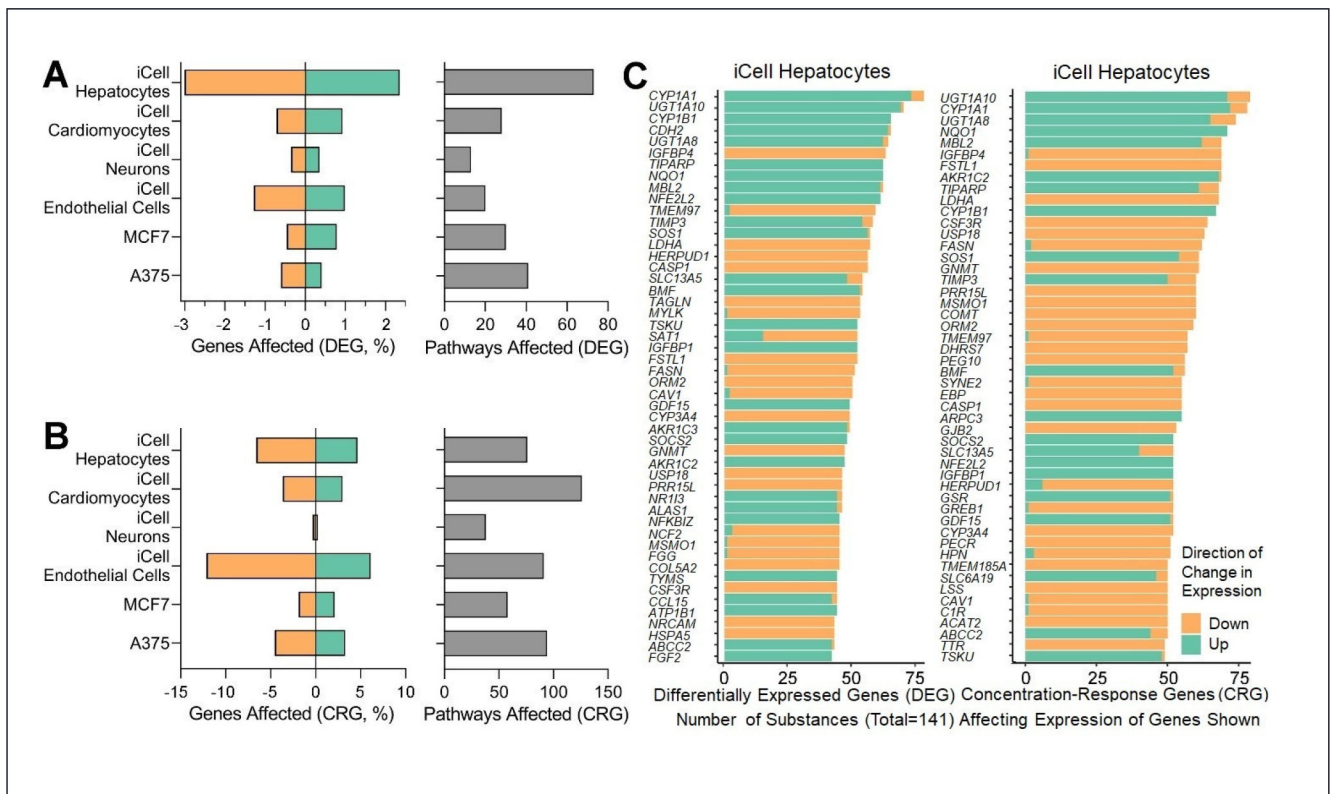
Fig. 2: Cell type-specific transcriptional responses to petroleum substances
(A) Differentially expressed genes (DEG, DeSeq2 analysis and false discovery q 0.01)
were derived by comparing expression between the highest concentration of each substance
($n = 2$) with that of vehicle-treated cells ($n\sim45$) for each expressed gene. Data are shown
as average percent (across the 141 substances) of the total number of genes expressed in
each cell type for up- (green) and down- (orange) regulated genes. Pathways that were
significantly (false discovery q 0.1) affected among these genes were derived using the
Reactome database in the xgr package. (B) Same as (A) but concentration-response genes
($n = 2$ for each concentration and ~45 vehicle-treated) using DeSeq trend analysis, (CRG,
false discovery q 10%) are shown. (C) Example of cell-specific (iCell hepatocytes) effects
of petroleum substances. Top 50 genes that were affected (up- or downregulated), ranked
by the number of substances that had a significant (false discovery q 0.1) effect in either
DEG (left) or CRG (right) analysis. The same data as shown in (C) for each cell type are in
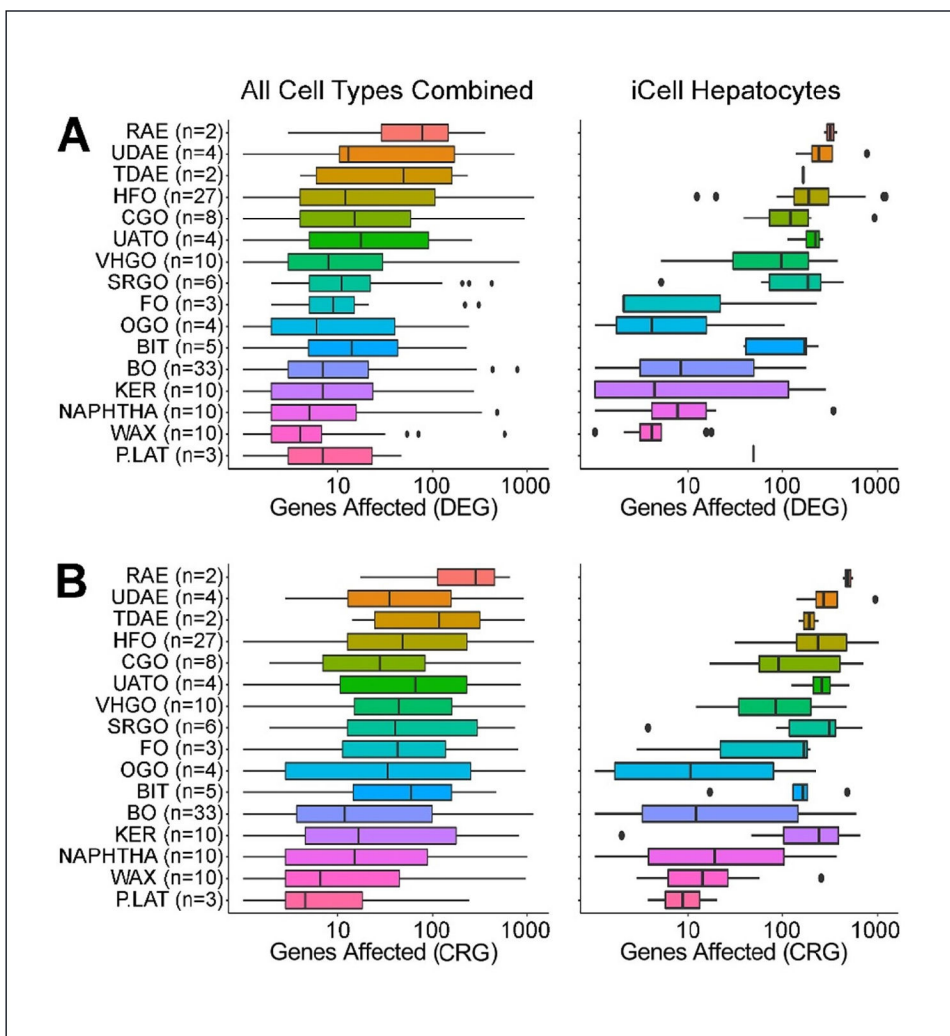Figures S1 and S2[1].

Fig. 3: Class-specific effects of petroleum substances on gene expression in the multi-cell *in vitro* transcriptomic analysis

Box and whiskers plots show the range in the number of genes significantly (false discovery q 0.1, DeSeq2 analysis) affected by the substances in each class (numbers in each class shown as n). (A) Differentially expressed genes (DEG) were derived by comparing expression between the highest concentration of each substance ($n = 2$) with that in vehicle-treated cells (n ~45). (B) Concentration-response genes (CRG) were derived by analyzing the slope in gene expression trend with increasing concentration ($n = 2$ for each concentration and ~45 vehicle-treated). Shown are effects in all cell types (left panels) or in iCell hepatocytes (right panels). Data for each cell type are shown in Figures S3 and S4[1].
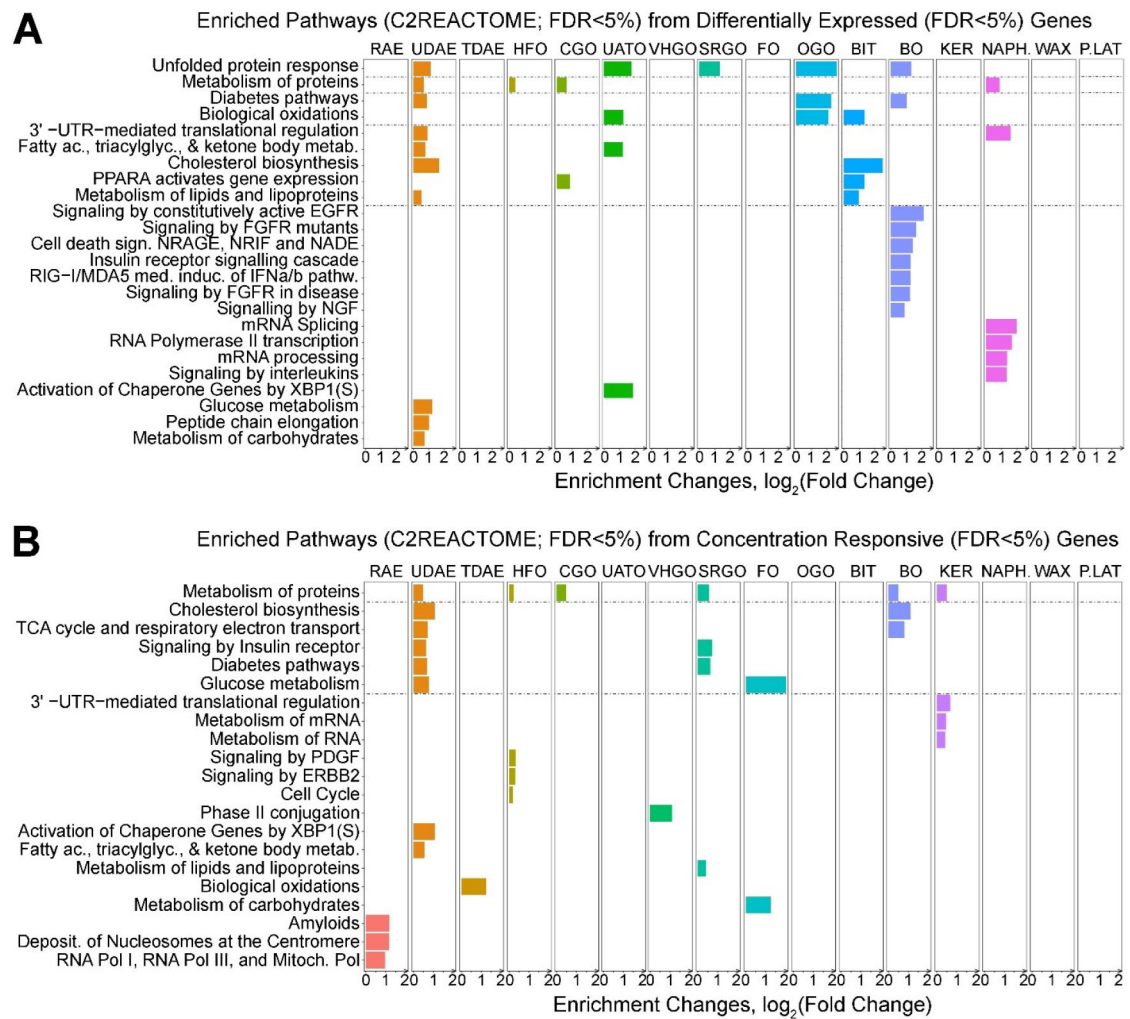
Fig. 4: Class-specific effects of petroleum substances on pathway enrichment (xgr package) in gene expression data from iPSC-derived hepatocytes

A false discovery q threshold of 0.05 was used for the gene set selection. For pathway enrichment, another false discovery q threshold of 0.05 on the pathway selection was used. Bar plots show enriched pathways (C2Reactome) at FDR 5% derived using either differentially expressed genes (A, DEG) or concentration-response genes (B, CRG) affected by the substances in each class. In both cases, the gene-level false discovery of q 0.05 was used. Shown are all substance classes regardless of whether any pathways were enriched. Pathways are ranked by the degree of overlap among classes. The same data for other cell types are shown in Figures S5 and S6[1].
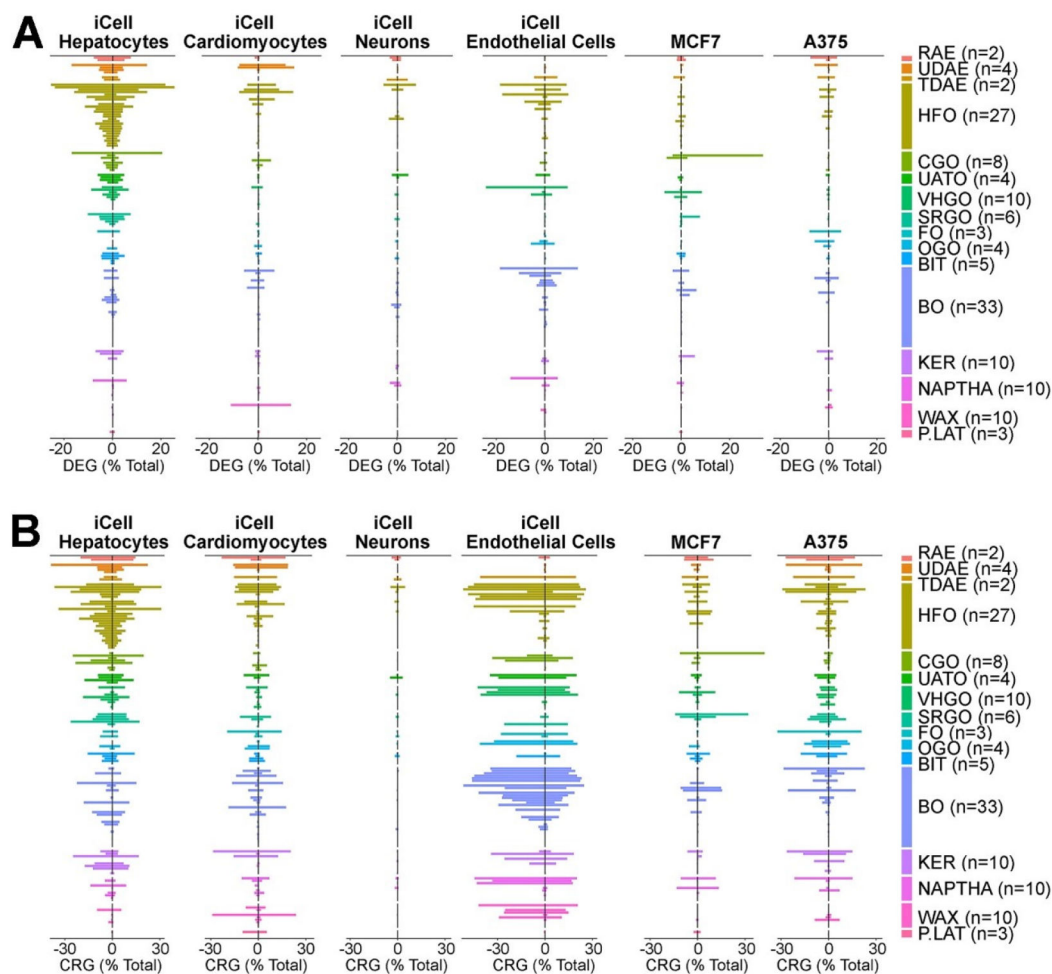
Fig. 5: Heterogeneity of the effects of petroleum substances in the multi-cell *in vitro* transcriptomic analysis – cell-specific effects

Two-sided bar plots show the percent of genes significantly (false discovery $q \leq 0.1$) affected (either up- or downregulated) by the substances in each class (numbers in each class shown as *n*). Within class, substances are ranked (top to bottom) based on their cumulative effect across all 6 cell types. (A) Differentially expressed genes (DEG) were derived by comparing expression between the highest concentration of each substance with that in vehicle-treated cells. (B) Concentration-response genes (CRG) were derived by analyzing the slope in gene expression trend with increasing concentration.
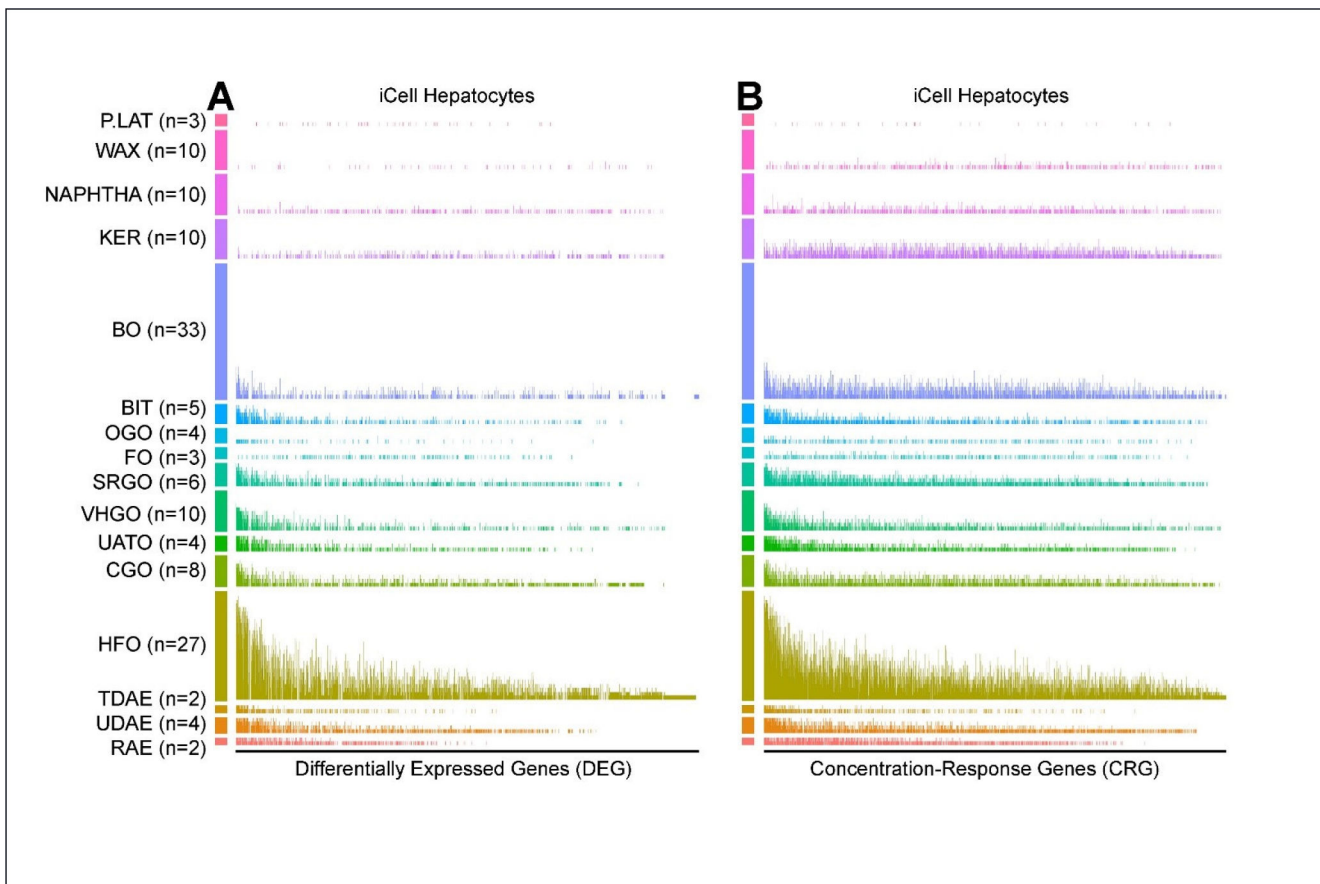
Fig. 6: Heterogeneity of the effects of petroleum substances in the multi-cell *in vitro* transcriptomic analysis

In the hepatocyte example, 2388 genes were assessed for DEGs and CRGs (false discovery $q < 0.1$) after removal of low-count genes. Each vertical line represents one of these genes, ordered left to right as genes most perturbed across all 141 evaluated substances. The height of each bar represents the proportion of times within the UVCB class (numbers in each class shown as *n*) the gene was either differentially expressed (A) or exhibited a concentration response (B).
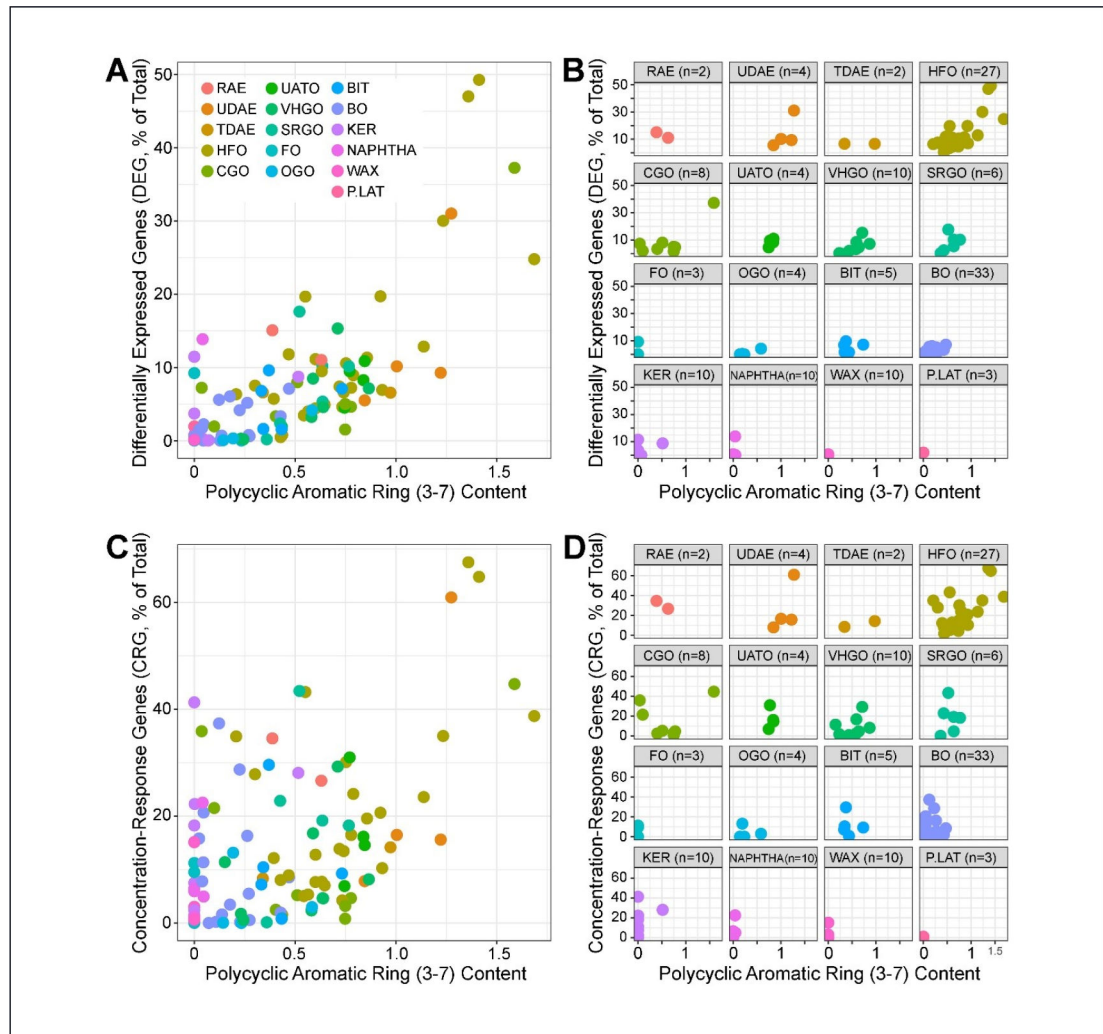
Fig. 7: Correlation between the number of DRGs (A,B) or CRGs (C,D) with the extractable 3–7 polycyclic aromatic ring content (PAH) in iCell hepatocytes with FDR 10%
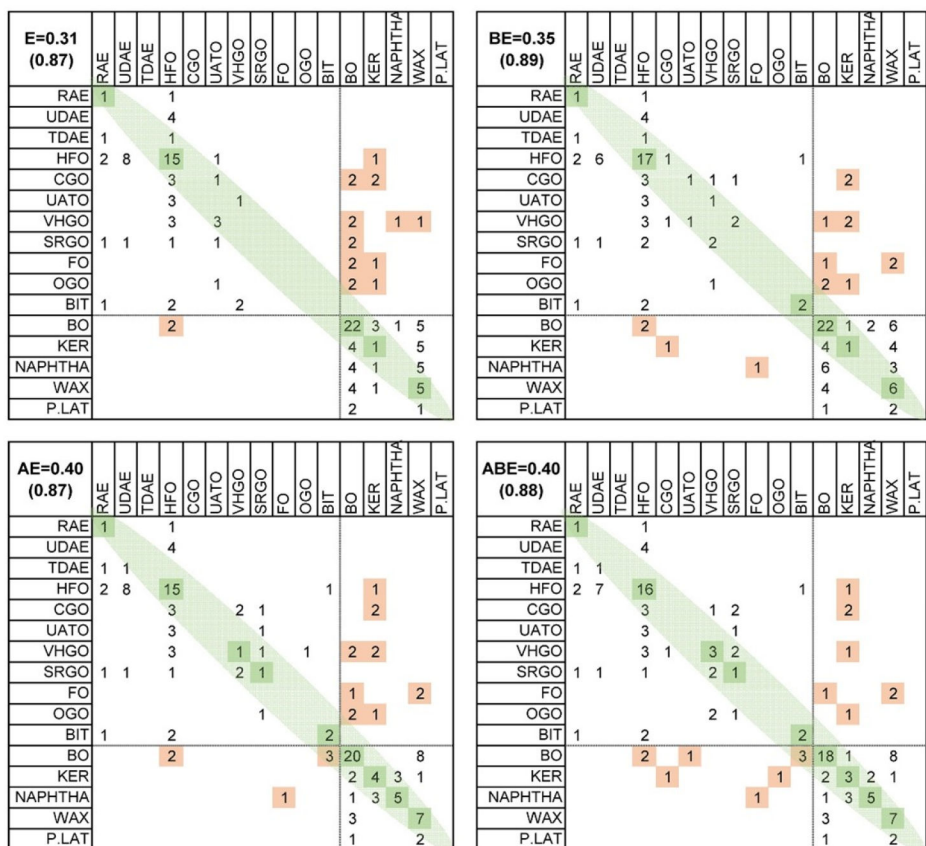Spearman's ρ = 0.77 (A) and 0.57 (C).

Fig. 8: Analysis of the relationship between gene expression data (iCell hepatocytes, "E"), analytical data (PAC, "A"), and summarized bioactivity data (from House et al., 2021, "B")

Top left (E): The results of supervised analysis in which the UVCB category is predicted from the pattern of gene expression data in iCell hepatocytes using the PAM classification procedure as described in Section 2. Rows refer to the true category and columns to predicted category. Correct classification counts are colored in green as values on the diagonal. Categories are ordered according to median bioactivity score, so misclassifications near the diagonal (not colored) are not severe, while misclassifications into categories with substantially different hazard profiles are colored in orange. Top right (BE): correct classifications and misclassifications using both expression and bioactivity patterns. Bottom left (AE): correct classifications and misclassifications using both expression and analytic data. Bottom right (ABE): correct classifications and misclassifications using expression, bioactivity, and analytic data. Numbers in top left corner of each matrix are correct classification rate when only exact matches are considered or (in parenthesis) when misclassifications fall into a hazard category that is not substantially different (P.LAT to BO, or BIT to RAE).

Tab. 1:

Petroleum substance categories and substances used in this study

See Table S1[1] of House et al. (2021) for a complete listing of substance names, CAS and EC numbers, and other information.

Petroleum substance category	Category abbreviation	N of samples in category
Residual aromatic extracts	RAE	2
Untreated distillate aromatic extracts	UDAE	4
Treated distillate aromatic extracts	TDAE	2
Heavy fuel oil components	HFO	27
Cracked gas oils	CGO	8
Unrefined/acid treated oils	UATO	4
Vacuum gas oils, hydrocracked gas oils & distillate fuels	VHGO	10
Straight-run gas oils	SRGO	6
Foots oils	FO	3
Other gas oils	OGO	4
Bitumens/oxidized asphalt	BIT	5
Other lubricant base oils/highly refined base oils	BO	33
Kerosines/MK1 diesel fuel	KER	10
Low boiling point naphthas (gasolines)	NAPHTHA	10
Paraffin and hydrocarbon waxes/slack waxes	WAX	10
Petrolatums	P.LAT	3

Tab. 2:

Cell-specific relationships between gene expression effects (either as differentially expressed genes, DEG; or concentration-response genes, CRG) and polycyclic aromatic compound (PAC, 3–7 ring) content of petroleum UVCBs tested in this study

Both Pearson *r* and Spearman ρ are shown with corresponding *p*-values for each correlation. See Figure 7 for cell type-specific correlation plots.

Cell type	Type	r	p-value	ρ	p-value
iCell hepatocytes	DEG	0.77	1.8E-26	0.76	5.9E-25
	CRG	0.57	2.6E-12	0.55	1.9E-11
iCell cardiomyocytes	DEG	0.57	1.0E-12	0.16	0.071
	CRG	0.21	0.014	0.14	n.s.
iCell neurons	DEG	0.39	1.4E-05	0.07	n.s.
	CRG	0.20	0.047	0.15	n.s.
iCell endothelial	DEG	0.13	n.s.	0.06	n.s.
	CRG	0.09	n.s.	0.08	n.s.
MCF7	DEG	−0.04	n.s.	0.26	0.005
	CRG	−0.04	n.s.	0.25	0.006
A375	DEG	0.11	n.s.	0.08	n.s.
	CRG	0.19	0.026	0.32	2.1E-04
All cells combined	DEG	0.34	5.9E-21	0.24	5.2E-11
	CRG	0.16	8.9E-06	0.23	1.4E-10