Electronic Supplementary Material (ESI) for Green Chemistry.

A chemical-biological similarity-based grouping of complex substances as a prototype approach for evaluating chemical alternatives

- Supplementary Information –

Fabian A. Grimm[1], Yasuhiro Iwata[1], Oksana Sirenko[2], Grace A. Chappell[1], Fred A. Wright[3],
David M. Reif[3], John Braisted[4], David L. Gerhold[4], Joanne M. Yeakley[5],
Peter Shepard[5], Bruce Seligmann[5], Tim Roy[6],
Peter J. Boogaard[7], Hans B. Ketelslegers[8], Arlean M. Rohde[9], Ivan Rusyn[1]

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

[2]Molecular Devices LLC, Sunnyvale, CA

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

[4]National Institutes of Health, National Center for Advancing Translational Sciences, Bethesda, MD

[5]BioSpyder Technologies Inc., Carlsbad, CA

[6]Department of Natural Science, University of South Carolina, Beaufort, SC

[7]SHELL International BV, The Hague, NL

[8]European Petroleum Refiners Association, Concawe Division, Brussels, BE

[9]ExxonMobil Chemical Co., Spring, TX

Address correspondence to:

Ivan Rusyn, MD, PhD
Department of Veterinary Integrative Biosciences
4458 TAMU
Texas A&M University
College Station, TX 77843
(979) 458-9866
irusyn@cvm.tamu.edu

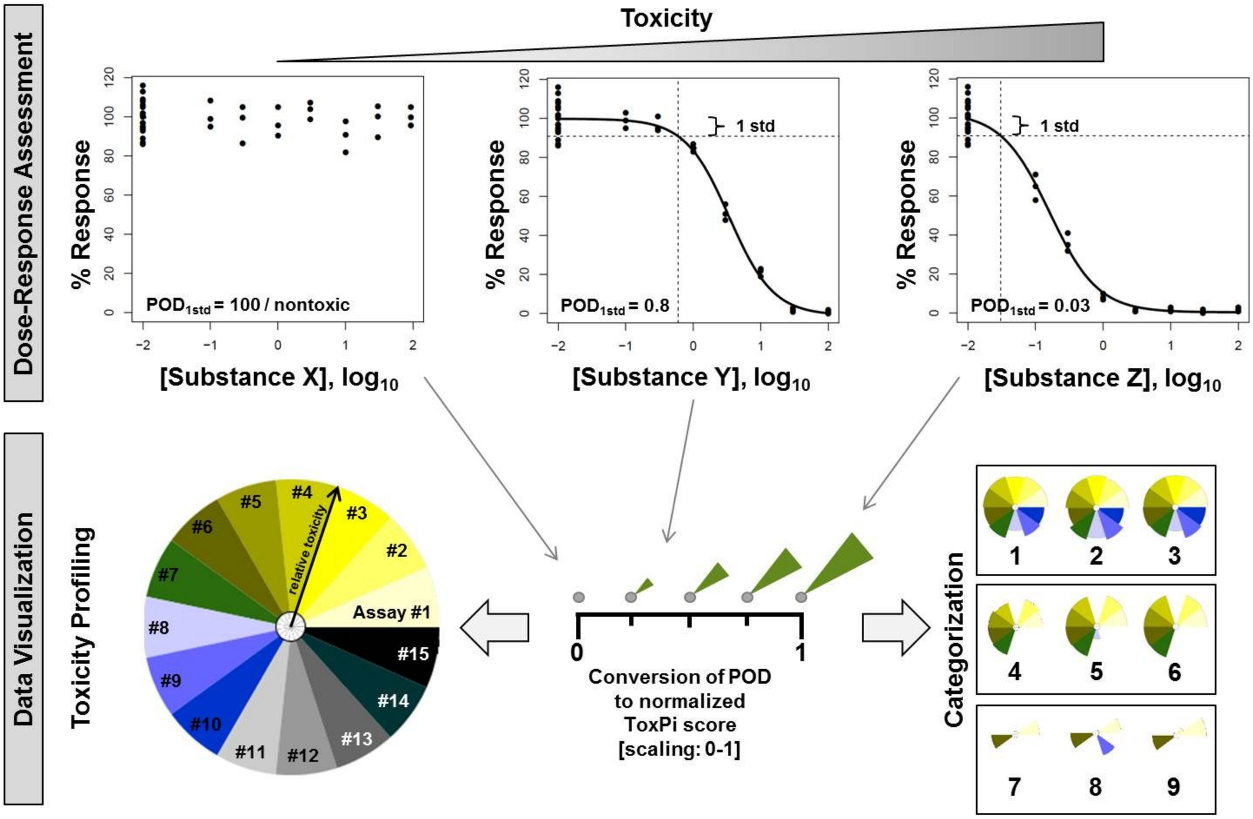

Figure S1: Computational data integration and bioactivity profiling in ToxPi.

Curve-fitting of quantitative outputs from HCS applications (**data acquisition, concentration-response assessment**) in R yielded point-of-departure (POD) benchmark values, which were utilized as relative measures of toxicity. Following normalization of POD values on a 0-1 scale, values derived from sixteen distinct assays or assay parameters were integrated for quantitative bioactivity profiling in ToxPi (**data visualization**). In the resulting ToxPi charts, each slice is representative of an individual assay and the area of the slice is proportional to the relative bioactivity of the chemical within the dataset. ToxPi also provides a cumulative ToxPi score, a relative measure of bioactivity that can be utilized to globally rank chemicals for potential clustering. Ranking chemicals according to their overall ToxPi score can be utilized to reveal clusters of similar compounds or substances.

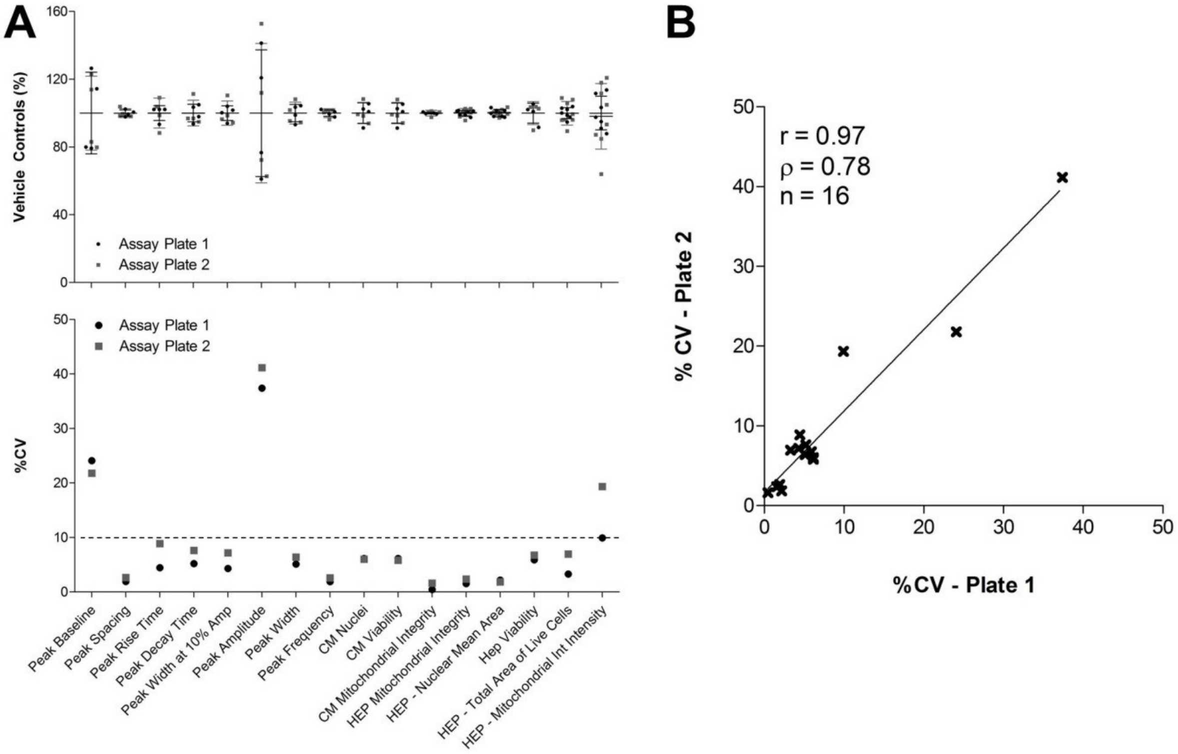

Figure S2: Inter-plate replicability parameters.

Assay performance between assay plates was evaluated using comparisons of normalized controls and coefficients of variation (%CV), an assay quality indicator herein defined as a single standard deviation across the mean of vehicle controls **(A)**. Pearson (r) and Spearman (ρ) correlation analysis was used as an additional quality control to assess the assay reproducibility between plates. [CM=cardiomyocytes; HEP=hepatocytes].

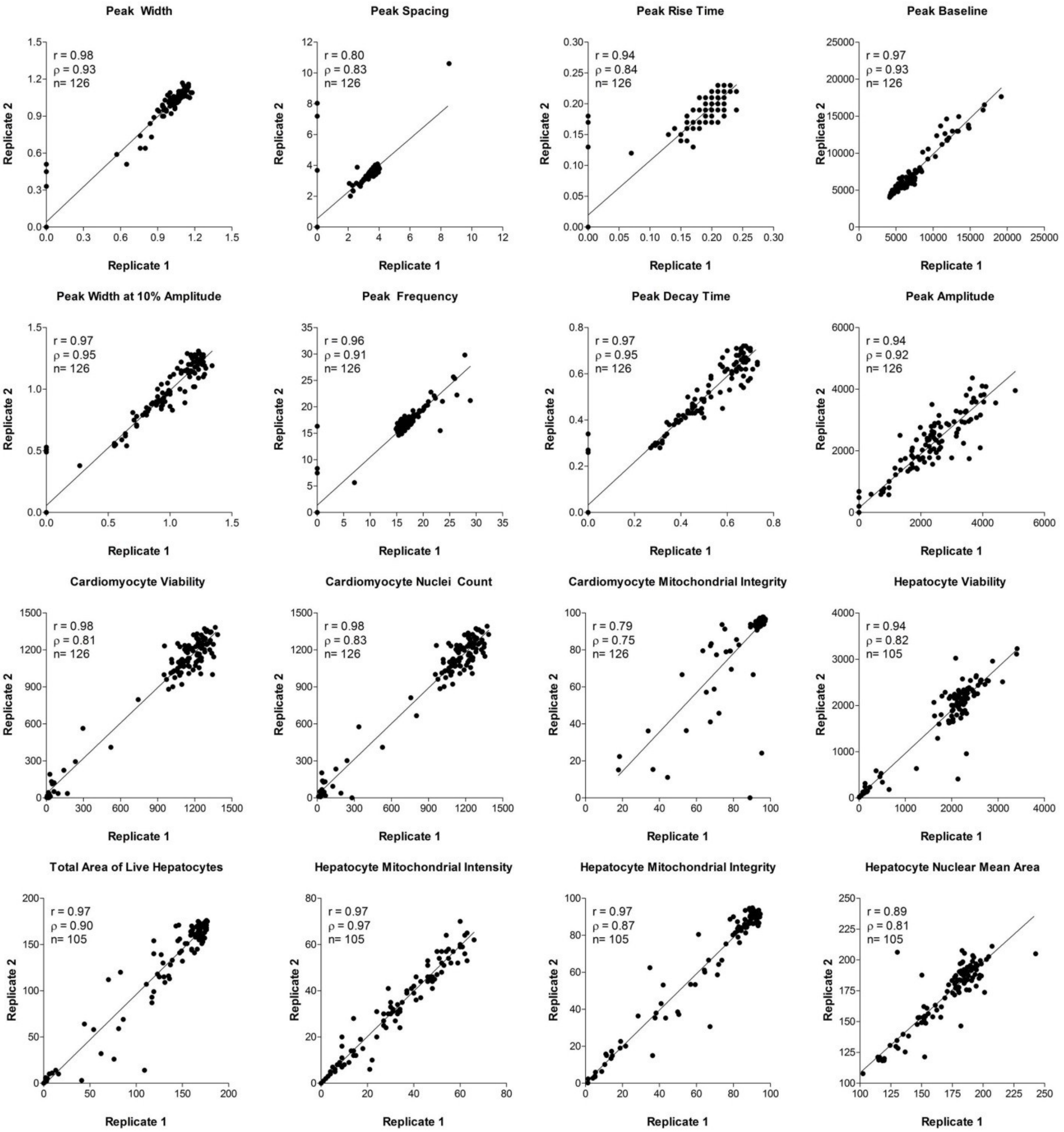

Figure S3: Intra-plate replicability parameters.

The consistency of replicates in high-content screening assays was determined using linear correlation analysis (r=Pearson; ρ=Spearman) for each assay parameter. Plots depict correlations for original, non-normalized duplicate (n=2) measurements in concentration-response (n=6 concentrations for cardiomyocytes, n=5 for hepatocytes) for all 21 Concawe extracts and sixteen assay parameters.

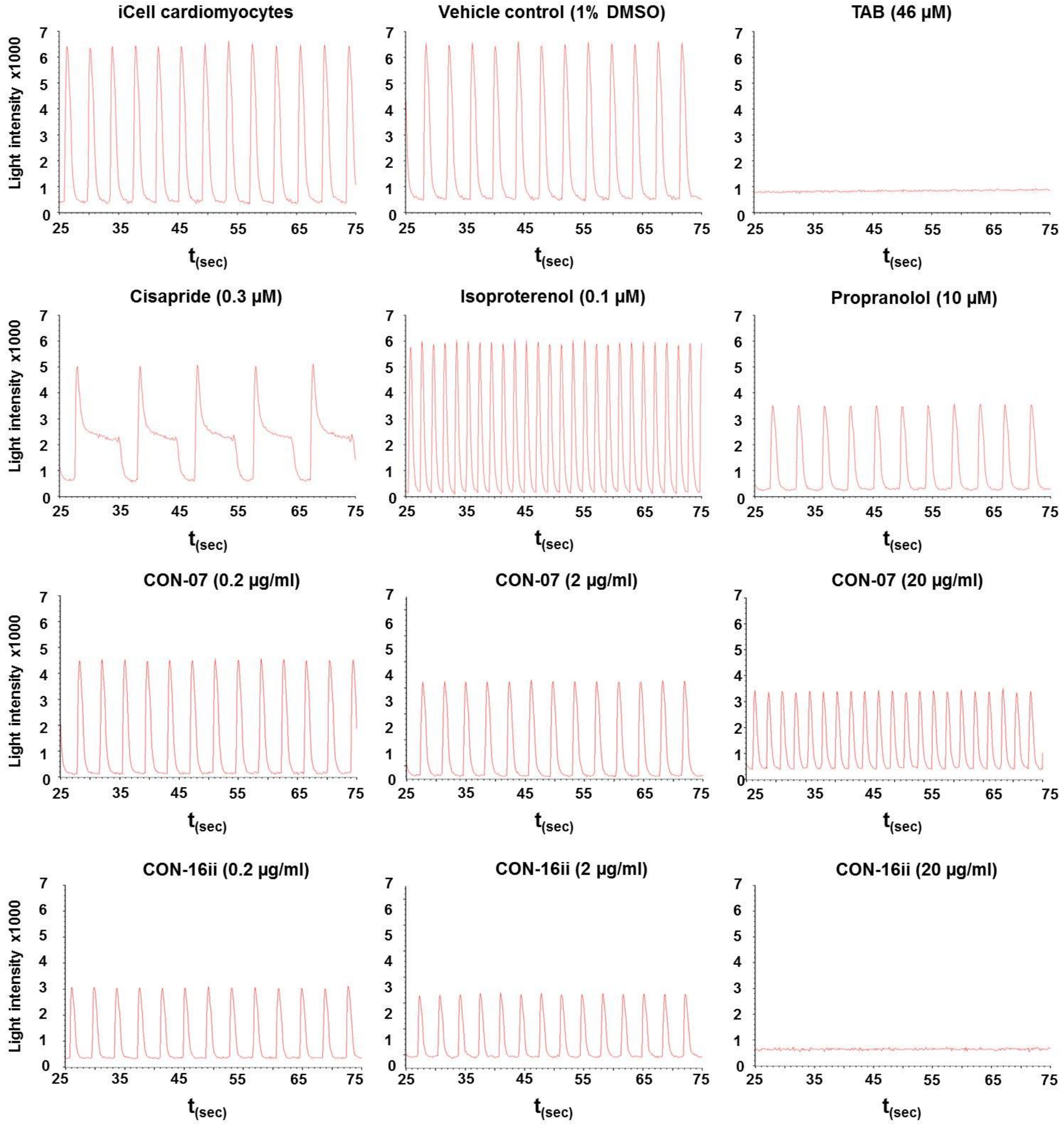

Figure S4: Calcium-flux traces for control chemicals and two petroleum substances. Representative calcium flux-traces for cultured iPSC cardiomyocytes in the presence and absence of vehicle (1% DMSO in media), reference chemicals (tetraoctyl ammonium bromide – cytotoxicity; cisapride – QT interval elongation; isoproterenol – positive chronotrope; propranolol – negative chronotrope) and two petroleum substances, CON-07 (OGO) and CON-16ii (VHGO), in concentration-response are shown.

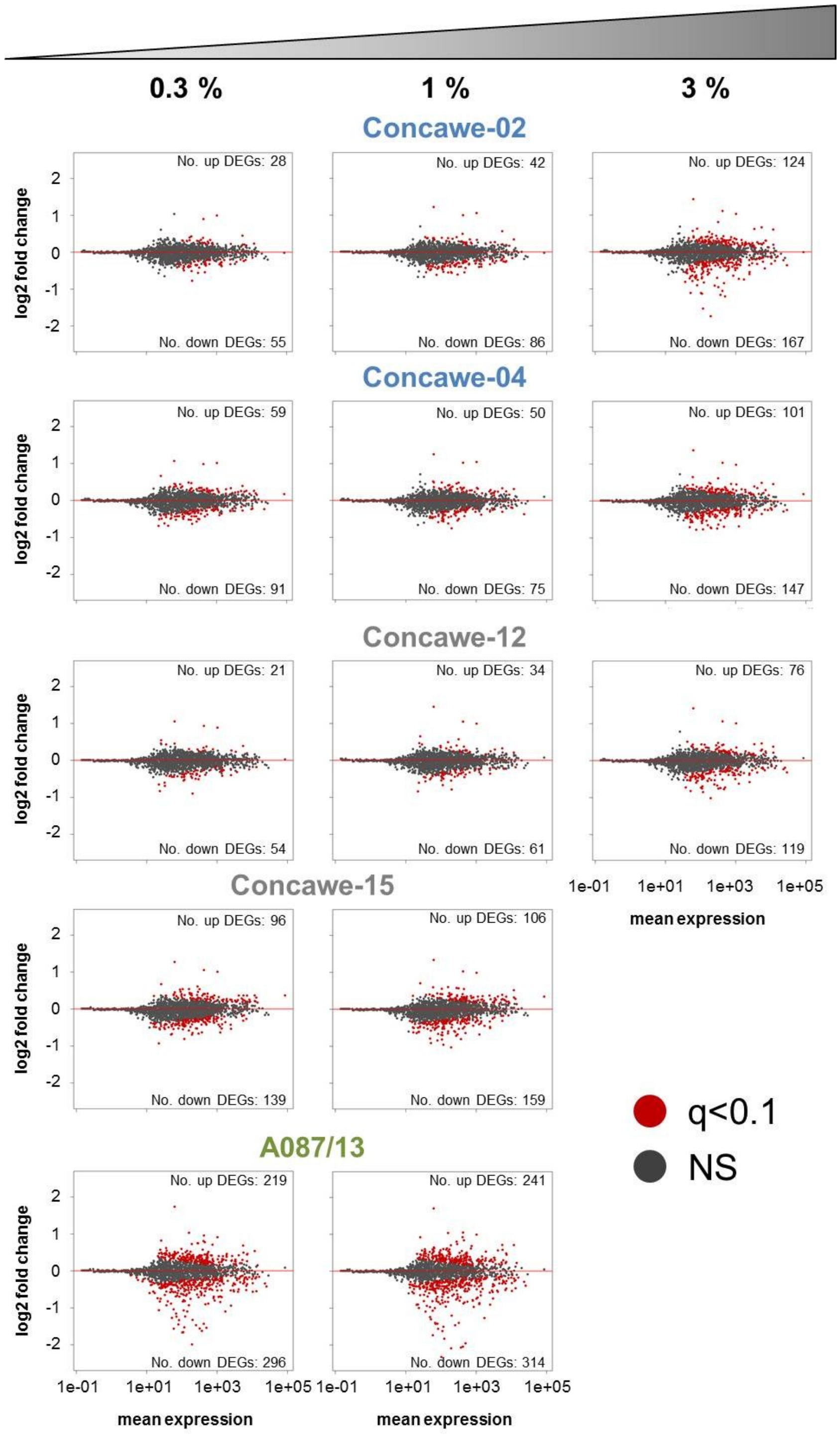

Figure S5: MA plots, which demonstrate the gene expression data transformed onto the M (log ratios) and the A (mean average) scale, show the change in the average level of each transcript in the S1500+ gene set for all chemicals at all concentrations relative to the averages determined for DMSO controls. Differentially expressed genes (*q*<0.1), determined as described in the Materials and Methods section, are indicated by red points.

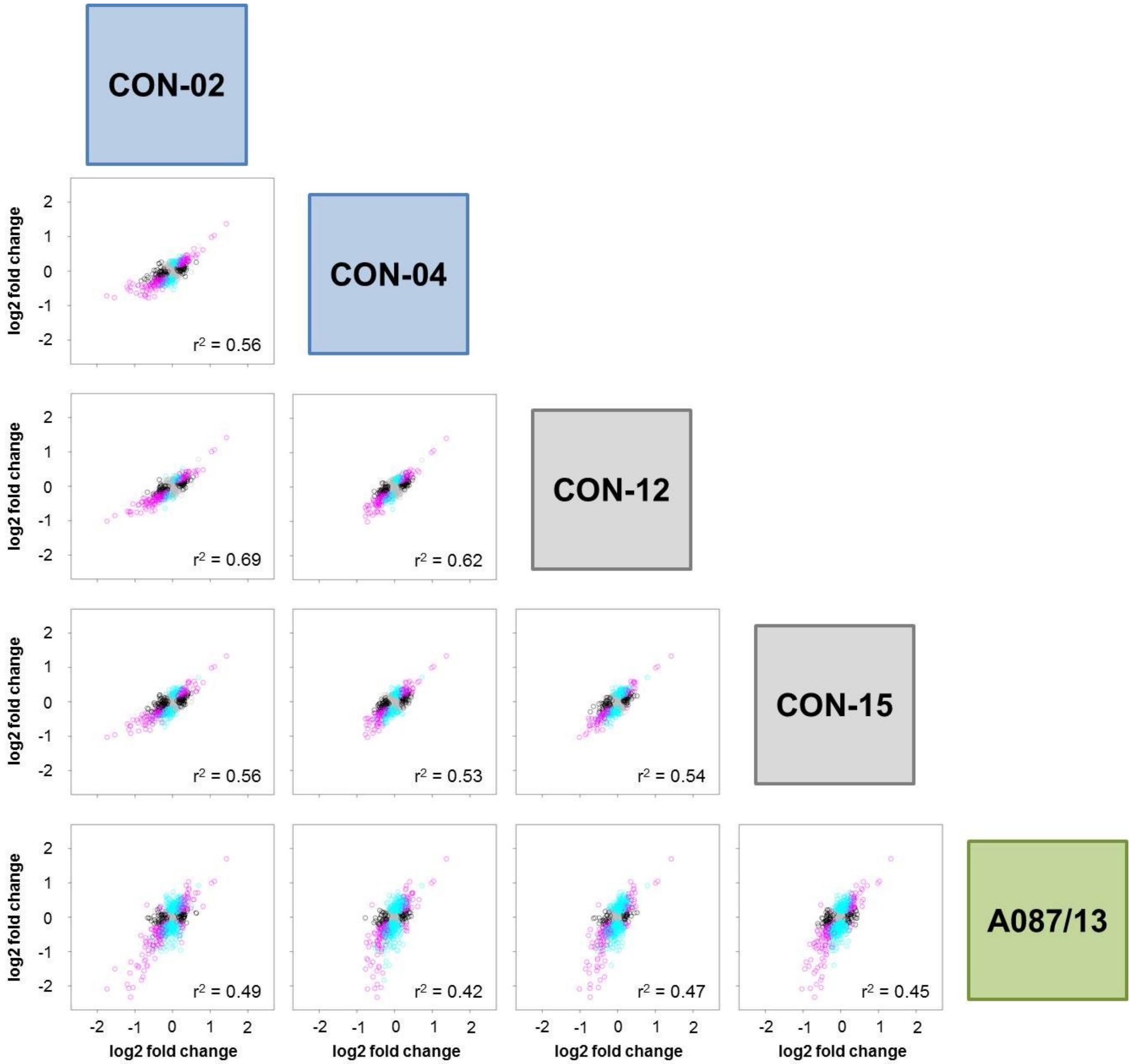

Figure S6: Correlation plots of log2(fold change) for combinations of all chemicals at the highest dose used. Colors indicate significantly differentially expressed genes for either (aqua and black) or both (pink) chemicals in each comparison. r^2=coefficient of determination (squared Pearson correlation).

Table S1: Common differentially expressed genes*

		Con-02		Con-04		Con-12		Con-15		A087/13	
Gene	**Mean count**	**log2FC**	**q-value**	**log2FC**	**q-value**	**log2FC**	**q-value**	**log2FC**	**q-value**	**log2FC**	**q-value**
TIPARP	61.95	1.43	2.55E-18	1.37	5.10E-16	1.42	3.68E-17	1.33	4.15E-15	1.70	1.53E-26
UGT1A1	424.06	1.11	2.22E-12	1.03	5.70E-10	1.06	1.32E-10	1.02	6.48E-10	1.05	2.23E-11
UGT1A6	1027.24	1.03	4.84E-11	0.98	2.84E-09	1.01	4.69E-10	0.99	1.72E-09	0.99	1.54E-10
CD44	317.20	0.81	4.90E-10	0.61	1.87E-05	0.48	1.61E-03	0.57	8.81E-05	0.31	3.34E-02
AKR1C1	5518.76	0.67	4.73E-14	0.49	4.44E-07	0.47	1.23E-06	0.39	8.81E-05	0.71	3.56E-16
SQSTM1	11720.95	0.61	8.20E-15	0.46	9.43E-08	0.38	1.21E-05	0.51	6.48E-10	0.71	1.91E-20
ATF3	60.12	0.57	6.16E-04	0.65	1.49E-04	0.44	2.01E-02	0.60	5.57E-04	0.51	1.56E-03
CDKN1A	279.08	0.44	4.22E-04	0.29	4.12E-02	0.41	2.03E-03	0.57	3.34E-06	0.80	6.95E-13
ASAH1	448.02	0.43	2.05E-05	0.30	7.67E-03	0.43	3.04E-05	0.55	6.58E-08	0.84	2.84E-20
AQP3	159.39	0.42	1.23E-03	0.40	4.43E-03	0.35	1.65E-02	0.60	3.23E-06	1.04	3.04E-20
LOXL1	105.84	0.40	6.97E-03	0.37	2.25E-02	0.44	5.10E-03	0.55	1.70E-04	0.46	1.23E-03
VMP1	740.70	0.40	9.61E-06	0.46	5.59E-07	0.29	4.29E-03	0.30	3.04E-03	0.68	2.47E-16
TSKU	413.15	0.36	8.00E-03	0.31	3.67E-02	0.42	2.89E-03	0.41	3.80E-03	0.92	9.21E-16
CTSA	7129.24	0.31	3.74E-06	0.30	3.05E-05	0.25	8.01E-04	0.34	6.37E-07	0.48	6.48E-15
CCDC92	834.83	0.27	5.23E-04	0.32	9.06E-05	0.24	5.72E-03	0.32	7.91E-05	-0.19	1.97E-02
HLA-A	6206.69	0.24	5.21E-03	0.22	2.12E-02	0.18	8.83E-02	0.26	4.17E-03	-0.28	4.89E-04
NFE2L2	2389.85	0.23	2.44E-02	0.22	5.13E-02	0.21	6.05E-02	0.37	1.33E-04	0.27	4.29E-03
CHCHD1	536.12	0.23	5.67E-02	0.33	3.79E-03	0.23	7.50E-02	0.23	6.99E-02	0.47	1.64E-06
ARPC4	684.54	0.20	1.55E-02	0.19	3.60E-02	0.21	1.99E-02	0.21	2.36E-02	0.29	1.48E-04
RPL27	5539.64	-0.17	9.61E-02	-0.35	7.78E-05	-0.20	5.04E-02	-0.31	5.57E-04	-0.33	5.68E-05
C1R	2956.12	-0.31	7.57E-03	-0.32	7.01E-03	-0.26	3.68E-02	-0.25	5.03E-02	-0.56	1.49E-08
YIF1A	541.26	-0.32	5.77E-04	-0.27	7.67E-03	-0.25	1.35E-02	-0.31	1.43E-03	-0.32	4.16E-04
HAX1	1031.13	-0.32	1.10E-04	-0.39	2.50E-06	-0.25	5.72E-03	-0.24	1.12E-02	-0.49	2.74E-10
ITPR2	112.26	-0.32	3.63E-02	-0.41	8.16E-03	-0.39	1.31E-02	-0.32	5.77E-02	-0.62	3.88E-06
HSPA4	115.12	-0.34	9.25E-03	-0.44	1.18E-03	-0.33	1.83E-02	-0.31	3.78E-02	-0.28	3.17E-02
TRIB3	1353.78	-0.36	3.24E-02	-0.41	1.40E-02	-0.39	2.20E-02	-0.38	3.01E-02	-0.66	3.76E-06
NRP1	116.82	-0.37	1.81E-02	-0.33	4.62E-02	-0.32	6.86E-02	-0.33	5.52E-02	-0.51	3.31E-04
FAIM	495.86	-0.39	1.42E-03	-0.29	2.92E-02	-0.29	3.42E-02	-0.37	4.04E-03	-0.48	3.05E-05
GGH	810.03	-0.40	3.99E-05	-0.23	4.56E-02	-0.31	3.68E-03	-0.21	8.02E-02	-0.80	7.50E-19
GNG5	958.66	-0.43	2.24E-05	-0.36	1.10E-03	-0.39	3.28E-04	-0.35	1.64E-03	-0.28	8.24E-03
KPNA2	1101.99	-0.44	2.40E-09	-0.20	1.96E-02	-0.27	9.65E-04	-0.31	1.28E-04	-0.72	1.42E-23
CKS1B	113.29	-0.44	6.59E-03	-0.31	8.38E-02	-0.42	1.55E-02	-0.39	3.00E-02	-0.53	4.13E-04
ICT1	630.07	-0.45	1.81E-04	-0.24	9.19E-02	-0.28	5.04E-02	-0.26	6.38E-02	-0.35	3.08E-03
PACSIN3	51.61	-0.47	1.19E-02	-0.49	1.16E-02	-0.54	5.72E-03	-0.58	2.19E-03	-0.50	4.51E-03
COL4A2	112.60	-0.47	9.90E-04	-0.46	2.44E-03	-0.70	5.55E-07	-0.49	1.20E-03	-0.86	1.47E-10
COL4A1	200.77	-0.47	2.24E-05	-0.74	4.44E-11	-0.53	2.95E-06	-0.49	3.24E-05	-1.20	4.55E-28
PSIP1	562.54	-0.51	1.13E-06	-0.41	3.64E-04	-0.28	2.13E-02	-0.26	3.86E-02	-0.83	1.61E-16
CENPW	57.55	-0.52	1.42E-03	-0.42	1.96E-02	-0.44	1.35E-02	-0.42	2.26E-02	-0.81	7.01E-08
PHGDH	157.88	-0.53	1.46E-03	-0.34	7.43E-02	-0.54	2.04E-03	-0.65	8.76E-05	-0.83	3.19E-08
TPX2	67.02	-0.54	9.90E-04	-0.35	6.76E-02	-0.43	1.89E-02	-0.55	1.43E-03	-1.00	2.55E-11
EIF4EBP1	2486.05	-0.54	1.03E-05	-0.45	7.99E-04	-0.58	3.29E-06	-0.51	8.46E-05	-0.91	1.12E-15
SMC4	41.74	-0.58	7.19E-04	-0.46	1.32E-02	-0.53	4.39E-03	-0.53	3.77E-03	-0.69	2.01E-05

CDH1	107.46	-0.60	2.76E-05	-0.55	3.45E-04	-0.40	1.47E-02	-0.82	2.13E-08	-0.79	1.29E-08
ORM1	160.07	-0.61	1.64E-04	-0.66	9.06E-05	-0.55	1.63E-03	-0.60	4.14E-04	-0.71	3.33E-06
HPR	110.13	-0.68	2.06E-04	-0.60	2.42E-03	-0.51	1.39E-02	-0.53	9.21E-03	0.36	6.50E-02
CA2	75.10	-0.69	9.51E-06	-0.44	1.13E-02	-0.45	1.03E-02	-0.40	2.96E-02	-0.81	5.14E-08
IGFBP4	297.00	-0.70	1.18E-06	-0.75	5.59E-07	-0.67	6.53E-06	-0.68	7.00E-06	-0.74	1.27E-07
SLC27A2	72.13	-0.71	5.05E-06	-0.40	2.61E-02	-0.49	5.65E-03	-0.32	9.87E-02	-0.60	1.37E-04
MYLK	49.32	-0.72	1.53E-05	-0.71	4.73E-05	-0.54	4.15E-03	-0.79	2.89E-06	-1.24	3.83E-16
FGFR2	108.58	-0.72	4.19E-07	-0.45	4.88E-03	-0.45	5.77E-03	-0.63	4.16E-05	-0.70	7.33E-07
ATF5	1231.10	-0.73	6.44E-06	-0.69	5.79E-05	-0.62	3.72E-04	-0.68	5.49E-05	-0.86	1.53E-08
TK1	83.16	-0.76	2.85E-07	-0.47	4.89E-03	-0.55	8.91E-04	-0.67	2.56E-05	-1.56	3.69E-28
HIST1H2BM	58.87	-0.79	1.00E-06	-0.57	1.24E-03	-0.35	8.62E-02	-0.50	7.29E-03	-1.74	1.61E-31
SHCBP1	36.59	-0.81	2.94E-06	-0.46	2.25E-02	-0.45	2.79E-02	-0.37	8.23E-02	-1.09	1.11E-11
C5	463.25	-0.85	2.10E-08	-0.70	1.87E-05	-0.77	9.00E-07	-0.48	6.86E-03	-0.53	5.92E-04
ASPM	79.40	-0.90	2.47E-09	-0.41	2.25E-02	-0.55	1.61E-03	-0.85	1.24E-07	-1.43	2.24E-22
DEPDC1B	253.48	-0.91	8.20E-15	-0.42	2.52E-03	-0.53	4.05E-05	-0.61	2.30E-06	-1.47	1.19E-34
CCNB1	321.54	-0.95	3.01E-17	-0.64	1.96E-07	-0.76	1.74E-10	-0.53	2.73E-05	-1.40	1.50E-34
CDK1	422.06	-1.06	2.29E-16	-0.36	2.37E-02	-0.51	6.85E-04	-0.52	4.25E-04	-2.05	1.52E-57
CENPF	51.91	-1.11	4.35E-12	-0.58	2.21E-03	-0.82	1.73E-06	-0.92	8.79E-08	-1.30	6.55E-17
AURKB	102.44	-1.11	4.80E-17	-0.47	2.57E-03	-0.73	2.40E-07	-0.69	2.20E-06	-2.33	1.28E-63
KIF2C	109.94	-1.13	1.18E-17	-0.32	5.53E-02	-0.76	8.31E-08	-0.70	1.44E-06	-1.62	2.51E-33
CCNA2	371.01	-1.17	2.55E-18	-0.52	7.99E-04	-0.74	2.63E-07	-0.81	2.13E-08	-2.07	1.01E-55
TOP2A	499.52	-1.20	2.03E-19	-0.46	4.10E-03	-0.72	5.19E-07	-0.64	1.55E-05	-1.96	1.10E-50
CCNB2	117.69	-1.53	1.80E-27	-0.77	5.81E-07	-0.85	1.87E-08	-0.96	3.73E-10	-1.51	1.50E-26
UBE2C	194.33	-1.74	5.86E-44	-0.72	1.55E-07	-1.02	3.34E-15	-1.03	4.15E-15	-2.09	1.44E-58

*Genes that were commonly significantly differentially expressed among the cells treated with one of the five petroleum substances at the highest concentration used are shown. A total of 66 genes were significantly differentially expressed across all chemicals at the highest concentration used.

Table S2: Top-most common, differentially expressed genes*

		Con-02		**Con-04**		**Con-12**		**Con-15**		**A087/13**	
Gene	**Mean count**	**log2FC**	**q-value**	**log2FC**	**q-value**	**log2FC**	**q-value**	**log2FC**	**q-value**	**log2FC**	**q-value**
TIPARP	61.95	1.43	2.55E-18	1.37	5.10E-16	1.42	3.68E-17	1.33	4.15E-15	1.70	1.53E-26
UGT1A1	424.06	1.11	2.22E-12	1.03	5.70E-10	1.06	1.32E-10	1.02	6.48E-10	1.05	2.23E-11
UGT1A6	1027.24	1.03	4.84E-11	0.98	2.84E-09	1.01	4.69E-10	0.99	1.72E-09	0.99	1.54E-10
CD44	317.20	0.81	4.90E-10	0.61	1.87E-05	0.48	1.61E-03	0.57	8.81E-05	0.31	3.34E-02
AKR1C1	5518.76	0.67	4.73E-14	0.49	4.44E-07	0.47	1.23E-06	0.39	8.81E-05	0.71	3.56E-16
SQSTM1	11720.95	0.61	8.20E-15	0.46	9.43E-08	0.38	1.21E-05	0.51	6.48E-10	0.71	1.91E-20
ATF3	60.12	0.57	6.16E-04	0.65	1.49E-04	0.44	2.01E-02	0.60	5.57E-04	0.51	1.56E-03
CDKN1A	279.08	0.44	4.22E-04	0.29	4.12E-02	0.41	2.03E-03	0.57	3.34E-06	0.80	6.95E-13
ASAH1	448.02	0.43	2.05E-05	0.30	7.67E-03	0.43	3.04E-05	0.55	6.58E-08	0.84	2.84E-20
AQP3	159.39	0.42	1.23E-03	0.40	4.43E-03	0.35	1.65E-02	0.60	3.23E-06	1.04	3.04E-20
DEPDC1B	253.48	-0.91	8.20E-15	-0.42	2.52E-03	-0.53	4.05E-05	-0.61	2.30E-06	-1.47	1.19E-34
CCNB1	321.54	-0.95	3.01E-17	-0.64	1.96E-07	-0.76	1.74E-10	-0.53	2.73E-05	-1.40	1.50E-34
CDK1	422.06	-1.06	2.29E-16	-0.36	2.37E-02	-0.51	6.85E-04	-0.52	4.25E-04	-2.05	1.52E-57
CENPF	51.91	-1.11	4.35E-12	-0.58	2.21E-03	-0.82	1.73E-06	-0.92	8.79E-08	-1.30	6.55E-17
AURKB	102.44	-1.11	4.80E-17	-0.47	2.57E-03	-0.73	2.40E-07	-0.69	2.20E-06	-2.33	1.28E-63
KIF2C	109.94	-1.13	1.18E-17	-0.32	5.53E-02	-0.76	8.31E-08	-0.70	1.44E-06	-1.62	2.51E-33
CCNA2	371.01	-1.17	2.55E-18	-0.52	7.99E-04	-0.74	2.63E-07	-0.81	2.13E-08	-2.07	1.01E-55
TOP2A	499.52	-1.20	2.03E-19	-0.46	4.10E-03	-0.72	5.19E-07	-0.64	1.55E-05	-1.96	1.10E-50
CCNB2	117.69	-1.53	1.80E-27	-0.77	5.81E-07	-0.85	1.87E-08	-0.96	3.73E-10	-1.51	1.50E-26
UBE2C	194.33	-1.74	5.86E-44	-0.72	1.55E-07	-1.02	3.34E-15	-1.03	4.15E-15	-2.09	1.44E-58

*Genes that were commonly significantly differentially expressed among the cells treated with one of the five petroleum substances at the highest concentration used are shown. The top-most differentially expressed (based on log2 fold change) are shown.

Table S3: Top-most common, differentially expressed genes across different concentrations*

		Con-02 low		Con-04 low		Con-12 low		Con-15 low		A087/13 low		Con-02 mid		Con-04 mid		Con-12 mid		Con-15 mid		A087/13 mid		Con-02 high		Con-04 high		Con-12 high	
Gene	Mean count	log2FC	q-value	log2FC	q-value	log2FC	q-value	log2FC	q-value	log2FC	q-value	log2FC	q-value	log2FC	q-value	log2FC	q-value	log2FC	q-value	log2FC	q-value	log2FC	q-value	log2FC	q-value	log2FC	q-value
UGT1A6	1027.24	0.99	2.29E-09	1.01	2.55E-09	0.89	2.45E-07	1.01	1.27E-09	0.91	7.59E-09	1.06	1.28E-10	1.05	1.38E-10	1.00	1.51E-09	0.99	1.72E-09	0.99	1.54E-10	1.03	4.84E-11	0.98	2.84E-09	1.01	4.69E-10
UGT1A1	424.06	0.90	2.39E-07	0.98	7.01E-09	0.93	7.19E-08	1.05	4.18E-10	0.96	2.00E-09	1.00	2.58E-09	1.02	6.34E-10	1.05	2.95E-10	1.02	6.48E-10	1.05	2.23E-11	1.11	2.22E-12	1.03	5.70E-10	1.06	1.32E-10
AKR1C1	5518.76	0.45	1.24E-05	0.39	0.000234	0.39	0.000437	0.25	0.032703	0.69	5.34E-15	0.56	2.58E-09	0.33	0.003092	0.33	0.005168	0.39	8.81E-05	0.71	3.56E-16	0.67	4.73E-14	0.49	4.44E-07	0.47	1.23E-06
ASAH1	448.02	0.25	0.058266	0.29	0.026679	0.27	0.061746	0.41	0.000252	0.78	2.65E-17	0.26	0.042478	0.34	0.003505	0.37	0.00197	0.55	6.58E-08	0.84	2.84E-20	0.43	2.05E-05	0.30	0.007667	0.43	3.04E-05
CTSA	7129.24	0.18	0.04167	0.26	0.000914	0.19	0.038094	0.36	1.45E-07	0.44	1.79E-12	0.19	0.027853	0.25	0.001645	0.18	0.047653	0.34	6.37E-07	0.48	6.48E-15	0.31	3.74E-06	0.30	3.05E-05	0.25	0.000801
CCNB1	321.54	-0.42	0.004858	-0.51	0.000217	-0.42	0.004526	-0.28	0.062509	-1.56	6.46E-42	-0.41	0.005406	-0.50	0.00034	-0.36	0.017778	-0.53	2.73E-05	-1.40	1.50E-34	-0.95	3.01E-17	-0.64	1.96E-07	-0.76	1.74E-10
COL4A2	112.60	-0.44	0.00987	-0.45	0.007973	-0.55	0.000782	-0.43	0.00779	-0.79	4.15E-09	-0.48	0.003568	-0.45	0.007059	-0.67	5.91E-06	-0.49	0.001202	-0.86	1.47E-10	-0.47	0.00099	-0.46	0.002443	-0.70	5.55E-07
C5	463.25	-0.54	0.004858	-0.43	0.035134	-0.39	0.083094	-0.36	0.063893	-0.55	0.000483	-0.55	0.003568	-0.42	0.029957	-0.46	0.026178	-0.48	0.006855	-0.53	0.000592	-0.85	2.10E-08	-0.70	1.87E-05	-0.77	9.00E-07
IGFBP4	297.00	-0.56	0.001073	-0.50	0.004336	-0.51	0.004761	-0.66	2.42E-05	-0.72	4.96E-07	-0.61	0.000284	-0.54	0.001565	-0.41	0.035019	-0.68	7.00E-06	-0.74	1.27E-07	-0.70	1.18E-06	-0.75	5.59E-07	-0.67	6.53E-06
COL4A1	200.77	-0.78	1.53E-12	-0.63	5.90E-08	-0.90	3.07E-16	-0.65	5.46E-08	-1.33	1.90E-33	-0.57	7.30E-07	-0.76	7.09E-12	-0.79	6.91E-13	-0.49	3.24E-05	-1.20	4.55E-28	-0.47	2.24E-05	-0.74	4.44E-11	-0.53	2.95E-06

*Genes that were commonly significantly differentially expressed among the cells treated with one of the five petroleum substances, at all concentrations, are shown. A total of 10 genes were significantly differentially expressed across all chemicals and concentrations.