Development and Use of High Content Tier 1 Screening Assays at the USEPA National Center for Computational Toxicology (NCCT)

Joshua A. Harrill, USEPA National Center for Computational Toxicology (NCCT)

Office of Research and Development
National Center for Computational Toxicology, RTP, NC

USEPA Endocrine Disruptors Screening Program (EDSP) Webinar
May 29, 2018

Disclaimer

The views expressed in this presentation are those of the author(s) and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency, nor does mention of trade names or products represent endorsement for use.

Outline

- **Background**
  - Computational Toxicology Strategic Vision
  - Tier 1 Screening Assays
- **High Throughput Transcriptomics (HTTr)**
  - TempO-Seq Technology Overview
  - Screening Format
  - Technical Reproducibility
  - Applications
- **High Throughput Phenotypic Profiling (HTPP)**
  - Cell Painting Assay Description
  - Laboratory & Analysis Workflows
  - Concentration-Response Modeling
  - Applications

EPA United States Environmental Protection Agency

NCCT HTTr Project Team

National Center for Computational Toxicology

Joshua **Harrill**
Toxicologist

Clinton **Willis**
NSSC (JH)

Imran **Shah**
Computational Systems Biologist

R. Woodrow **Setzer**
Mathematical Statistician

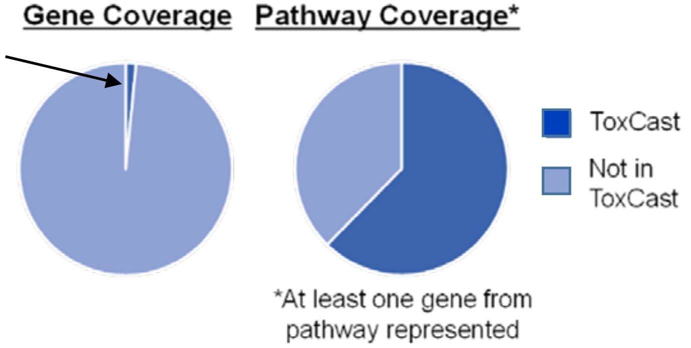
Derik **Haggard**
ORISE Fellow

Richard **Judson**
Bioinformatician

Russell **Thomas**
Director

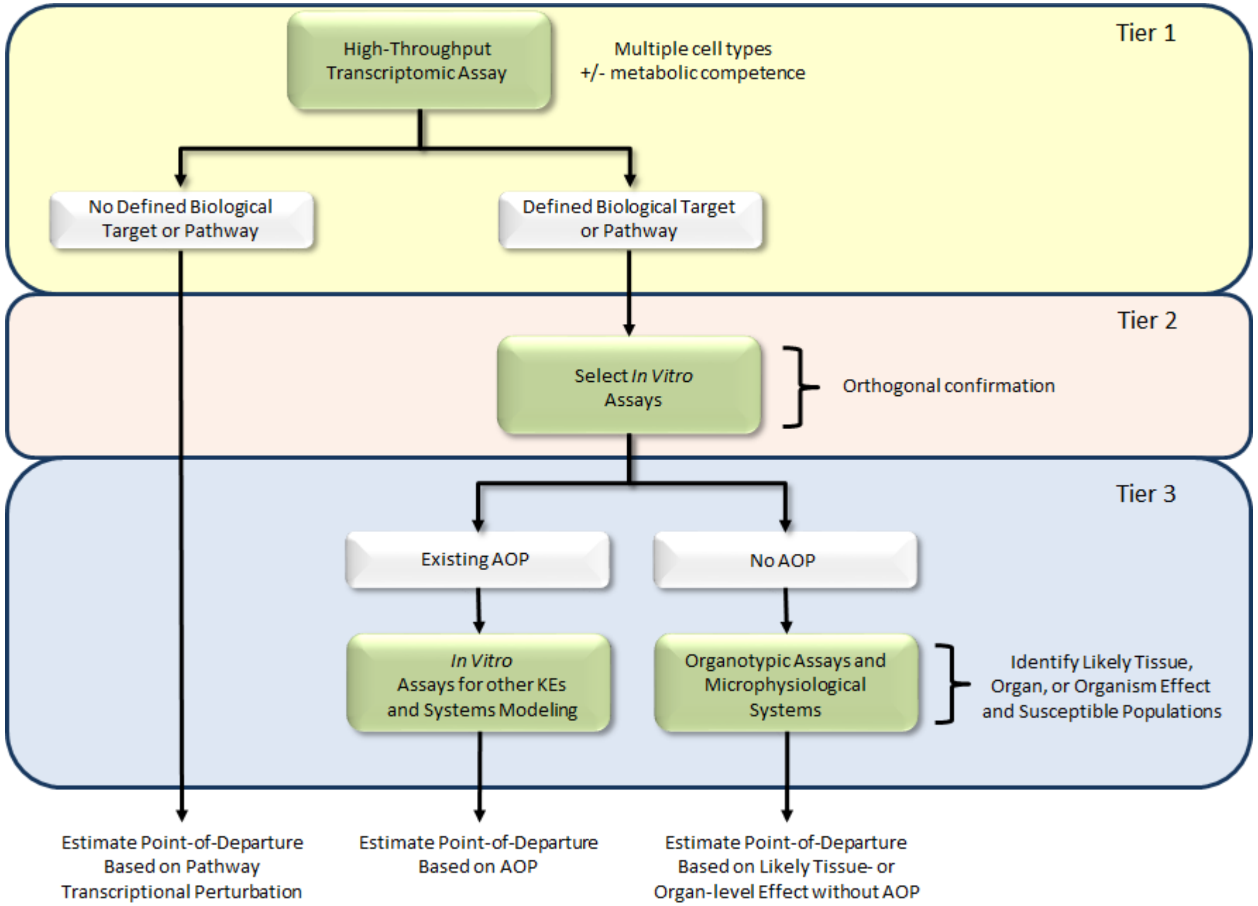
Background

- ToxCast assays cover many genes and pathways, but do not provide complete coverage of biological space.



- **USEPA Strategic Vision and Operational Roadmap:**
  - Tier 1 strategy must cast the broadest net possible for capturing hazards associated with chemical exposure.
  - Global gene expression provides a robust and comprehensive evaluation of chemically induced changes in biological processes.
  - Increasing efficiency and declining cost of generating whole transcriptome profiles has made high-throughput transcriptomics (HTTr) a practical option for determining bioactivity thresholds in *in vitro* models.

A strategic vision and operational road map for computational toxicology at the U.S. Environmental Protection Agency

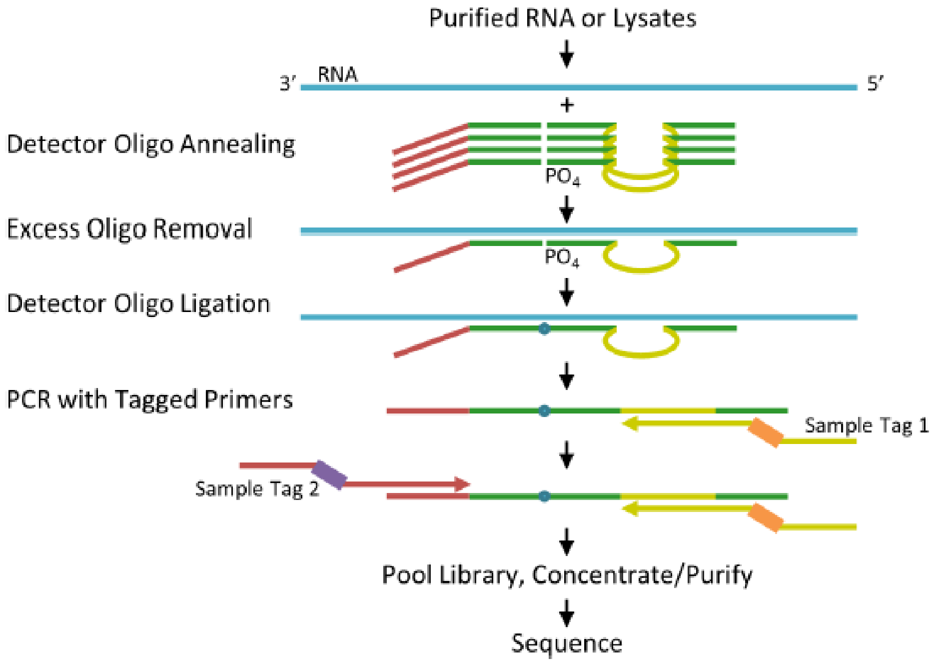


Background

Technology

- The **TempO-Seq** human whole transcriptome assay measures the expression of greater than 20,000 transcripts.
- Requires only picogram amounts of total RNA per sample.
- Compatible with purified RNA samples or **cell lysates**.
- Transcripts in cell lysates generated in 384-well format are barcoded according to well position and combined in a single library for sequencing using industry standard instrumentation.
- Scalable, targeted assay:
  - 1) specifically measures transcripts of interest
  - 2) ~50-bp reads for all genes
  - 3) requires less flow cell capacity than RNA-Seq
- Per sample fastq files are generated and aligned to BioSpyder sequence manifest to generate integer count tables.

TempO-Seq Assay Illustration



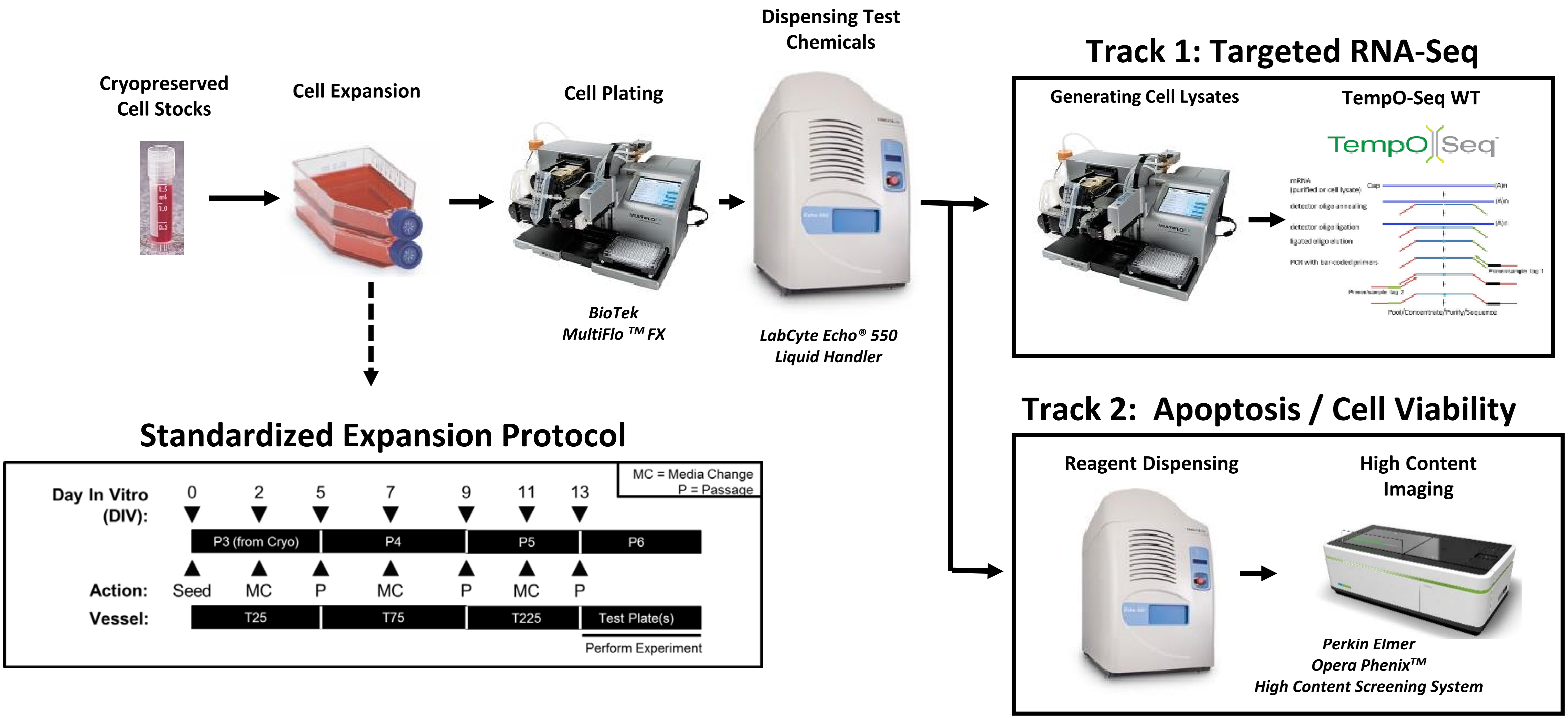
HTTr MCF Screen: Experimental Design

Parameter	Multiplier	Notes
Cell Type(s)	1	MCF7
Culture Condition	1	DMEM + 10% HI-FBS [a]
Chemicals	2,112	ToxCast ph1, ph2 Nominated chemicals from e1k / ph3
Time Points:	1	6 hours
Assay Formats:	2	TempO-Seq HCI Cell Viability & Apoptosis
Concentrations:	8	3.5 $\log_{10}$ units; semi $\log_{10}$ spacing
Biological Replicates:	3	--

- **Total number of samples:** 54,432
- **Total number of endpoint readouts:** 1.15×10^9
- **Total size of fastq files:** 32.5 to 54.4 TB

[a] MCF7 cells cultured in DMEM + 10% HI-FBS was selected as the test system to facilitate comparability to the Broad Institute Connectivity Map (CMAP) database (http://portals.broadinstitute.org/cmap/).

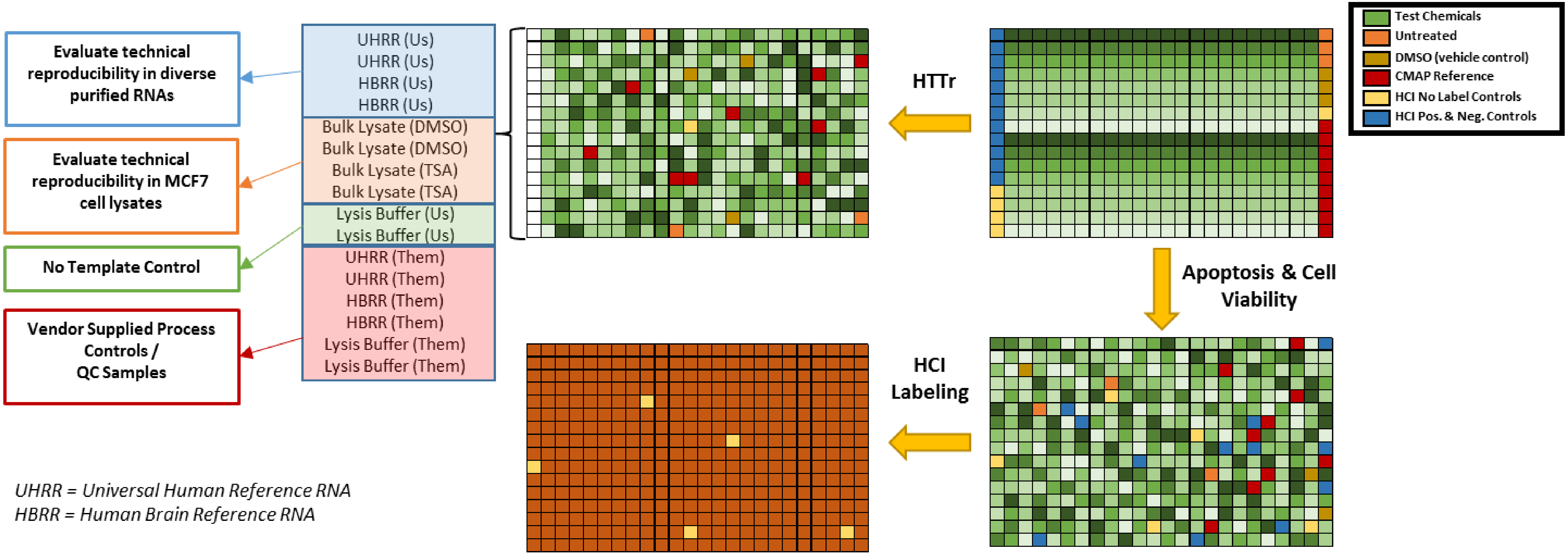
Experimental Workflow



Treatment Randomization & Quality Control Samples

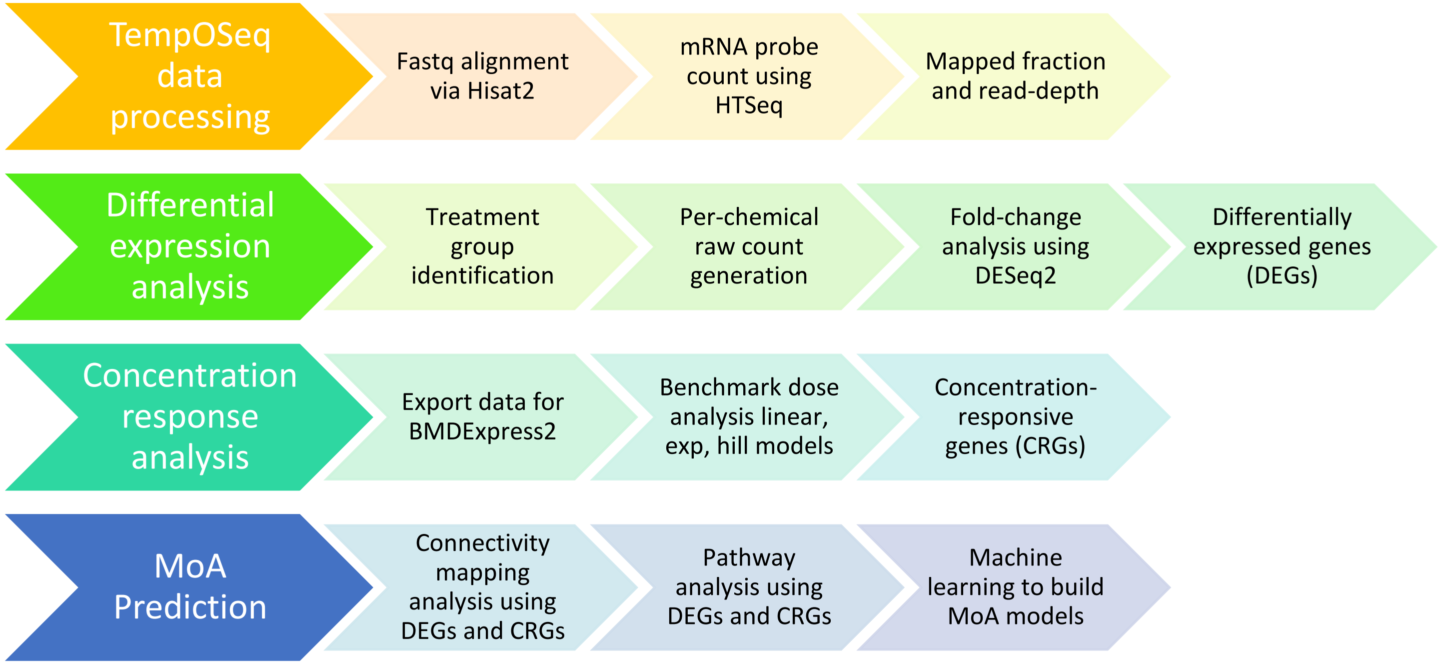
Treatment Randomization: *Each test plate uniquely randomized with respect to treatment.*

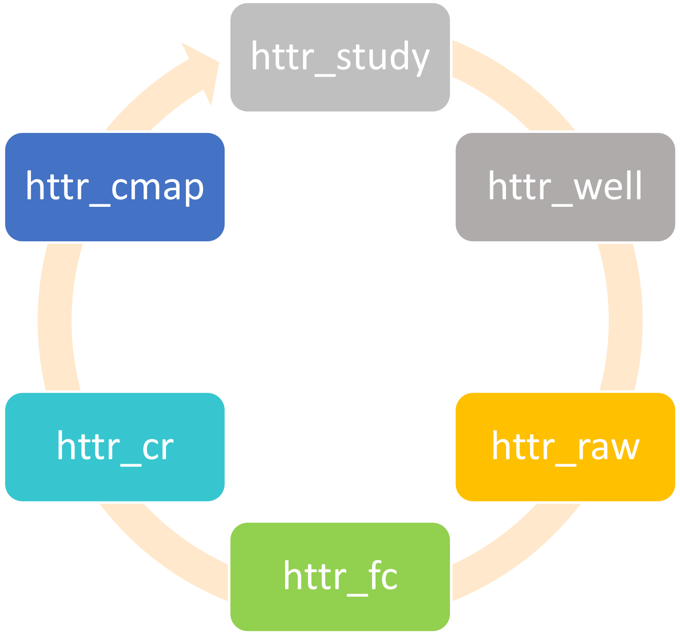
QC Samples: *Quality Control samples included on each plate*



UHRR = Universal Human Reference RNA
HBRR = Human Brain Reference RNA

HTTr Analysis Pipeline & Infrastructure





Python & R analysis pipeline

http://bitbucket.zn.epa.gov/projects/HTTR/repos/httr-wf-dev

Imran Shan
Josh Harrill
Woodrow Setzer
Richard Judson
Derik Haggard

MongoDB

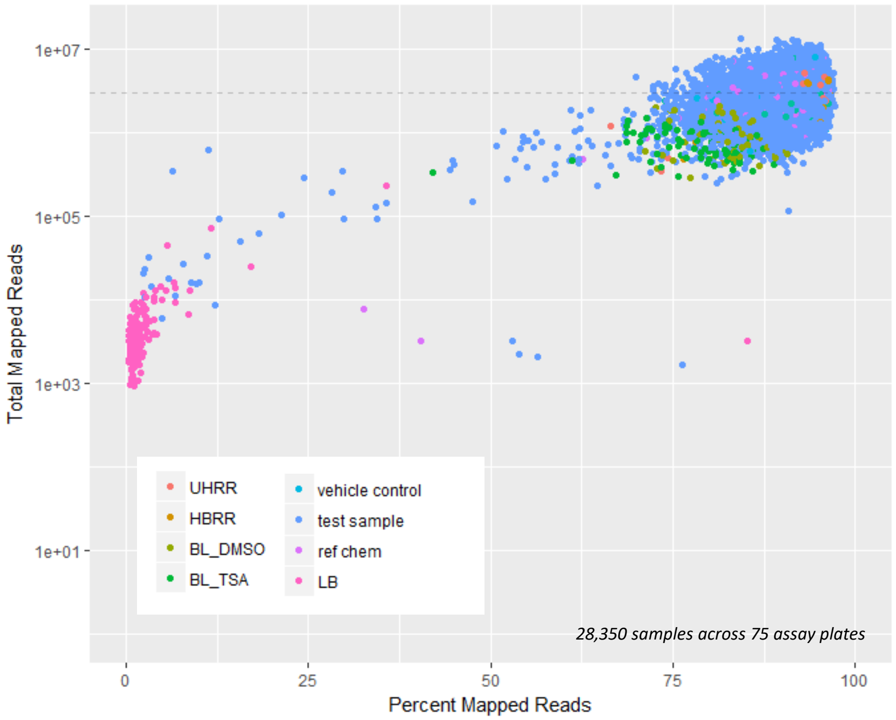
mongodb://pb.epa.gov/httr_v1
mongodb://pb.epa.gov/cmap_v2

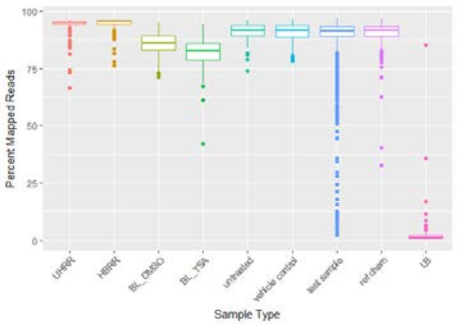
NCCT Investigators
NCCT IT

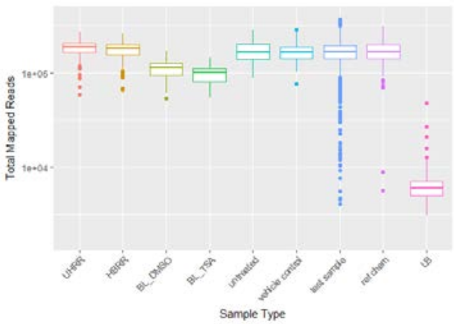
Technical Reproducibility

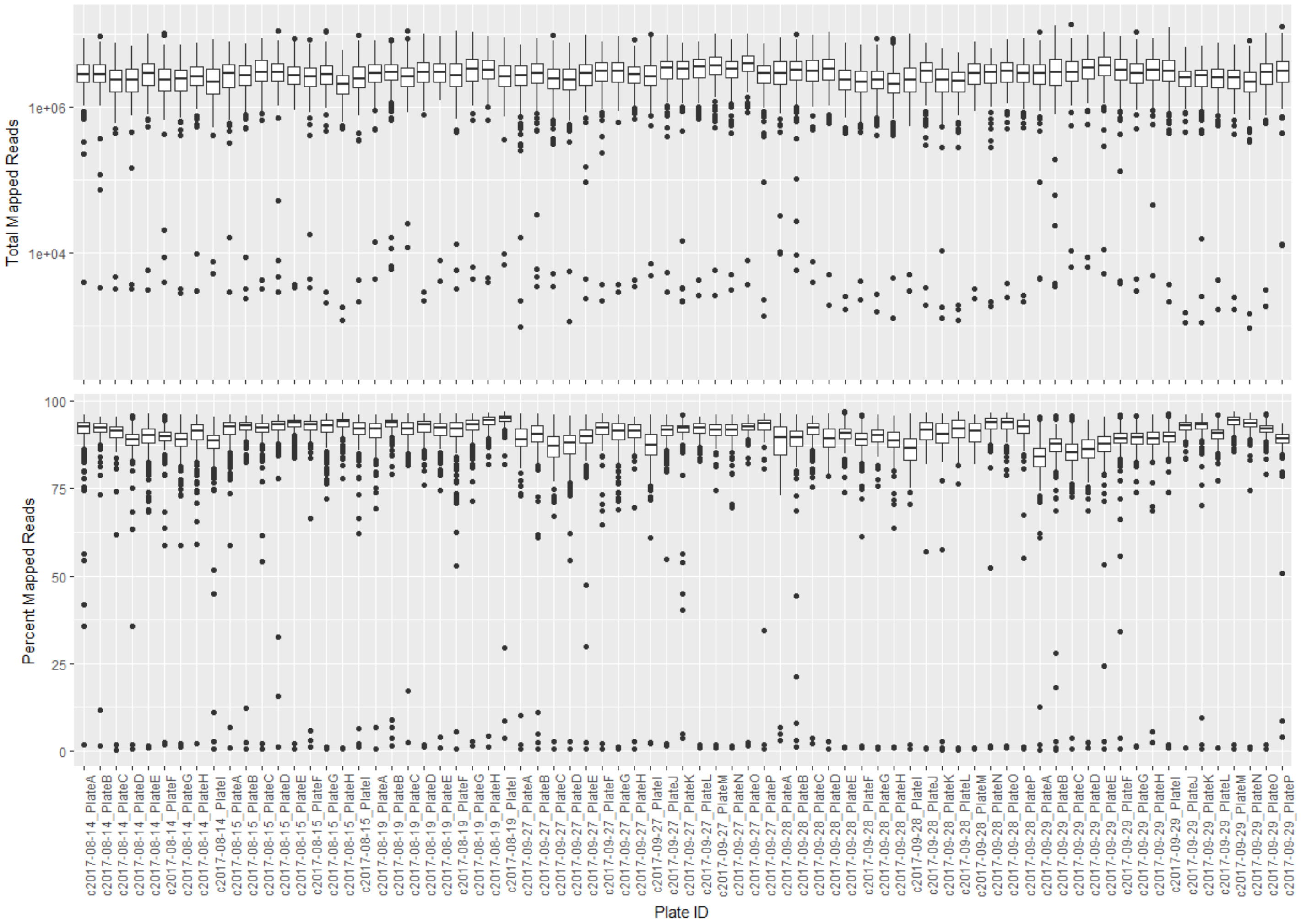
United States Environmental Protection Agency

Reproducibility of Read Depths and Mapping Rates



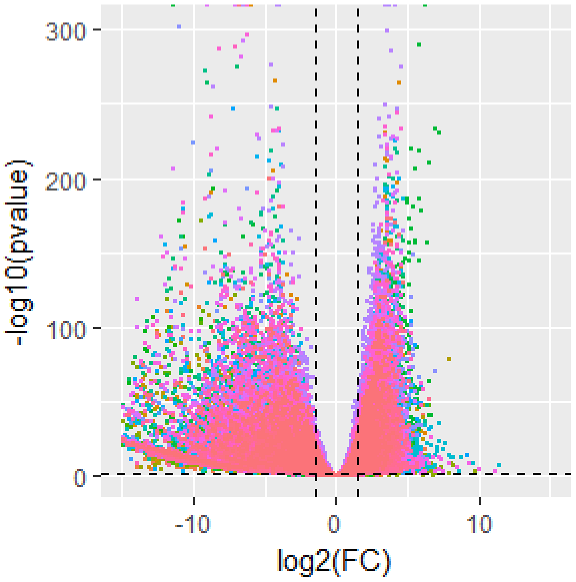




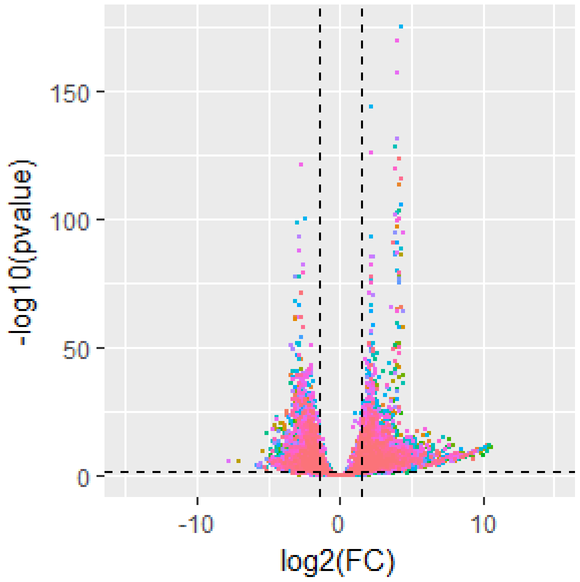


Reproducibility of Log_2(FC) Estimates

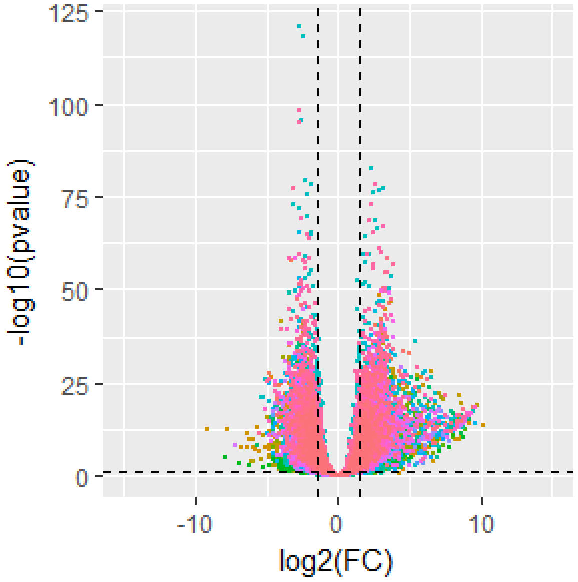
UHRR v HBRR

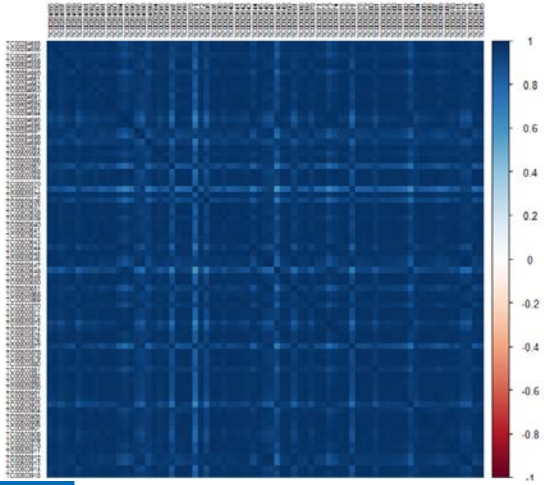


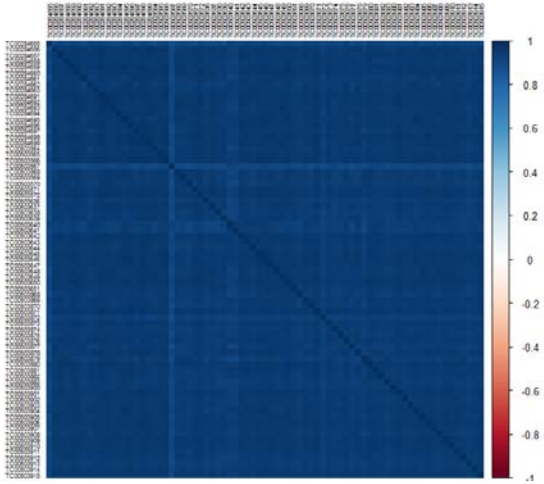
DMSO vs TSA (Bulk Lysate)

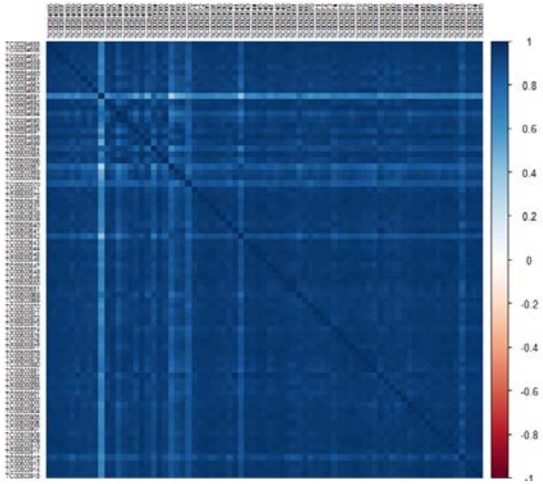


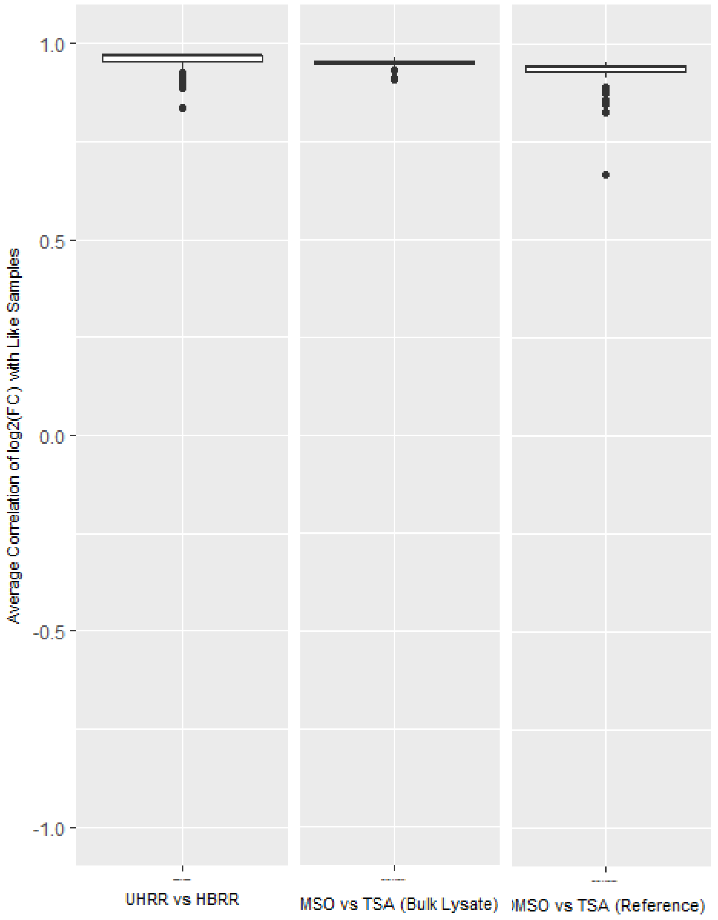
DMSO vs TSA (Treatment)











Concentration-Response Modeling / Bioactivity Thresholds

Benchmark Dose Modeling

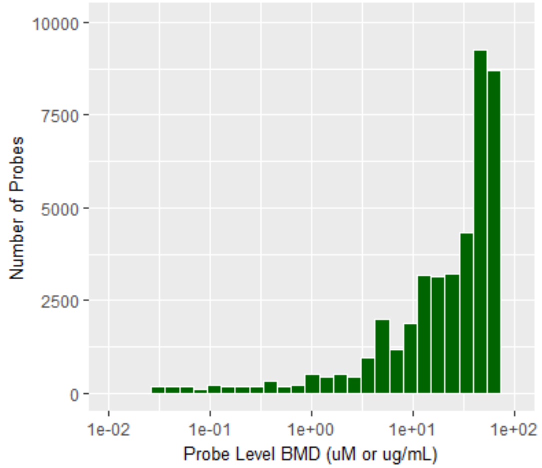
Parameter	Criteria [a]
Pre-filter:	ANOVA ($p_{raw} < 0.05$ & $\lvert FC \rvert \geq 2$)
Models	Hill, Exponential 2, *poly2, power, linear*
BMR Factor:	1.349 (10 %)
Best Model Selection:	Lowest AIC
Hill Model Flagging [b]:	'k' < 1/3 Lowest Positive Dose Retain Flagged Models
Pathway Analysis:	Genes with BMD <= Highest Dose ≥ 3 ≥ 1% Gene Set Coverage
Gene Set Collections [c]:	MSigDB_C2 MSigDB_H Reactome BioPlanet KEGG

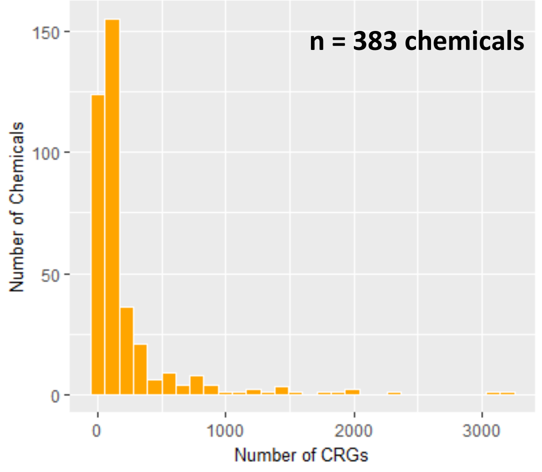
[a] *Exploratory analysis – modeling criteria not finalized*

[c] Gene Set Collections:

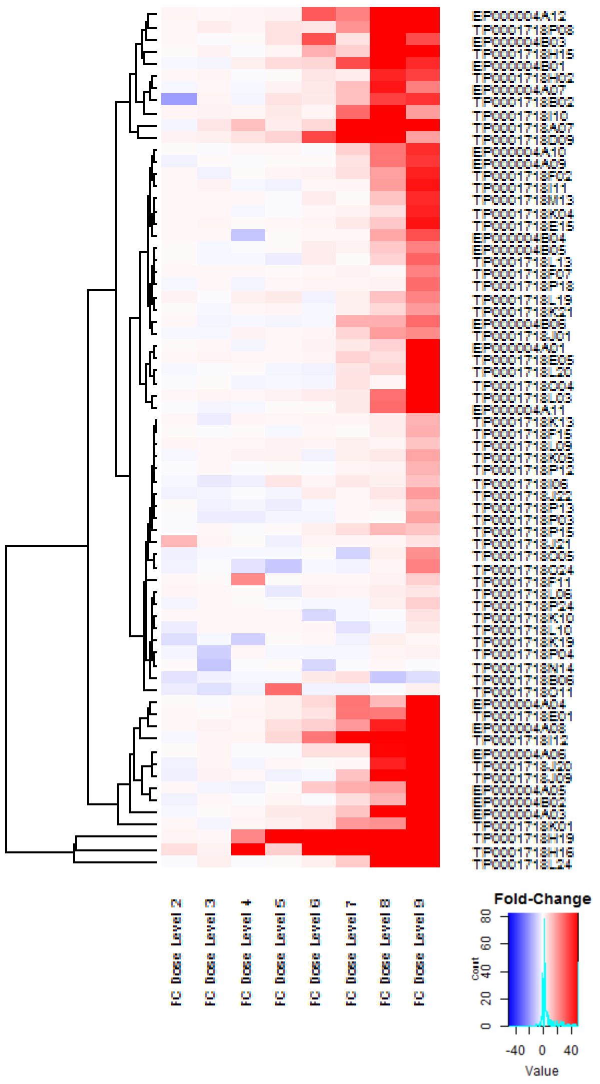
- **MSigDB_C2:** Curated gene sets from online pathway databases, publications and knowledge of domain experts (n = 4738).
- **MSigDB_H:** Coherently expressed signatures derived by aggregating many MSigDB gene sets to represent well-defined biological states or processes (n = 50).
- **Reactome:** Open-source, curated and peer reviewed pathway database with hierarchical pathway relationships in specific domains of biology. (n = 1764). Some pathways included in MSigDB_C2.
- **BioPlanet** (n = 1700): Curated pathway set developed by National Toxicology Program.

Benchmark Dose Modeling Summary

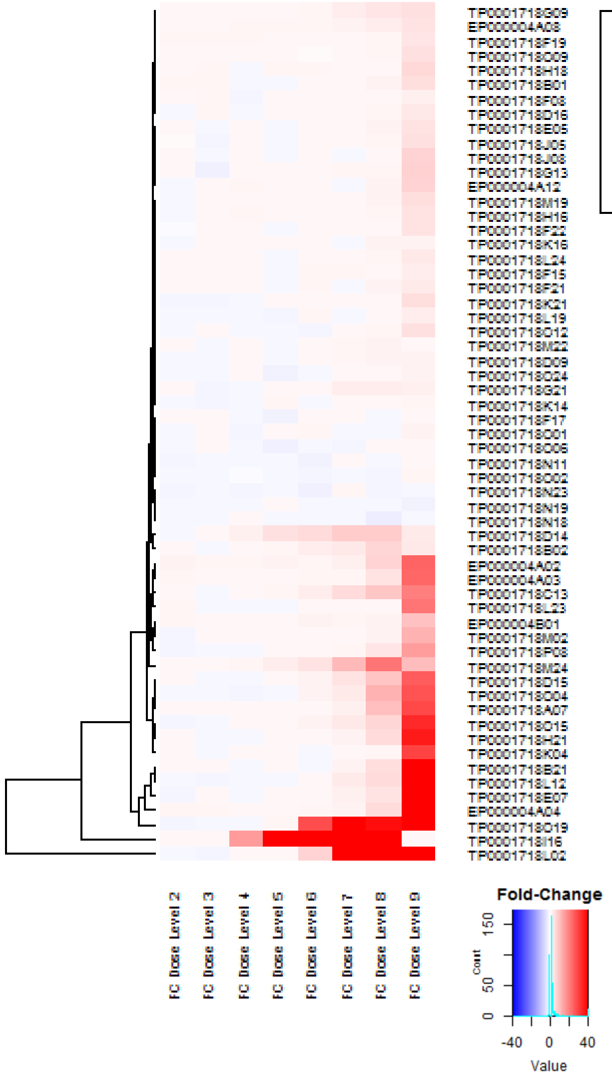




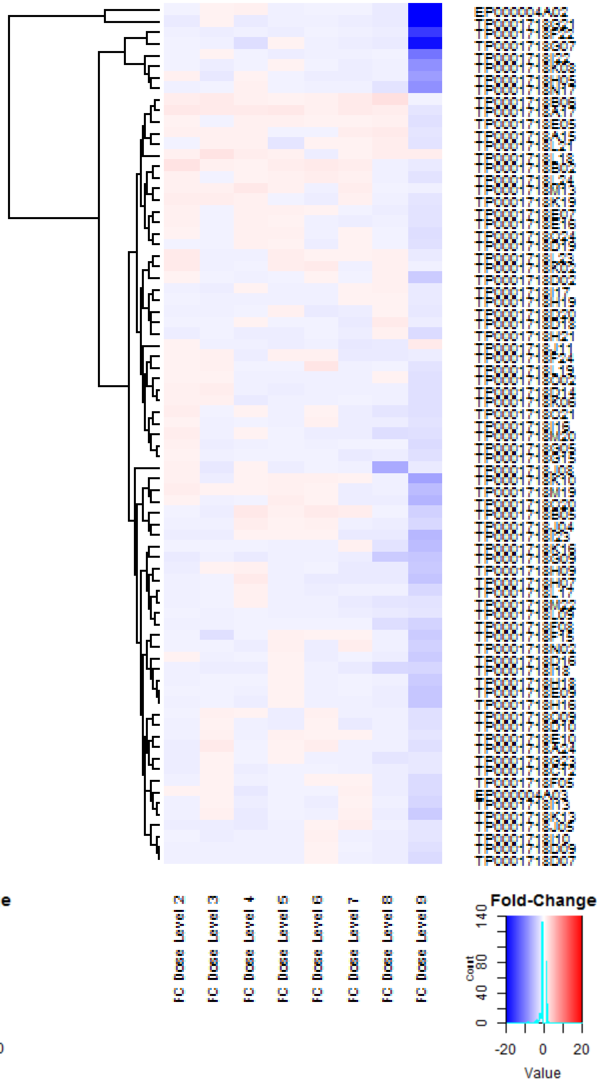
CYP1A1_10775



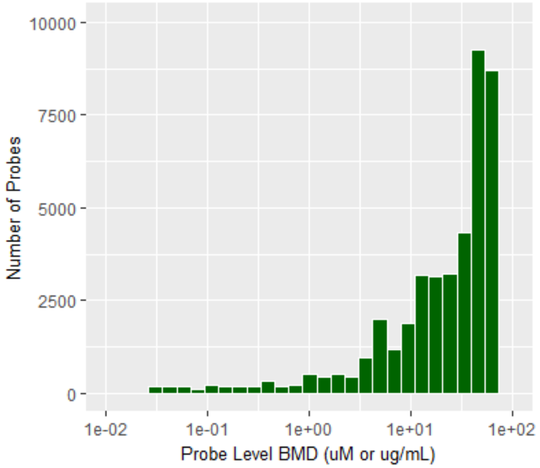
HMOX1_3041

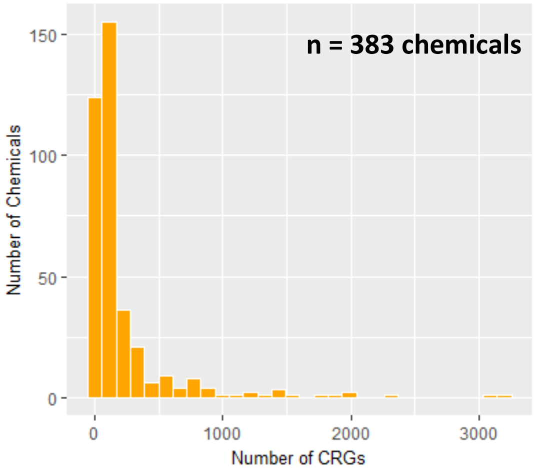


EIF4E2_16592

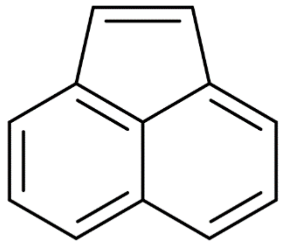


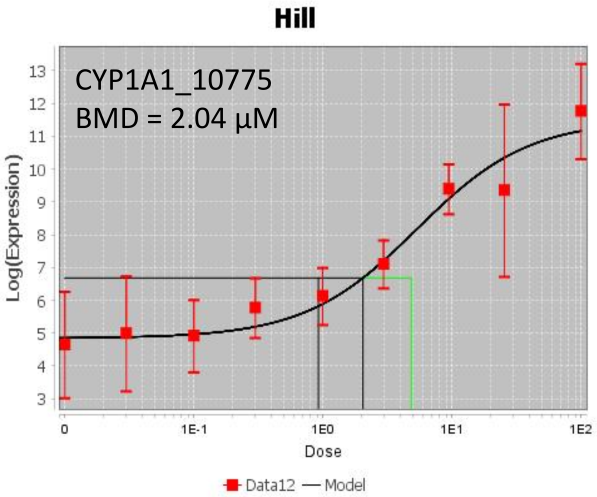
Benchmark Dose Modeling Summary



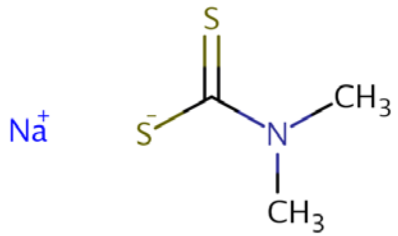


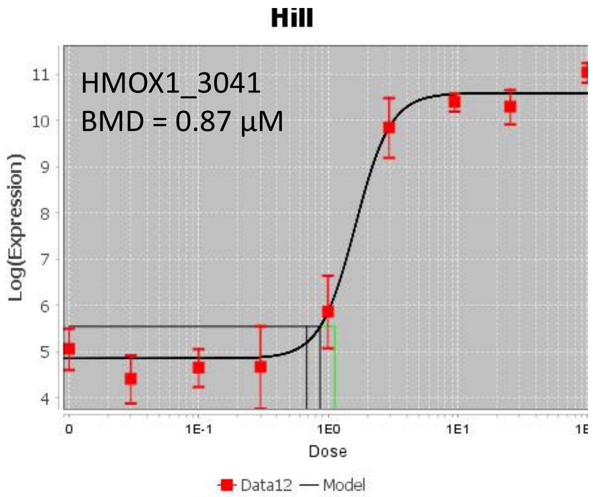
Acenaphthylene
208-96-8 | DTXSID3023845



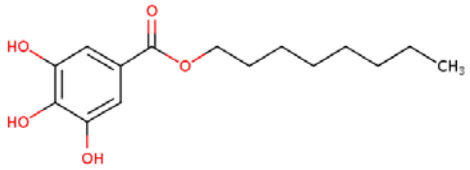


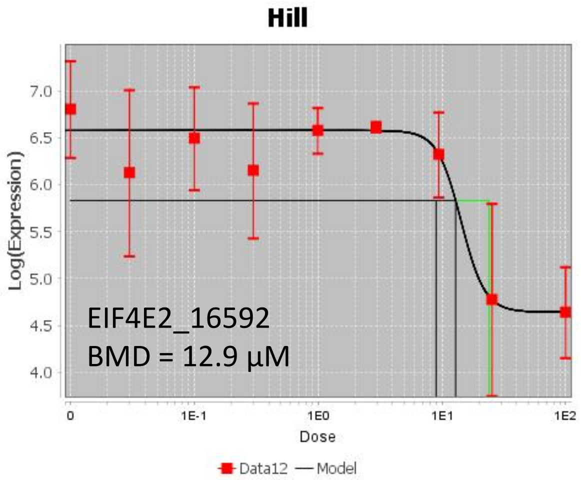
Sodium dimethyldithiocarbamate
128-04-1 | DTXSID6027050





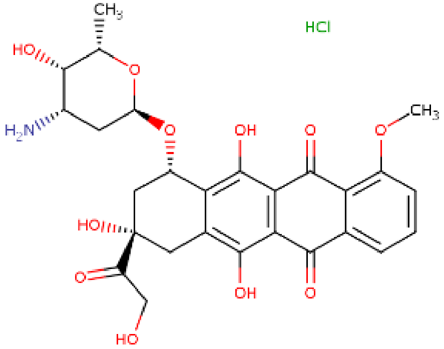
Octyl gallate
1034-01-1 | DTXSID4040713



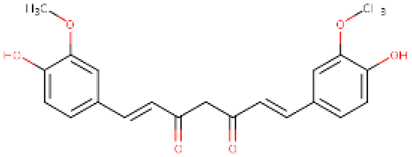


United States Environmental Protection Agency

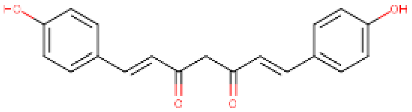
Case Study Chemicals



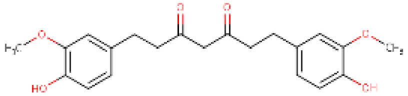
Adriamycin hydrochloride | DTXSID3030636



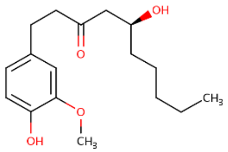
Curcumin | DTXSID8031077



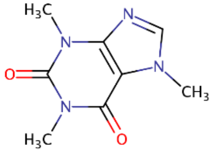
Bisdemethoxycurcumin | DTXSID00872663



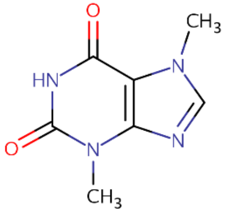
Tetrahydrocurcumin | DTXSID30865801



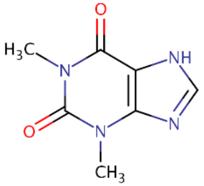
(6)-Gingerol | DTXSID3041035



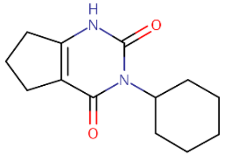
Caffeine | DTXSID0020232



Theobromine | DTXSID9026132

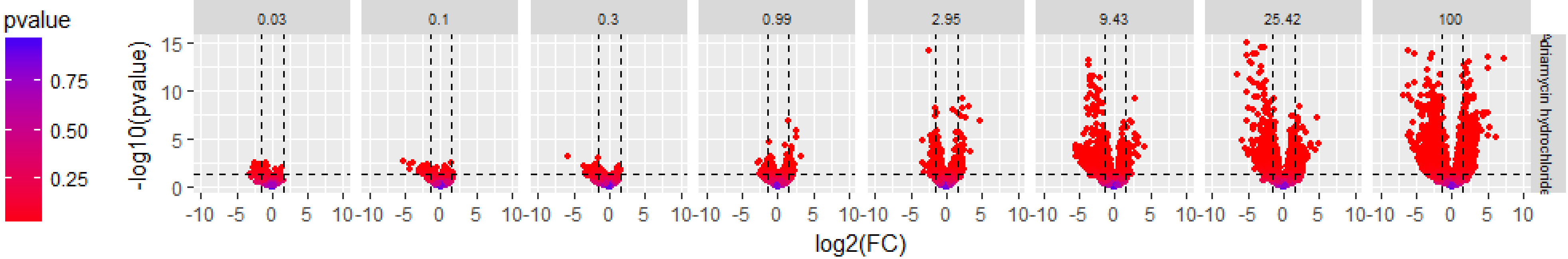


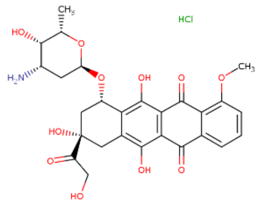
Theophylline | DTXSID5021336



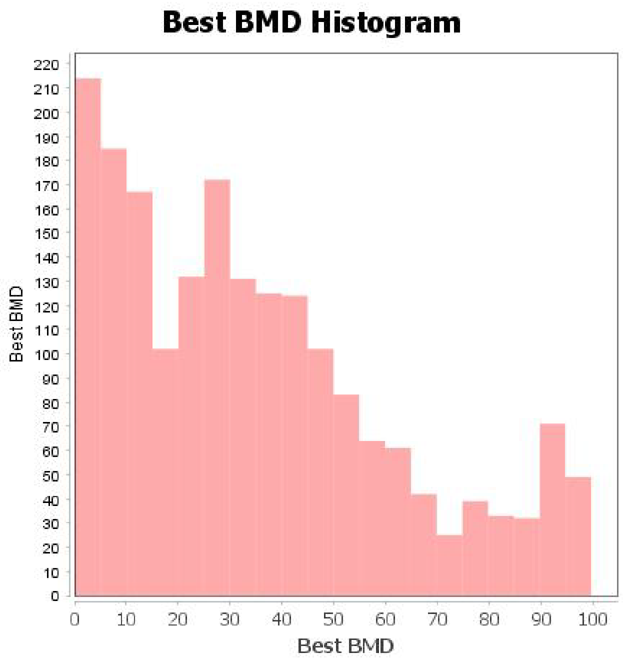
Lenacil | DTXSID9042093

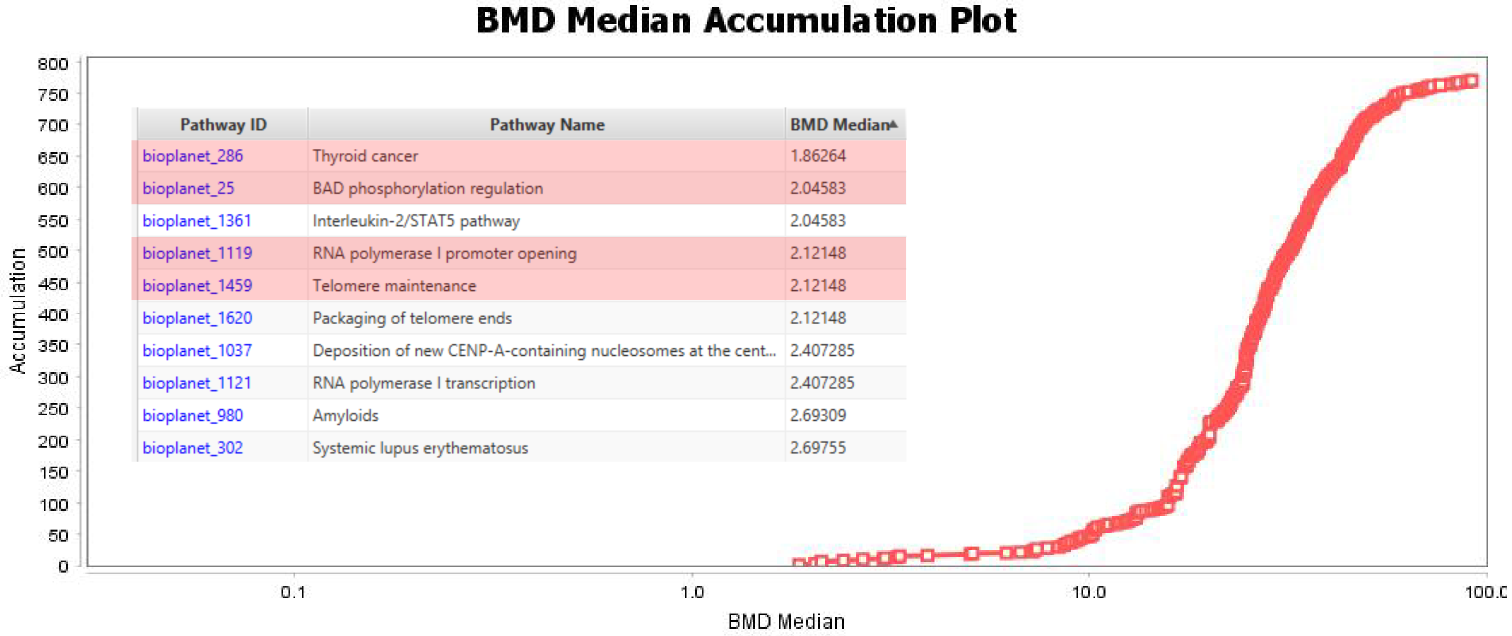
Doxorubicin (aka Adriamycin hydrochloride)





Adriamycin hydrochloride | DTXSID3030636

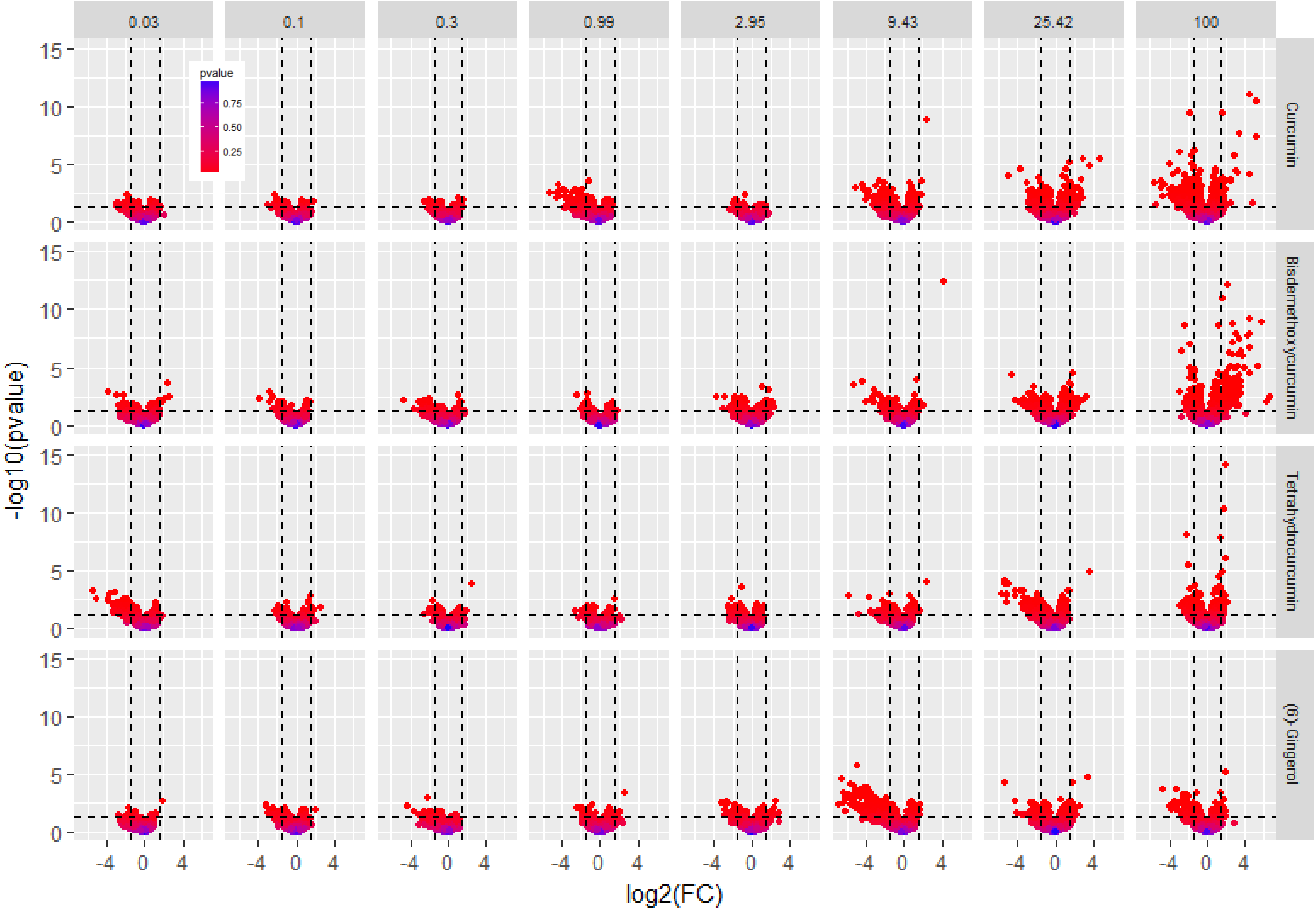


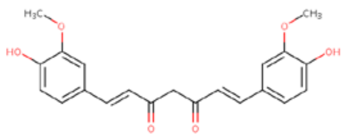


Pathway ID	Pathway Name	BMD Median
bioplanet_286	Thyroid cancer	1.86264
bioplanet_25	BAD phosphorylation regulation	2.04583
bioplanet_1361	Interleukin-2/STAT5 pathway	2.04583
bioplanet_1119	RNA polymerase I promoter opening	2.12148
bioplanet_1459	Telomere maintenance	2.12148
bioplanet_1620	Packaging of telomere ends	2.12148
bioplanet_1037	Deposition of new CENP-A-containing nucleosomes at the cent...	2.407285
bioplanet_1121	RNA polymerase I transcription	2.407285
bioplanet_980	Amyloids	2.69309
bioplanet_302	Systemic lupus erythematosus	2.69755

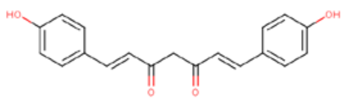
Int J Mol Cell Med. 2015 Spring;4(2):94-102.
Elevation of cAMP Levels Inhibits Doxorubicin-Induced Apoptosis in Pre- B ALL NALM- 6 Cells Through Induction of BAD Phosphorylation and Inhibition of P53 Accumulation.
Fatemi A[1], Kazemi A[1], Kashiri M[1], Safa M[2].

Evaluating Structurally Related Chemicals

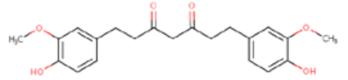




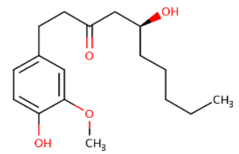
Curcumin | DTXSID8031077



Bisdemethoxycurcumin | DTXSID00872663



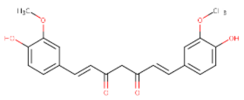
Tetrahydrocurcumin | DTXSID30865801



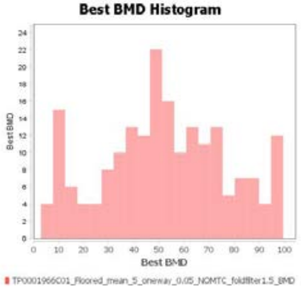
(6)-Gingerol | DTXSID3041035

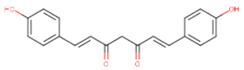
United States Environmental Protection Agency

Gene Set Analysis Using BMD Modeling Results

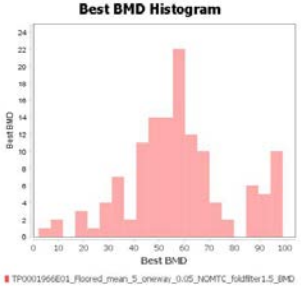


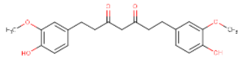
Curcumin | DTXSID8031077



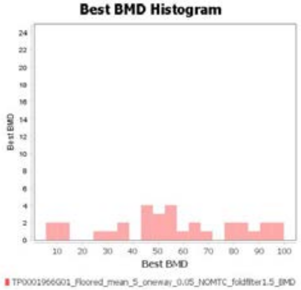


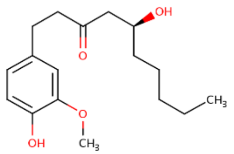
Bisdemethoxycurcumin | DTXSID00872663



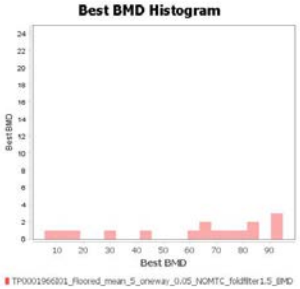


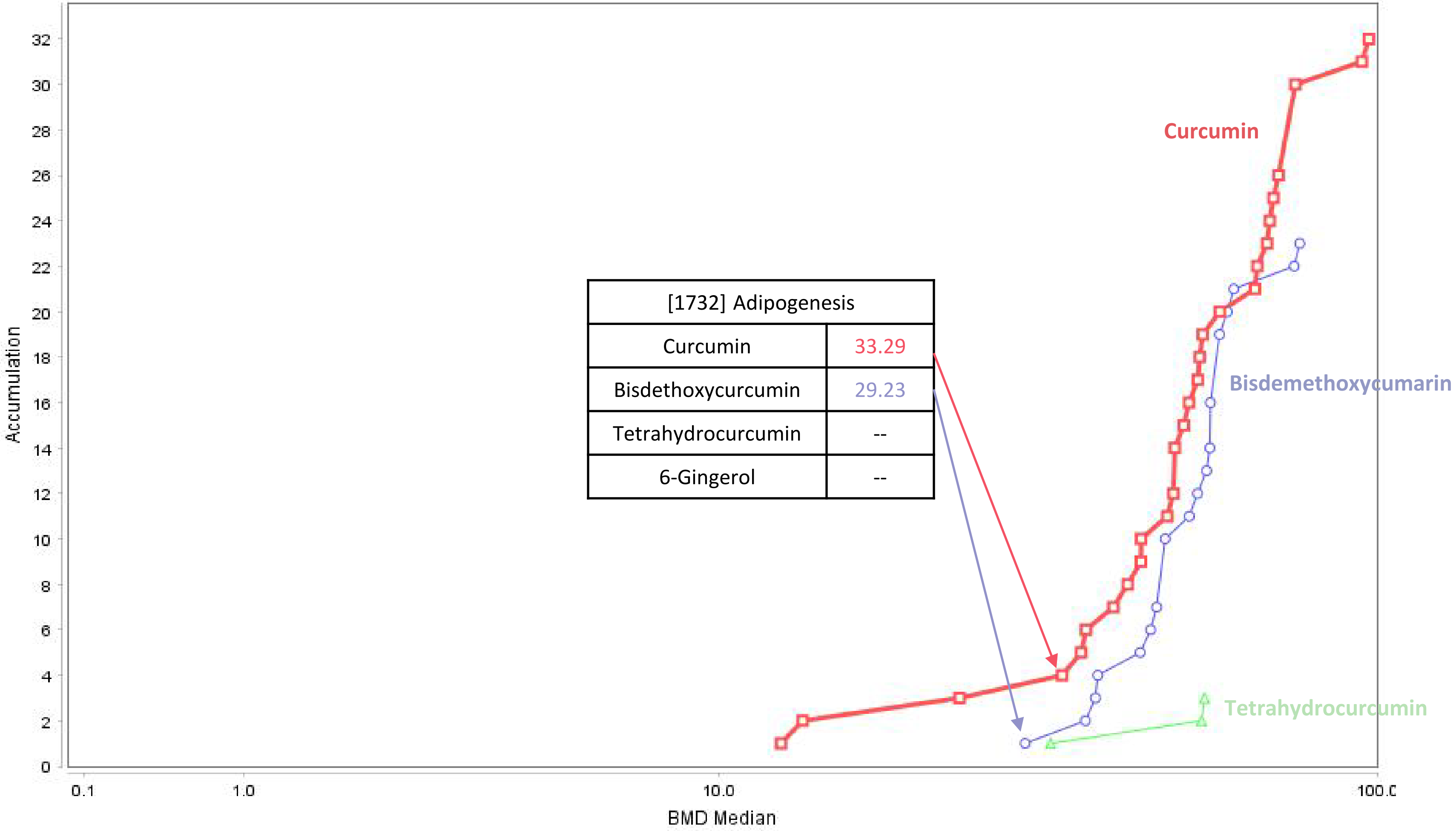
Tetrahydrocurcumin | DTXSID30865801





(6)-Gingerol | DTXSID3041035



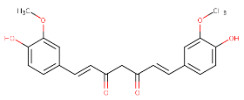


[1732] Adipogenesis	
Curcumin	33.29
Bisdethoxycurcumin	29.23
Tetrahydrocurcumin	--
6-Gingerol	--

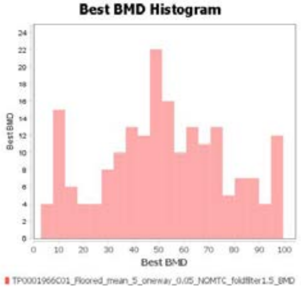
Toxicol Appl Pharmacol. 2017 Aug 15;329:158-164. doi: 10.1016/j.taap.2017.05.036.
Curcumin inhibits adipogenesis induced by benzyl butyl phthalate in 3T3-L1 cells.
Sakuma S[1], Sumida M[2], Endoh Y[2], Kurita A[2], Yamaguchi A[2], Watanabe T[2], Kohda T[2], Tsukiyama Y[2], Fujimoto Y[3].

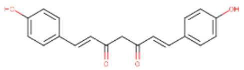
J Agric Food Chem. 2016 Feb 3;64(4):821-30. doi: 10.1021/acs.jafc.5b05577. Epub 2016 Jan 25.
Bisdemethoxycurcumin Inhibits Adipogenesis in 3T3-L1 Preadipocytes and Suppresses Obesity in High-Fat Diet-Fed C57BL/6 Mice.
Lai CS[1,2], Chen YY[1], Lee PS[1], Kalyanam N[3], Ho CT[4], Liou WS[5], Yu RC[1], Pan MH[1,6,7,8].

Gene Set Analysis Using BMD Modeling Results

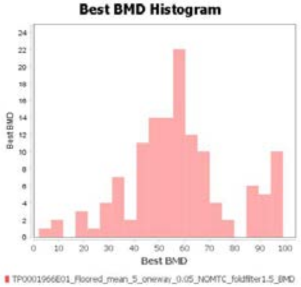


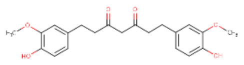
Curcumin | DTXSID8031077



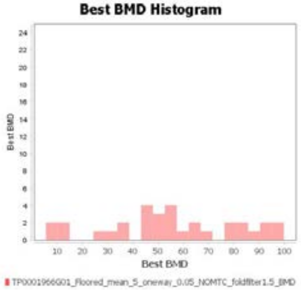


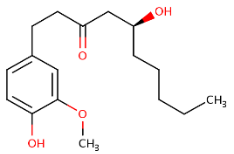
Bisdemethoxycurcumin | DTXSID00872663



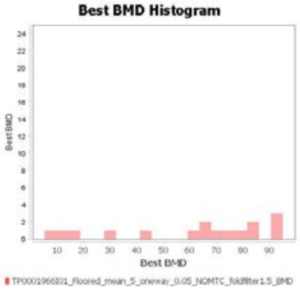


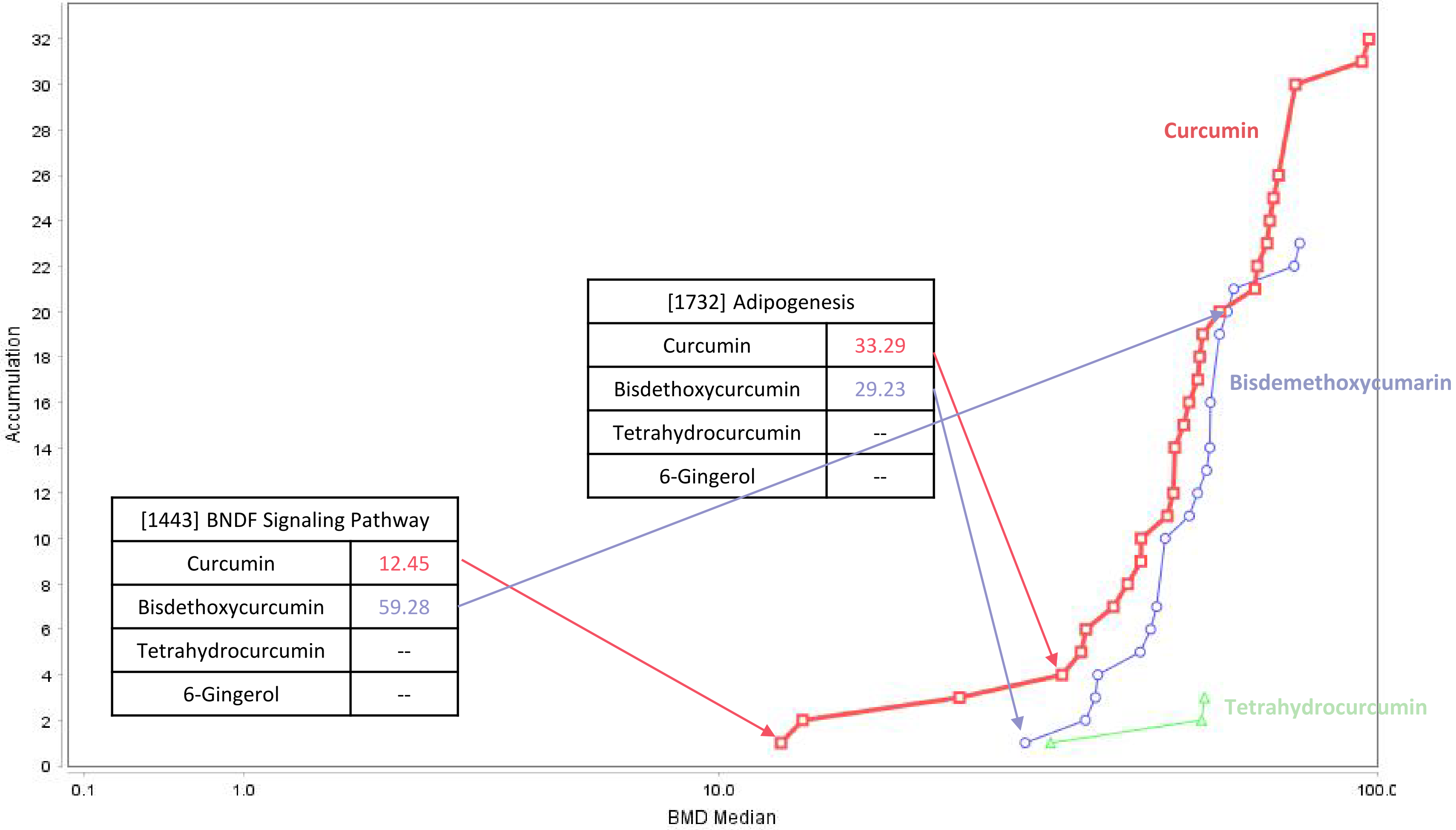
Tetrahydrocurcumin | DTXSID30865801





(6)-Gingerol | DTXSID3041035





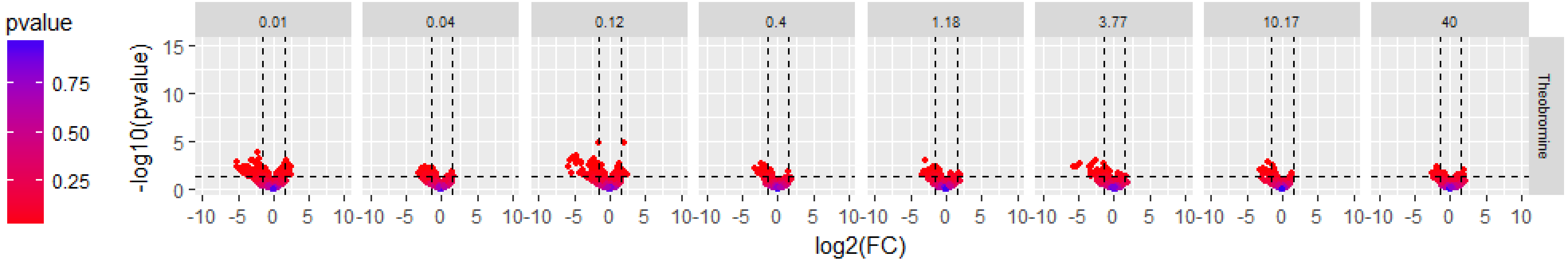
[1732] Adipogenesis	
Curcumin	33.29
Bisdethoxycurcumin	29.23
Tetrahydrocurcumin	--
6-Gingerol	--

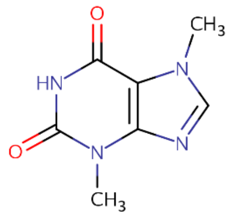
[1443] BNDF Signaling Pathway	
Curcumin	12.45
Bisdethoxycurcumin	59.28
Tetrahydrocurcumin	--
6-Gingerol	--

Neuropeptides. 2016 Apr;56:25-31. doi: 10.1016/j.npep.2015.11.003. Epub 2015 Nov 11.
Effect of curcumin on serum brain-derived neurotrophic factor levels in women with premenstrual syndrome: A randomized, double-blind, placebo-controlled trial. Fanaei H[1], Khayat S[2], Kasaeian A[3], Javadimehr M[4].

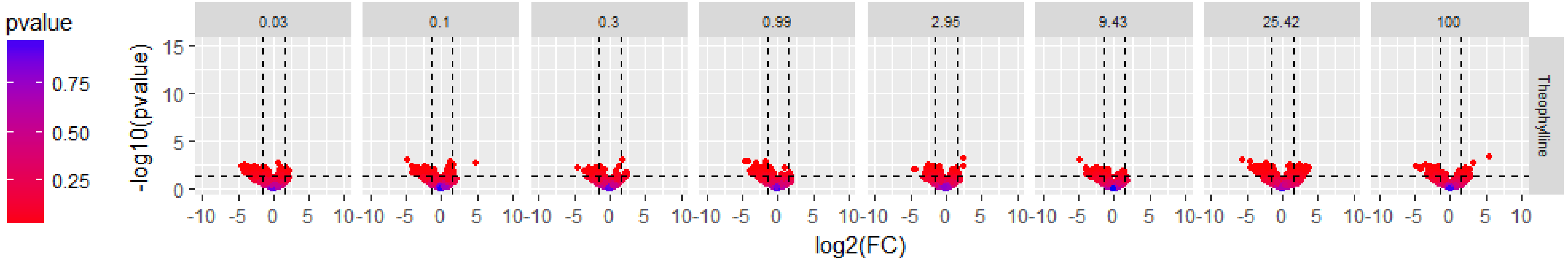
Biomed Pharmacother. 2017 Mar;87:721-740. doi: 10.1016/j.biopha.2016.12.020. Epub 2017 Jan 14.
Curcumin confers neuroprotection against alcohol-induced hippocampal neurodegeneration via CREB-BDNF pathway in rats. Motaghinejad M[1], Motevalian M[2], Fatima S[3], Hashemi H[1], Gholami M[4].

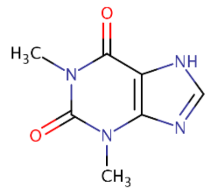
Caffeine Analogs



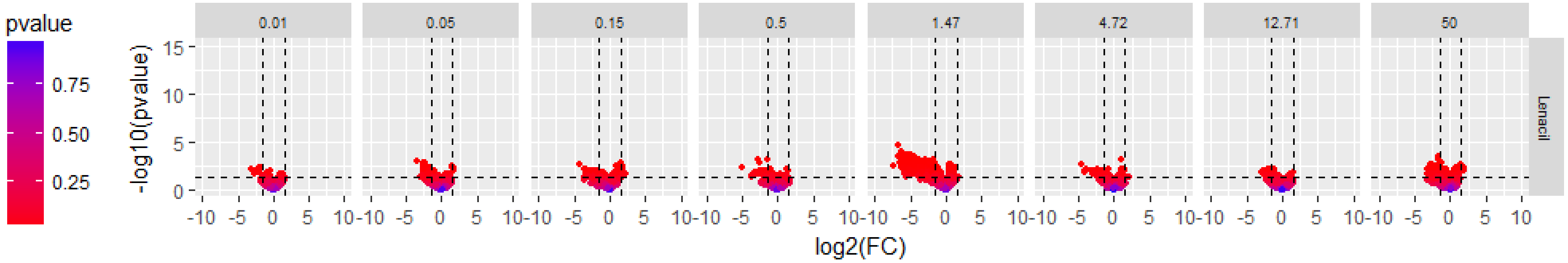


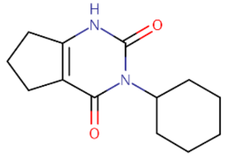
Theobromine | DTXSID9026132





Theophylline | DTXSID5021336





Lenacil | DTXSID9042093

- Very few dose-responsive genes
- No significantly enriched pathways using BMDe standard workflow

Alternative Approach for Gene Set Analysis

Step 1: Calculate Response

- A gene set is a list / bag of genes
- Under one condition (chemical x dose) calculate "gene set response" separately for genes in the set and out of the set:

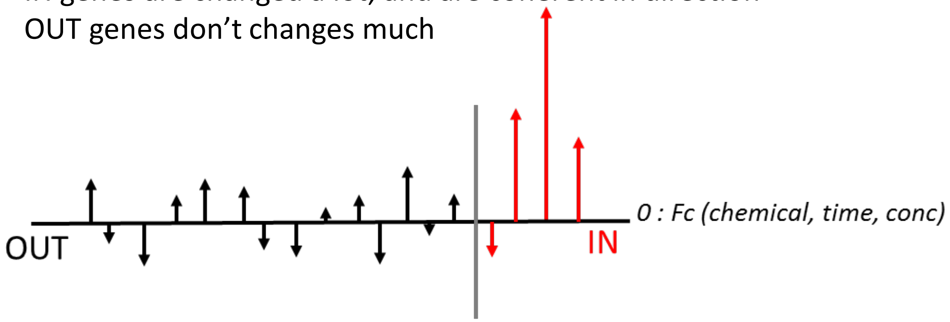
$$M = \sum_{i=1}^{ngene} \frac{\mathrm{fc}_i}{\mathrm{sc}_i^2} \Bigg/ \sum_{i=1}^{ngene} \frac{1}{\mathrm{sc}_i^2}$$

$$\mathrm{R} = M_{in} - M_{out}$$

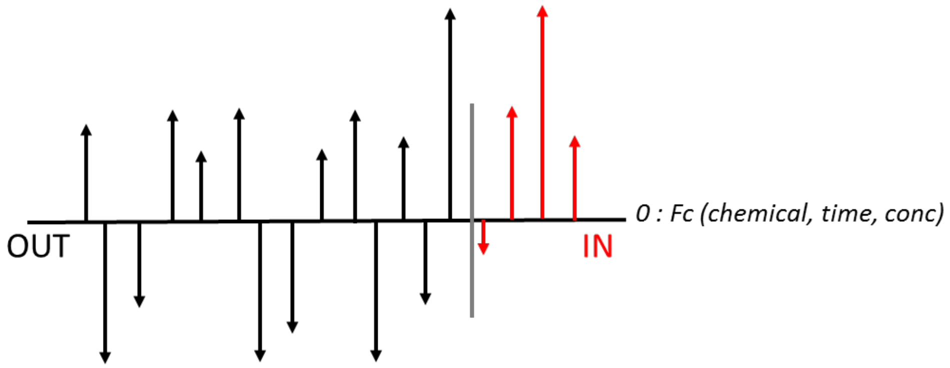
Step 2: CR Modeling

- For each chemical, fit using tcplFit
  - Constant, Hill , Gain-Loss methods
  - BMAD(pathway) = MAD of response for the pathway across the two lowest concentrations across all chemicals and times
- Hitcall:
  - tcplFit calls a hit
  - Top > 3*BMAD

- IN genes are changed a lot, and are coherent in direction
- OUT genes don't changes much

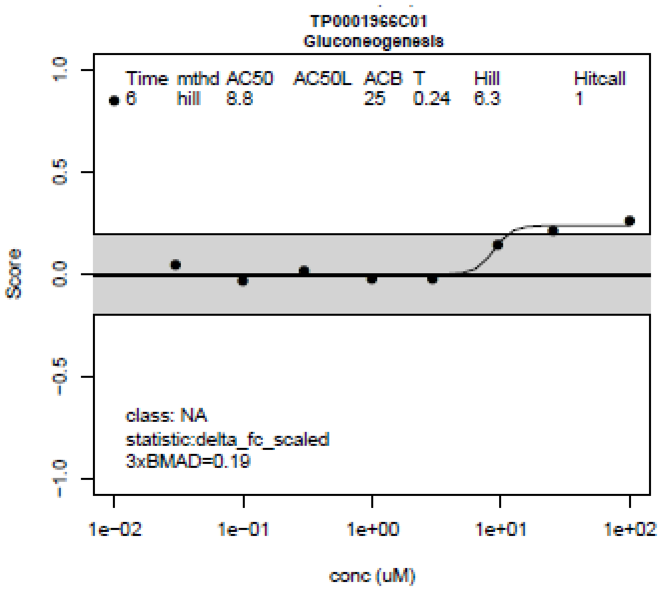


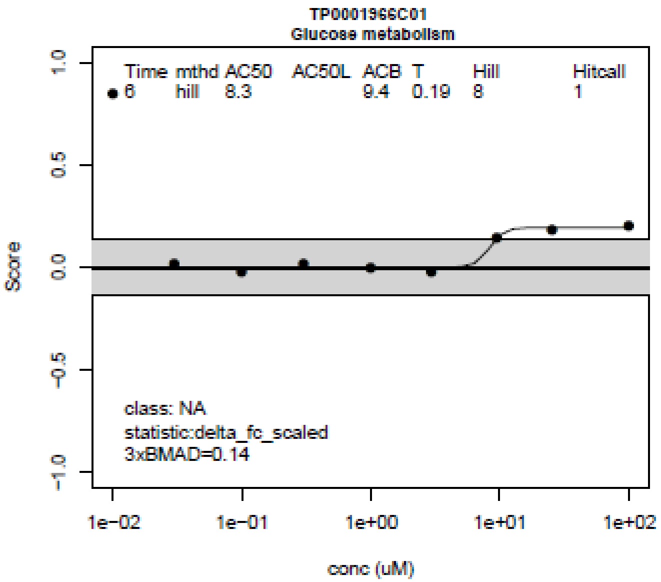
- IN genes are changed a lot, and are coherent in direction
- OUT genes change a lot but are not coherent (mean ~ 0)

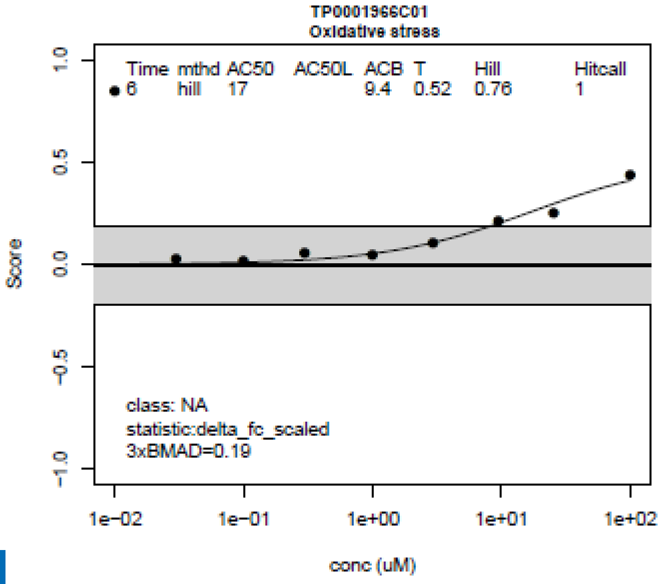


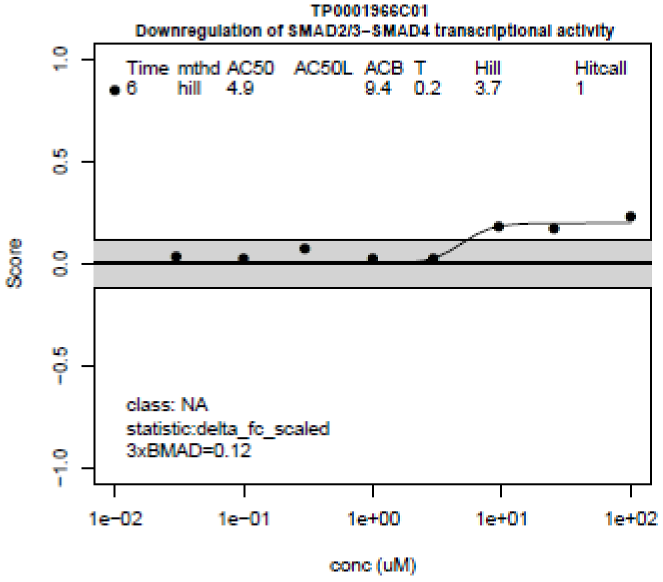
EPA United States Environmental Protection Agency

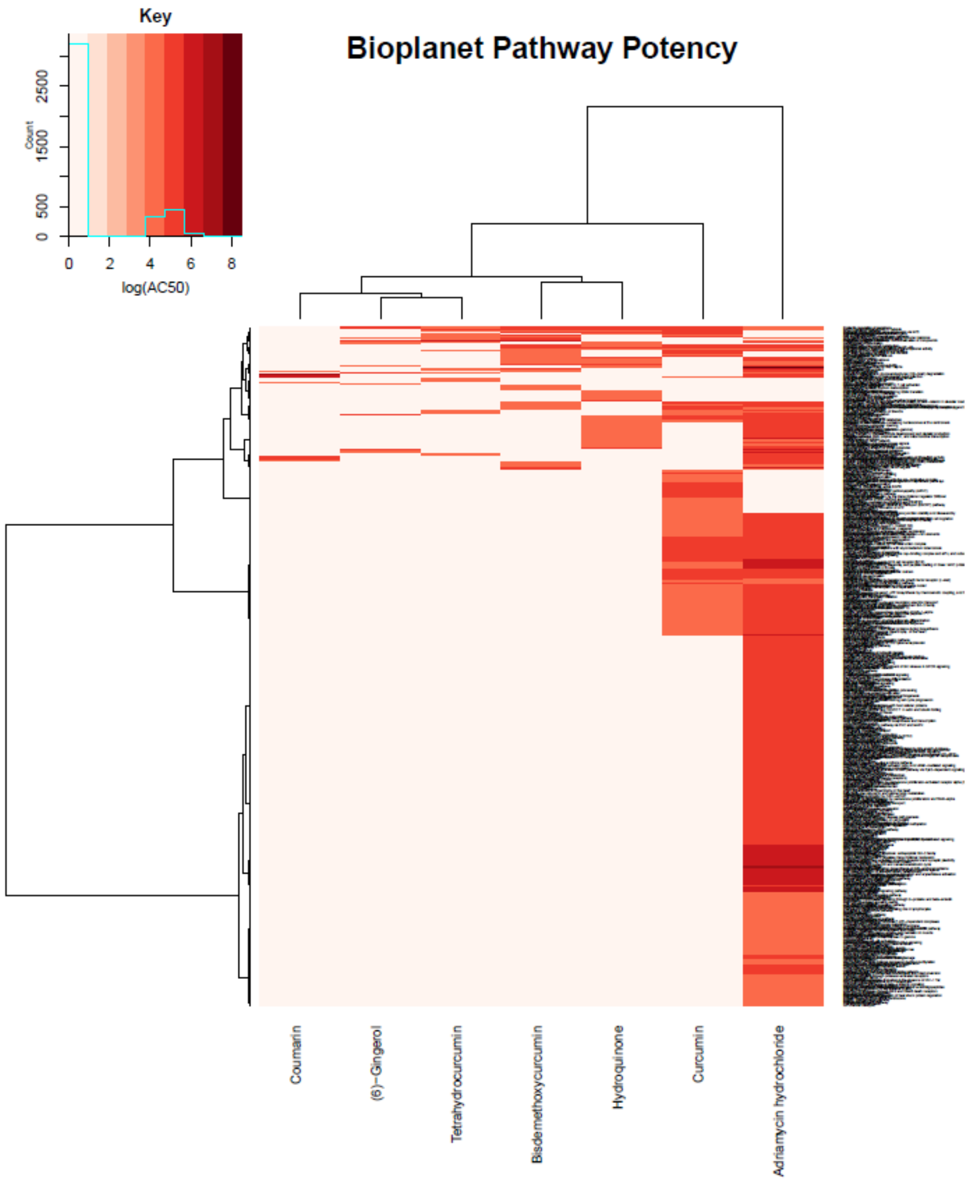
Comparing Activities Across Chemicals











High Throughput Phenotypic Profiling via High Content Image Analysis

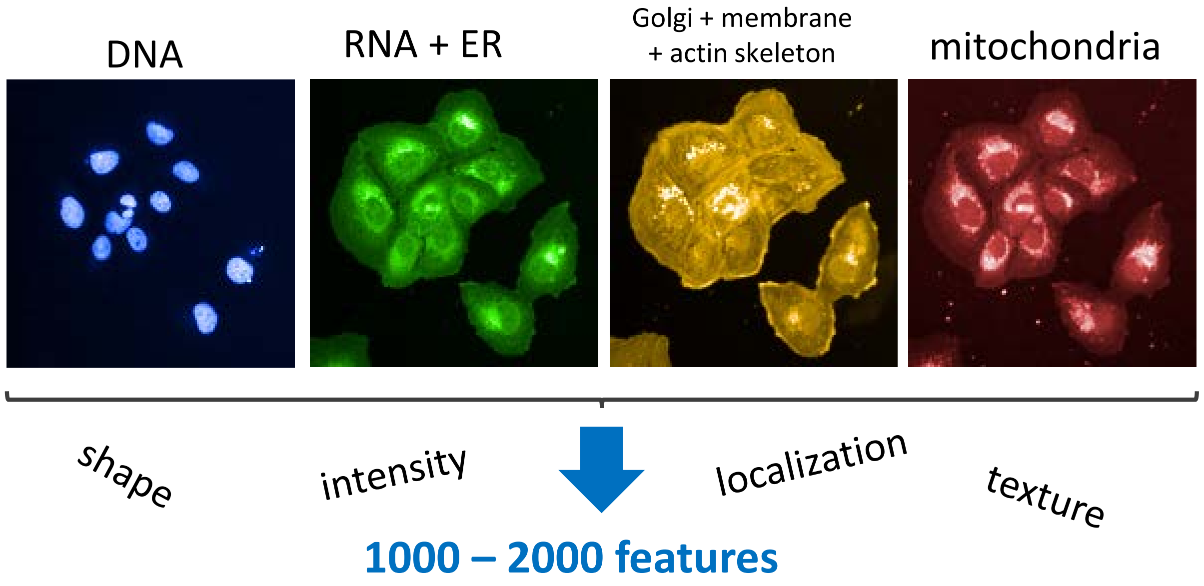
Phenotypic Profiling

- Image-based phenotypic profiling is a chemical screening method that measures a large variety of morphological features of individual cells in *in vitro* cultures.

- Successfully used for functional genomic studies and in the pharmaceutical industry for compound efficacy and toxicity screening.

- No requirement for *a priori* knowledge of molecular targets.

- **May be used as an efficient and cost-effective method for evaluating the chemical bioactivity.**



- **Cell Painting (Bray et al., 2016, *Nature Protocols*):** A cell morphology-based phenotypic profiling assay multiplexing six fluorescent "non-antibody" labels, imaged in five channels, to evaluate multiple cellular compartments and organelles.

Experimental Workflow

Marker	Cellular Component	Labeling Chemistry	Labeling Phase	Opera Phenix	
				Excitation	Emission
Hoechst 33342	Nucleus	Bisbenzamide probe that binds to dsDNA	Fixed	405	480
Concanavalin A – AlexaFluor 488	Endoplasmic reticulum	Lectin that selectively binds to α-mannopyranosyl and α-glucopyranosyl residues enriched in rough endoplasmic reticulum		435	550
SYTO 14 nucleic acid stain	Nucleoli	Cyanine probe that binds to ssRNA		435	550
Wheat germ agglutinin (WGA) – AlexaFluor 555	Golgi Apparatus and Plasma Membrane	Lectin that selectively binds to sialic acid and N-acetylglucosaminyl residues enriched in the trans-Golgi network and plasma membrane		570	630
Phalloidin –AlexaFluor 568	F-actin (cytoskeleton)	Phallotoxin (bicyclic heptapeptide) that binds filamentous actin			
MitoTracker Deep Red	Mitochondria	Accumulates in active mitochondria	Live	650	760

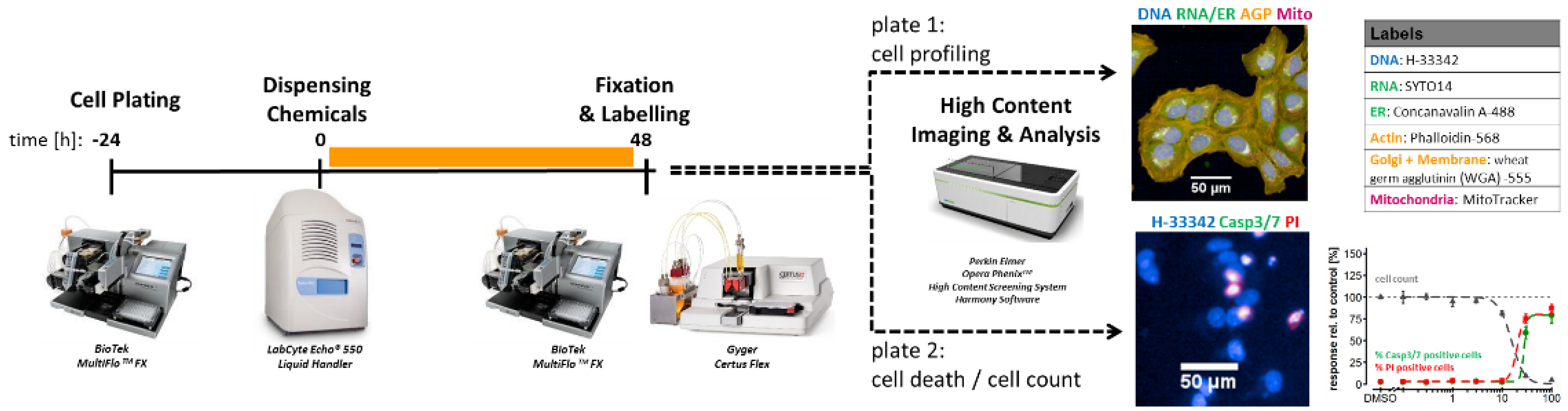


Image Analysis Workflow

Image Acquisition

<u>Image Acquisition</u>

- Perkin Elmer Opera Phenix
- 20x Water Immersion Objective
- Confocal Mode, Single Z
- CellCarrier-384 Ultra Microplates

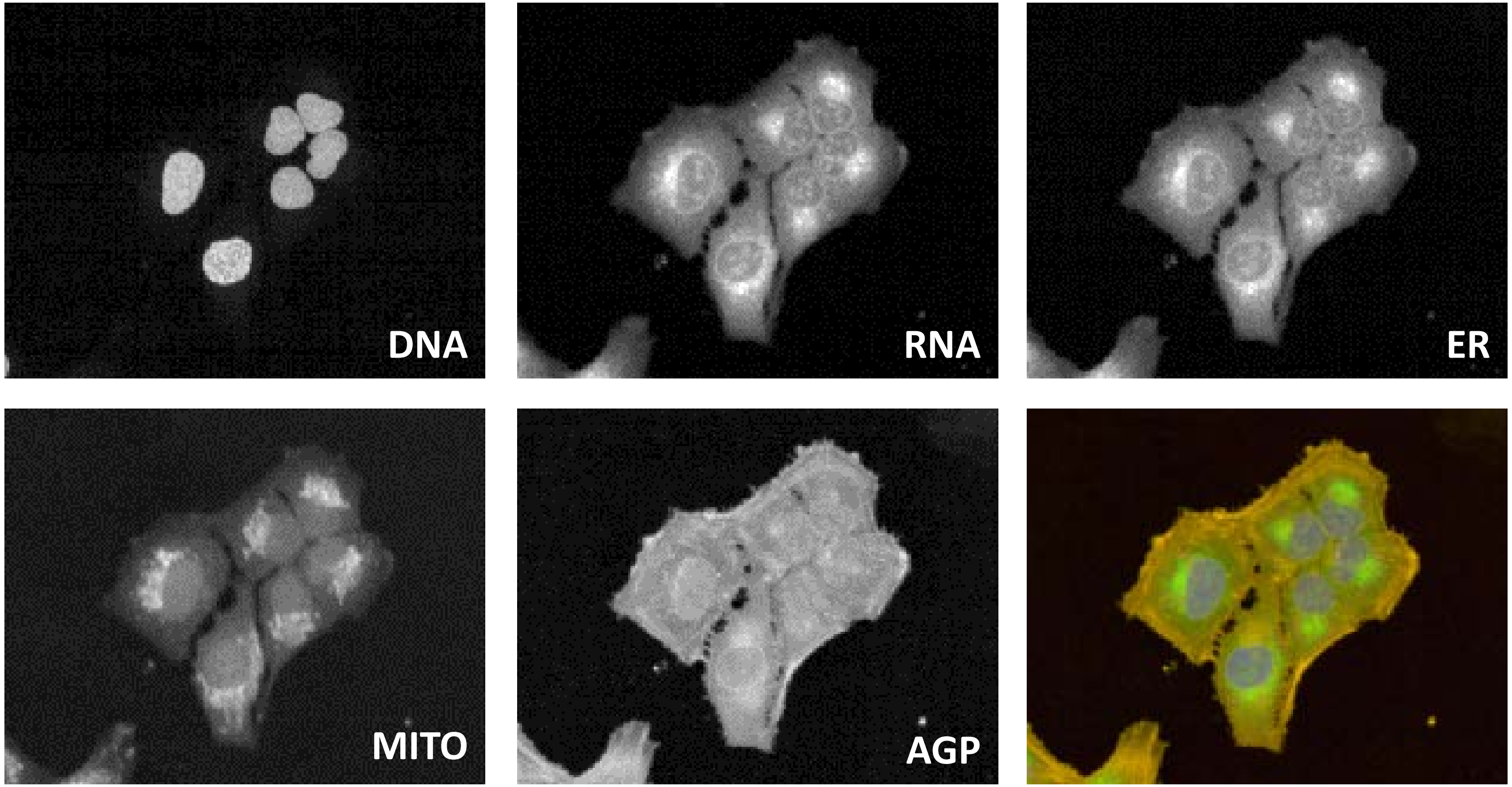


Image Analysis Workflow:
Nucleus and Cell Segmentation

1. find nuclei

2. find cell outline

3. reject border objects

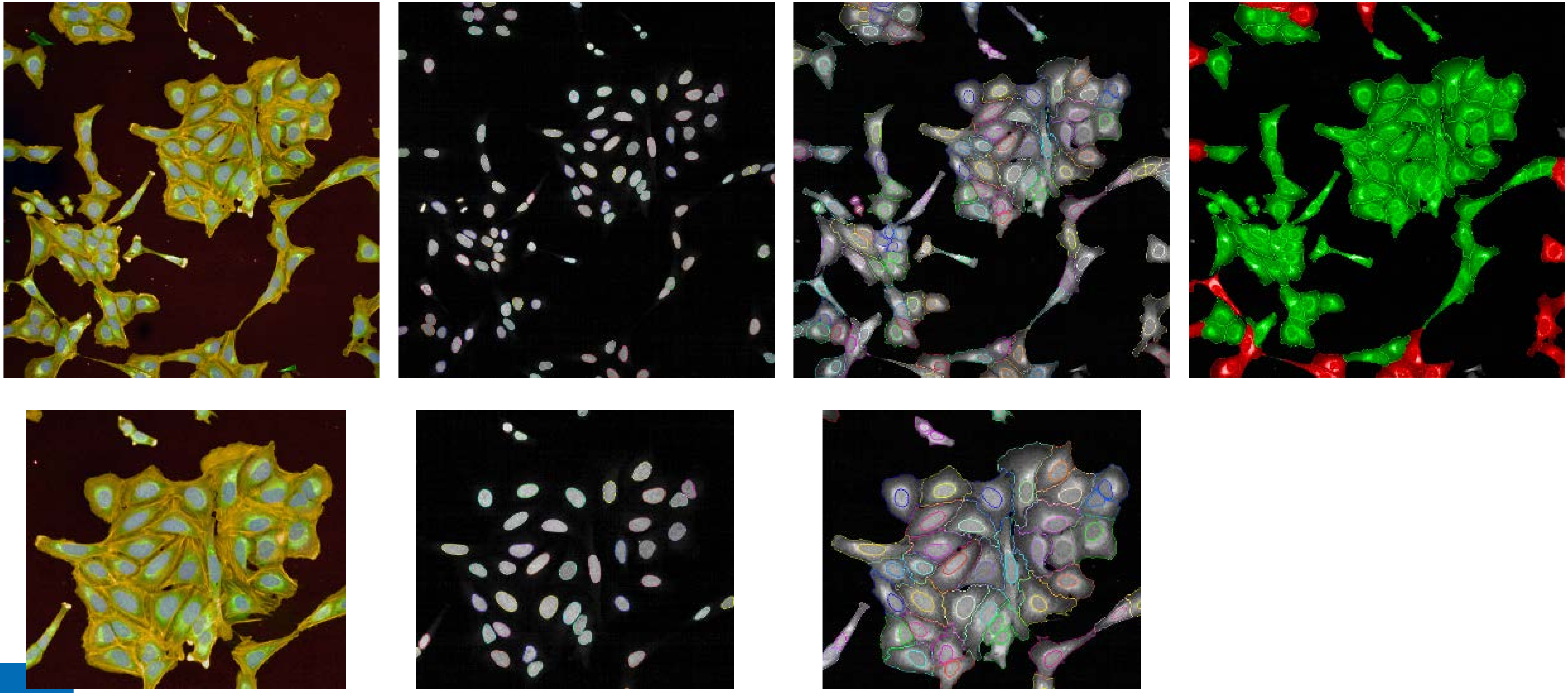
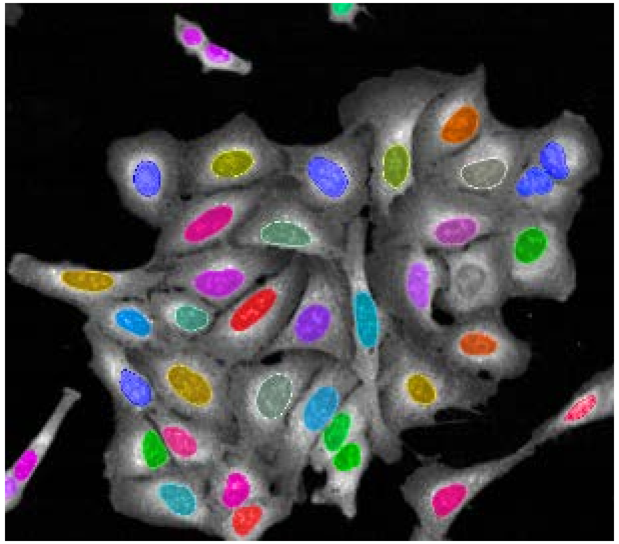


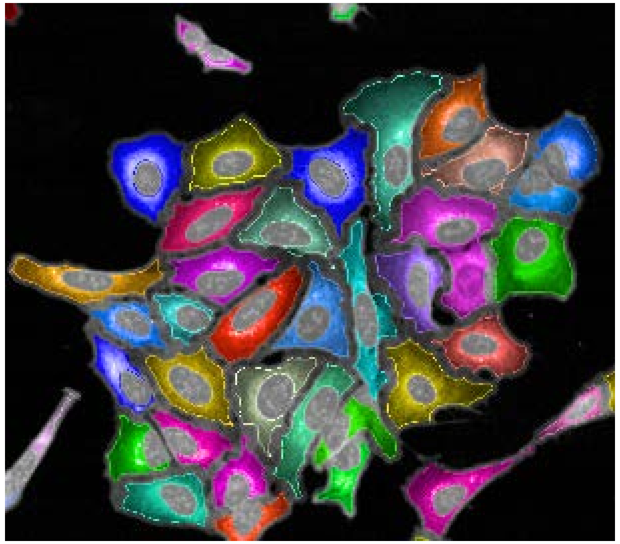
Image Analysis Workflow

Define Cellular Compartments

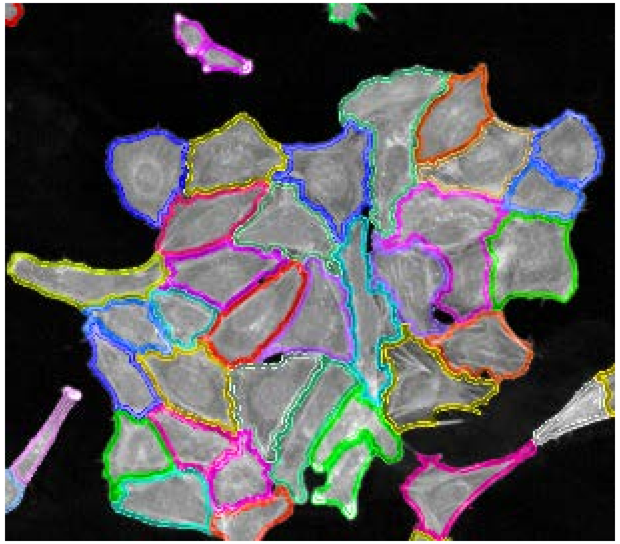
nuclei



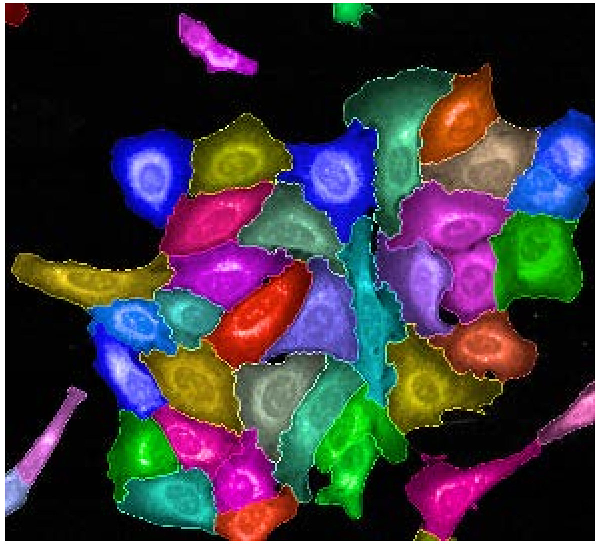
cytoplasm



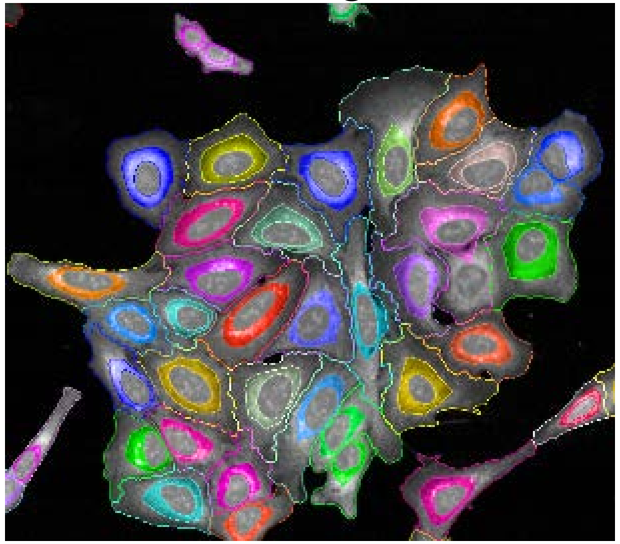
membrane



cell



ring



Assay Outputs & Data Analysis

compartments:
- nuclei
- ring
- cytoplasm
- membrane
- cell

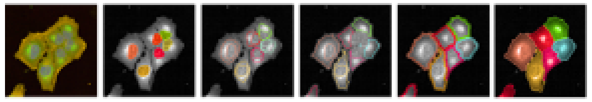
X

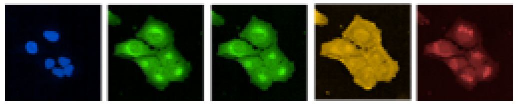
channels (organelles):
- DNA
- RNA
- ER
- AGP (actin skeleton / Golgi/ plasma membrane)
- mitochondria

X

parameters:
- intensity
- texture
- morphology:
  - symmetry
  - compactness
  - axial
  - radial
  - profile

= 1700 endpoints



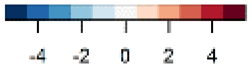


Data Reduction

cell-level data (~500 cells/well)

$$\frac{\text{cell value} - \text{median}_{DMSO}}{1.4826\ \text{MAD}_{DMSO}}$$

normalized cell-level data



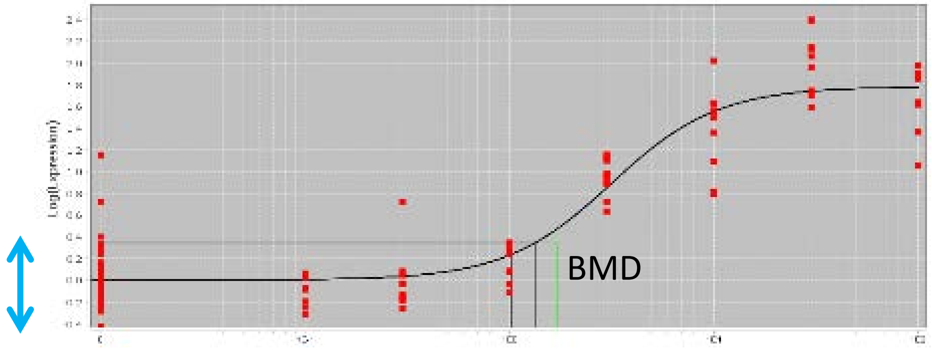
median

well-level data

BMD Modelling

- Well-level data x 3 technical replicates x 3 biological replicates = 9 values
- Filtered for affected parameters using ANOVA (p $\leq$ 0.01, FDR adjusted)
- BMD modelling with BMDExpress 2.0
  - 3 models: Hill, Power, Poly2
  - model with best logLikelihood selected

BMR: 10%



Objectives & Experimental Design

1. Miniaturize an existing assay (Bray et al. 2016) and establish a microfluidics-based laboratory workflow suitable for high-throughput screening purposes.

2. Test a set of 14 phenotypic reference and 2 negative compounds in seven cell lines.

3. Evaluate the applicability of the assay for:
   a) grouping of chemicals with similar biological effects
   b) derivation of *in vitro* point-of-departures (POD)

4. Identify chemicals & test concentrations for use as reference chemical controls in screening applications.

Parameter	Multiplier	Notes
Cell Type(s)	7	U2 OS, MCF7, HepG2, LNCap, A549, ARPE-19, HTB-9
Culture Condition	1	DMEM + 10% HI-FBS [a]
Chemicals	16	14 phenotypic reference chemicals, 2 negative controls
Time Points:	1	48 hours
Assay Formats:	2	Cell Painting HCI Cell Viability & Apoptosis
Concentrations:	8	3.5 $\log_{10}$ units; semi $\log_{10}$ spacing
Biological Replicates:	3	Independent cultures

Reference Chemical Set

- Reference chemicals (n=14) with narrative descriptions of observed phenotypes were identified from Gustafdottit et al. 2013.
- Candidate negative control chemicals (n=2) with no anticipated affect on cell phenotype were included in the reference set.

Compound Name	Chemical Use	Expected Phenotype
Amperozide	Atypical antipsychotic	Toroid nuclei
Berberine Chloride	Mitochondria complex I inhibitor	Redistribution of mitochondria
Ca-074-Me	Cathepsin B inhibitor	Bright, abundant golgi staining
Etoposide	Chemotherapeutic	Large, flat nucleoli
Fenbendazole	Anthelmintic	Giant, multi-nucleated cells
Fluphenazine	Typical antipsychotic	Enhanced golgi staining and some cells with fused nucleoli
Latrunculin B	Actin cytoskeleton disruptor	Actin breaks
Metoclopramide	D_2 dompaine receptor antagonist	Enhanced golgi staining and some cells with fused nucleoli
NPPD	Chloride channel blocker	Redistribution of ER to one side of the nucleus
Oxibendazole	Anthelmintic	Large, multi-nucleated cells with fused nucleoli
Rapamycin	Macrolide antibiotic / antifungal	Reduced nucleolar size
Rotenone	Mitochondria complex I inhibitor	Mitochondrial stressor
Saccharin	Artificial Sweetener	Negative Control
Sorbitol	Artificial Sweetener	Negative Control
Taxol	Microtubule Stabilizer	Large, multi-nucleated cells with fused nucleoli
Tetrandrine	Calcium channel blocker	Abundant ER

Gustafsdottir, et al. 2013

Phenotypic Profiles in U-2 OS Cells

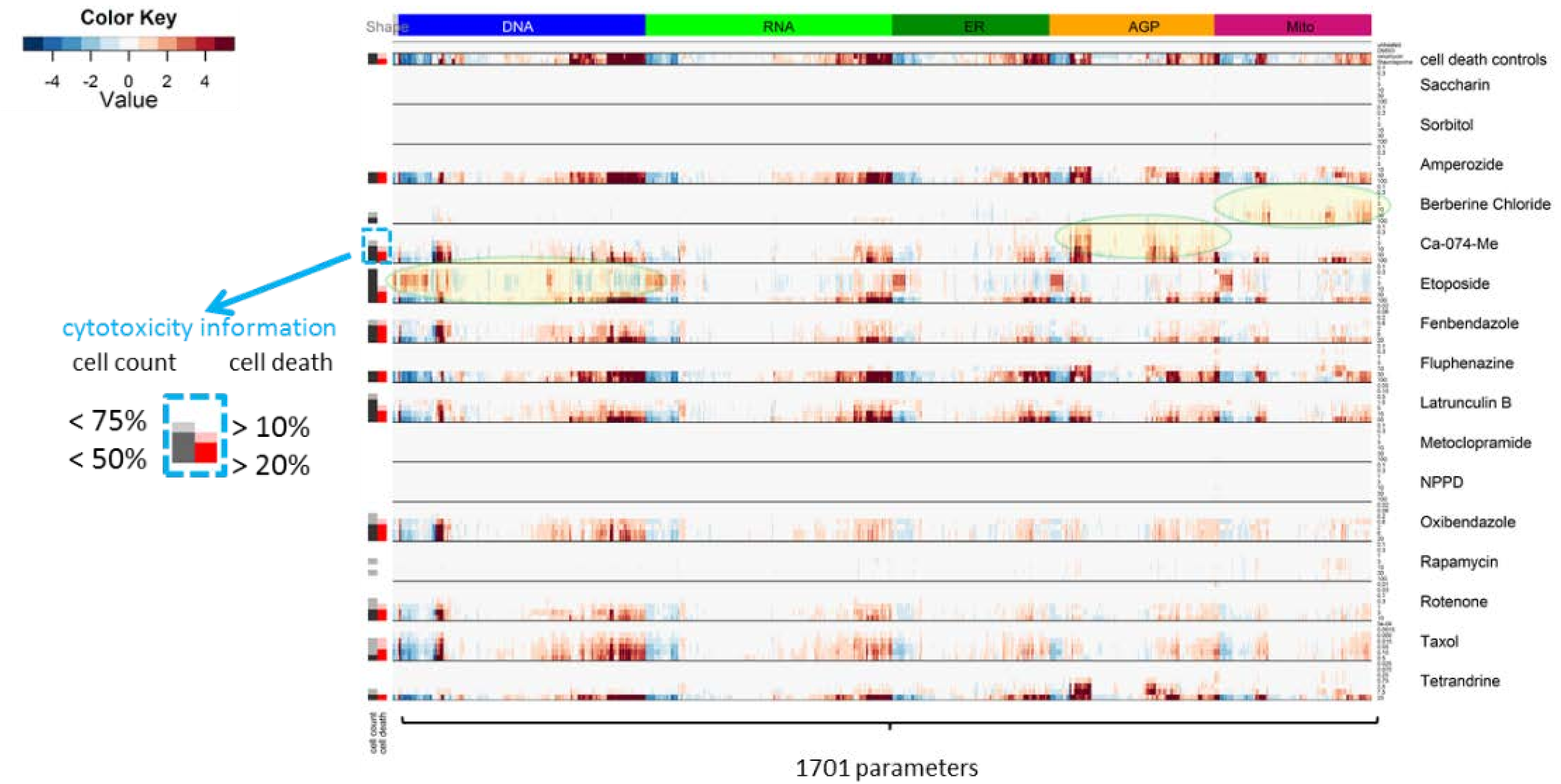


Fig 1.: MAD normalized well-level data of U-2 OS cells were averaged across 3 technical and 3 biological replicates. Endpoints are ordered according to the corresponding channel/organelle. The color key on the left indicates reductions in cell count and increases in cell death. Treatment with different chemicals resulted in distinct profiles. Some effects observed at non-cytotoxic concentrations.

EPA United States Environmental Protection Agency

Phenotypic Profiles Are Consistent with Observed Effects

Parameters with marked effects:

Channel	Compartment	Domain
Mito	Cytoplasm	Texture
Mito	Cytoplasm + Ring	Intensity: Maximum
Mito	Entire Cell	Compactness (of the bright spots)

Literature: redistribution of mitochondria

Channel	Compartment	Domain
AGP	Cytoplasm + Ring	Texture
AGP	Cytoplasm + Ring	Intensity: Maximum
AGP	Entire Cell	Profile (=distribution of intensity)

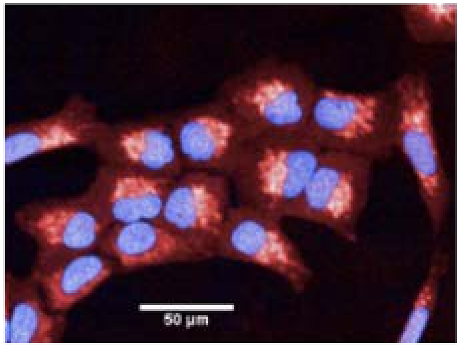
Literature: bright, abundant Golgi stain

Channel	Compartment	Domain
“Shape”	Entire Cell	Morphology: Area, Length, Width
DNA + RNA	Nuclei	Compactness (of the bright spots) Texture
ER + AGP	Cytoplasm + Ring	Intensity: Sum
all	Entire Cell	Morphology: intensity distribution

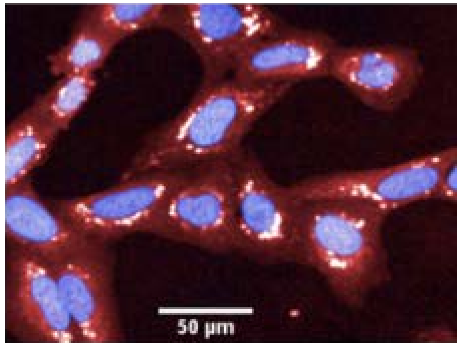
Literature: large, flat nucleoli

DNA Mitochondria

solvent control (0.5% DMSO)

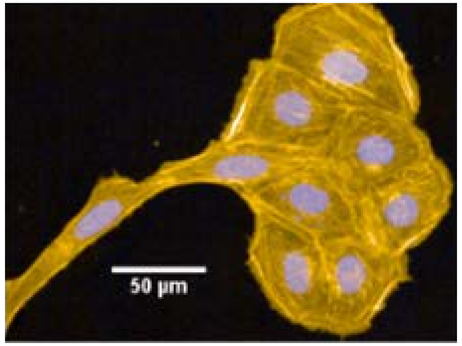


Berberine Chloride (10 µM)

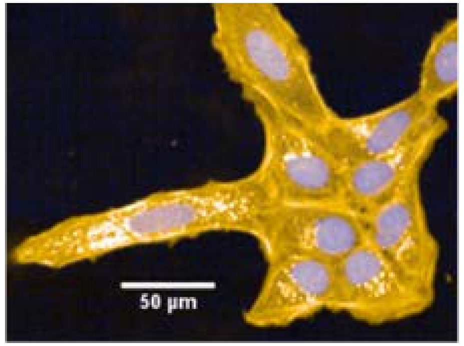


DNA AGP

solvent control (0.5% DMSO)

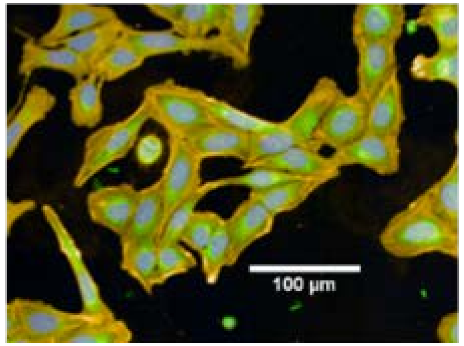


Ca-074-Me (1 µM)

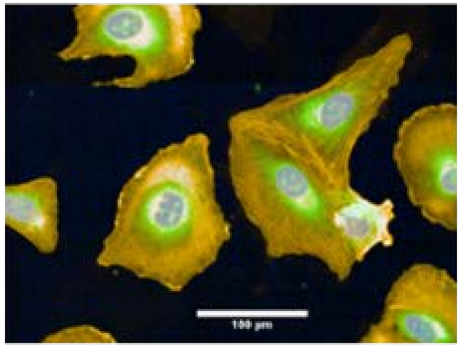


DNA RNA/ER AGP

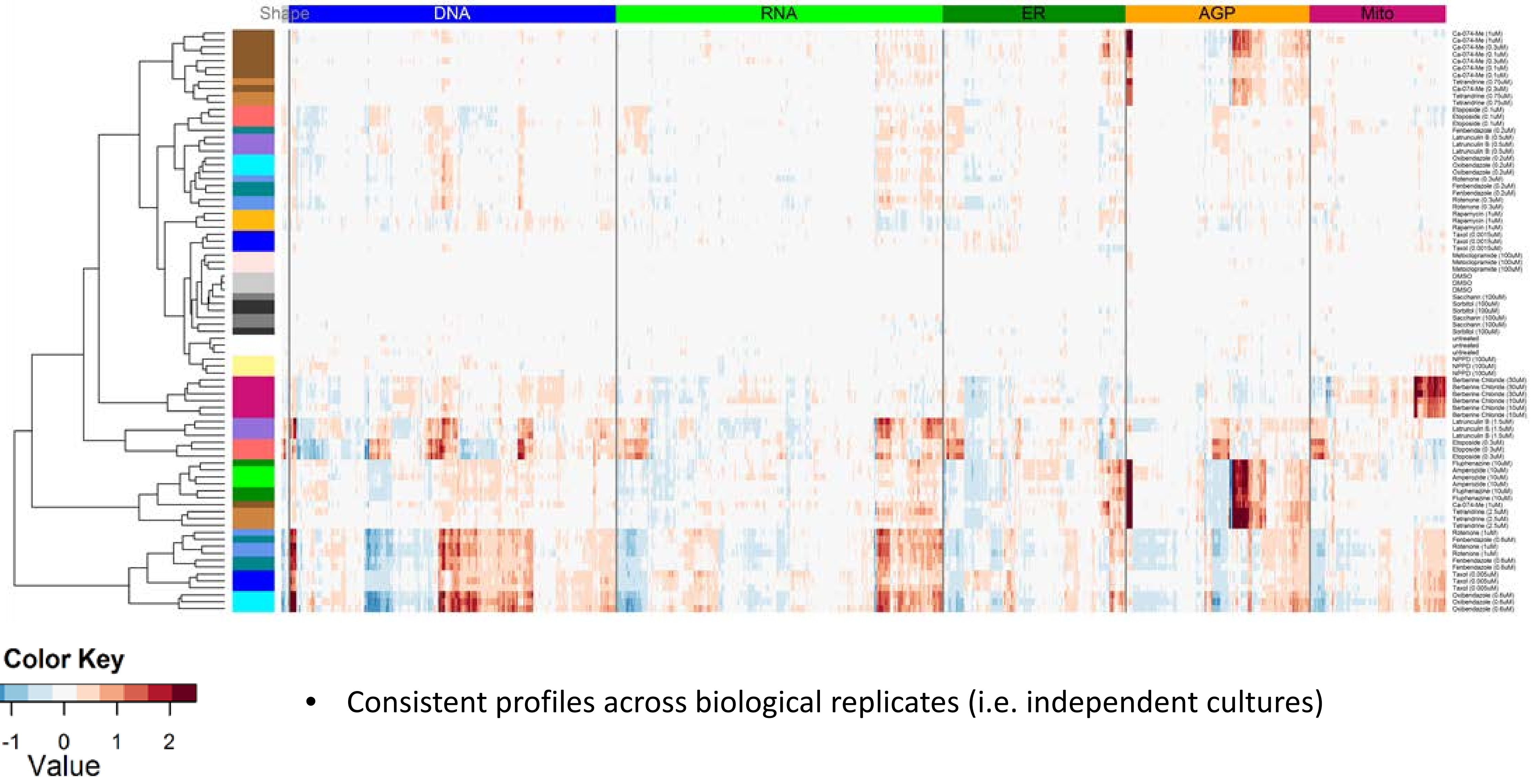
solvent control (0.5% DMSO)



Etoposide (1 µM)

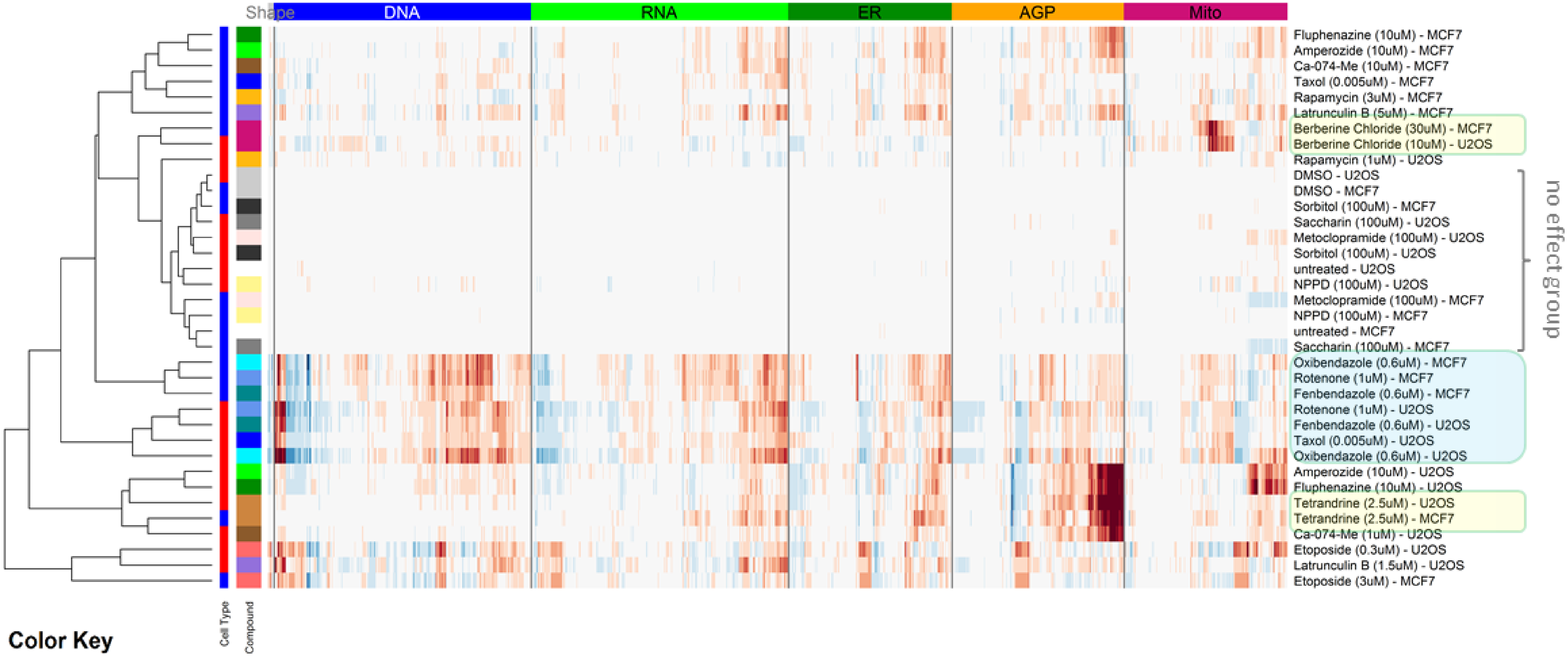


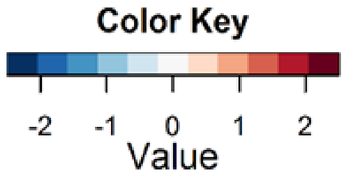
Reproducibility Across Biological Replicates



- Consistent profiles across biological replicates (i.e. independent cultures)

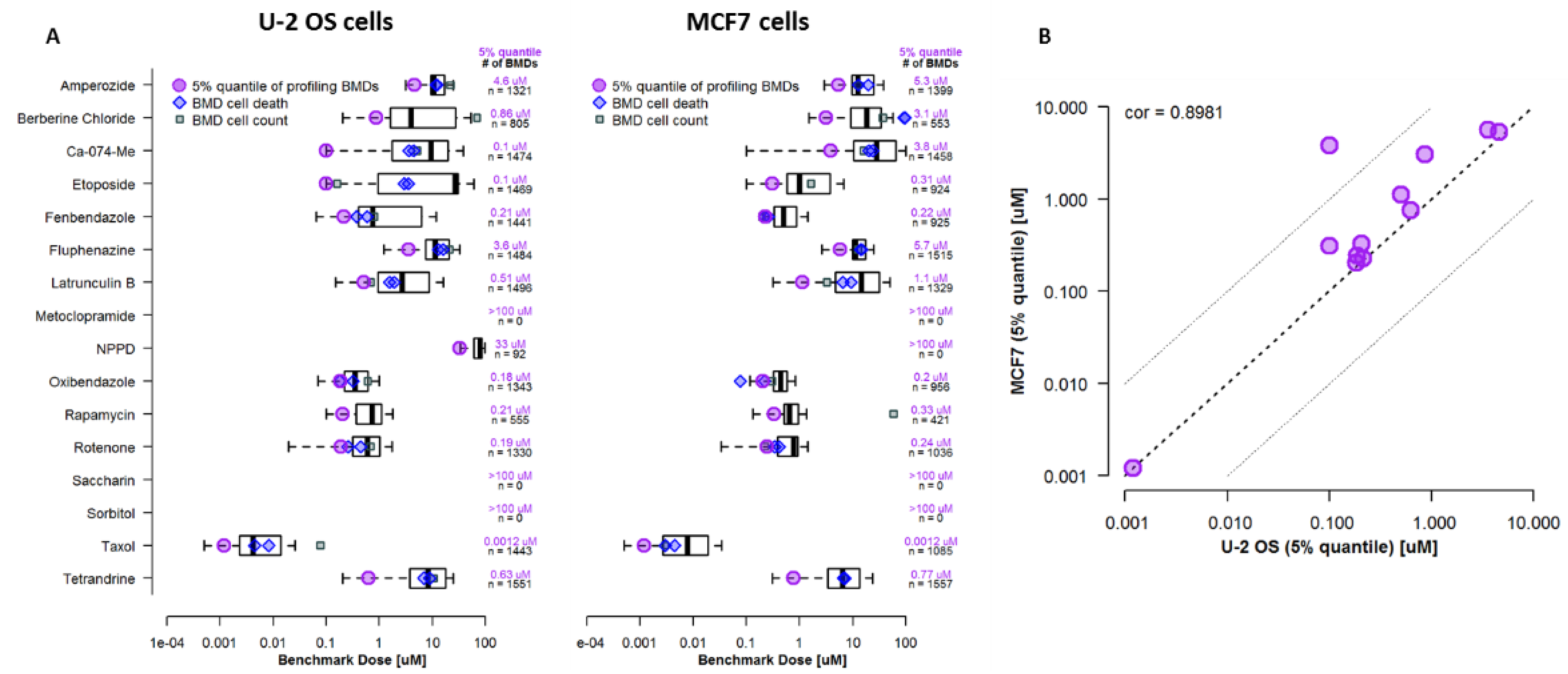
Chemical Profiles Across Two Cell Types





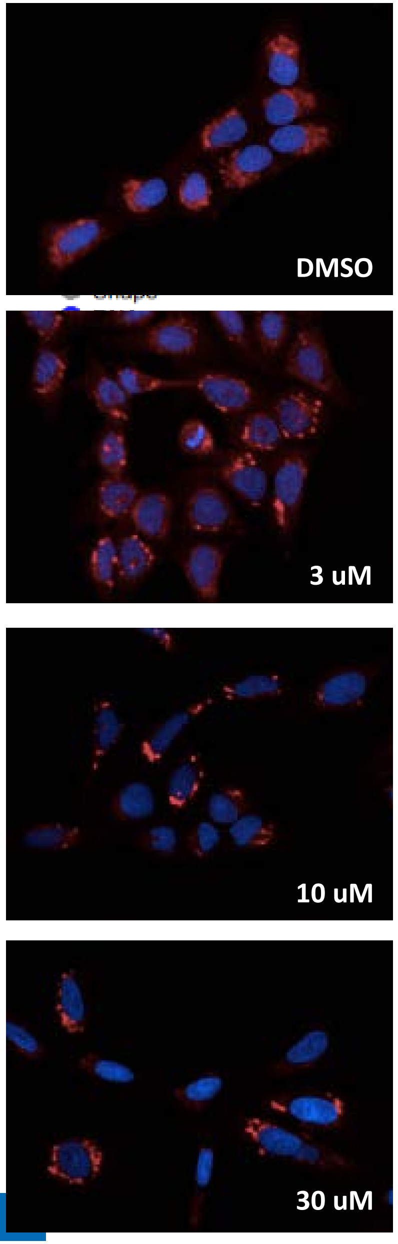
- Sub-cytotoxic concentrations of chemicals tend to produce similar profiles across cell types.

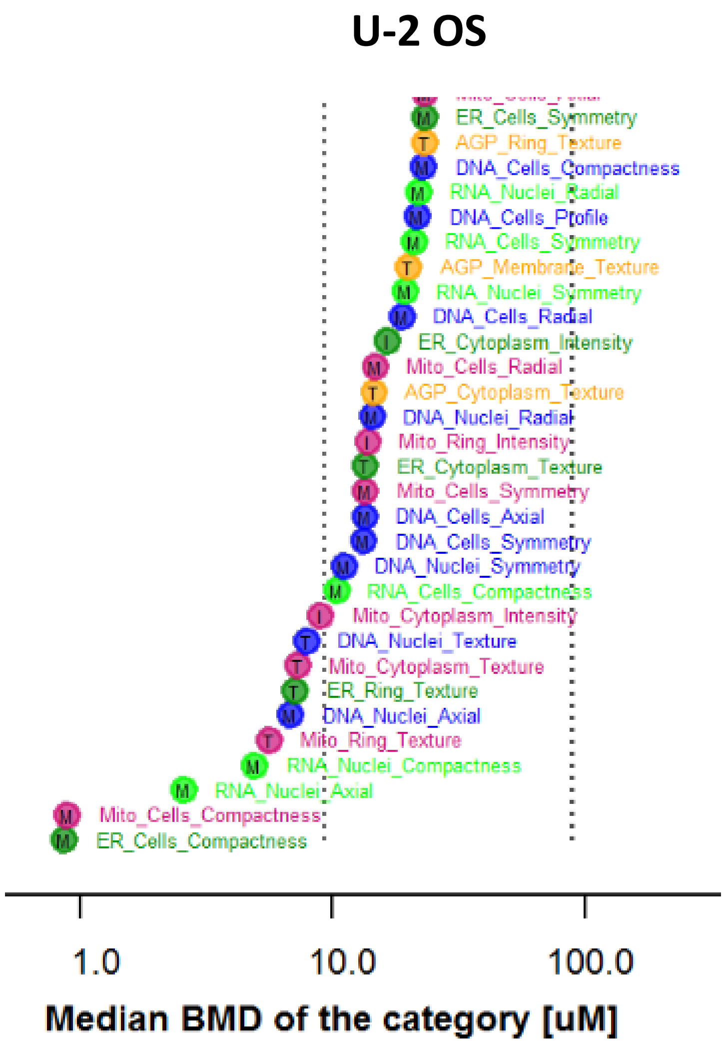
Determination of In Vitro Points of Departure (POD)

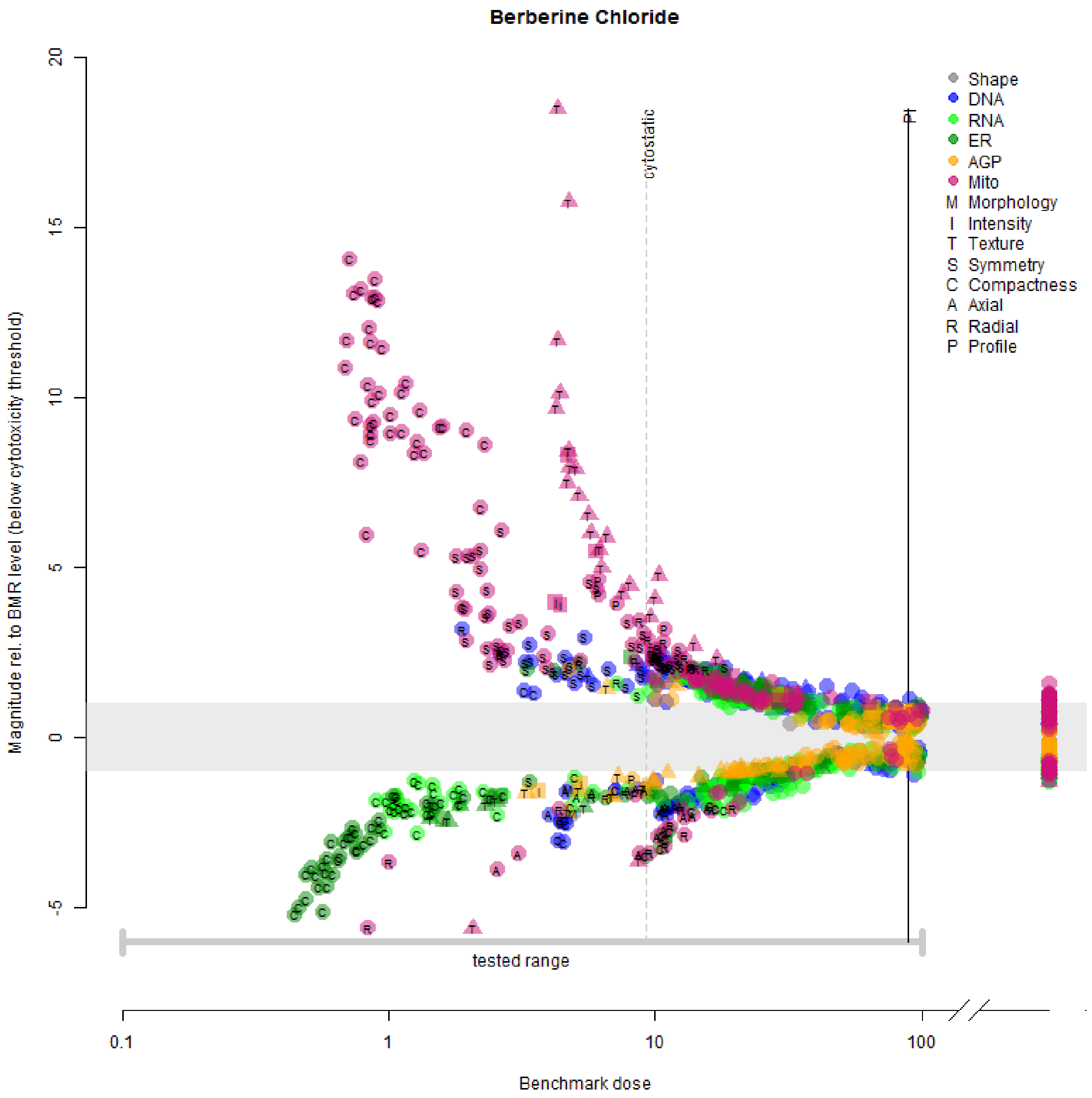


- Thresholds for effects on cell morphology are often below cytotoxicity thresholds, with separation of ≥1 order of magnitude in many cases.
- Potency estimates are largely consistent across cell types, with some exceptions.

Phenotype Ontogeny Analysis

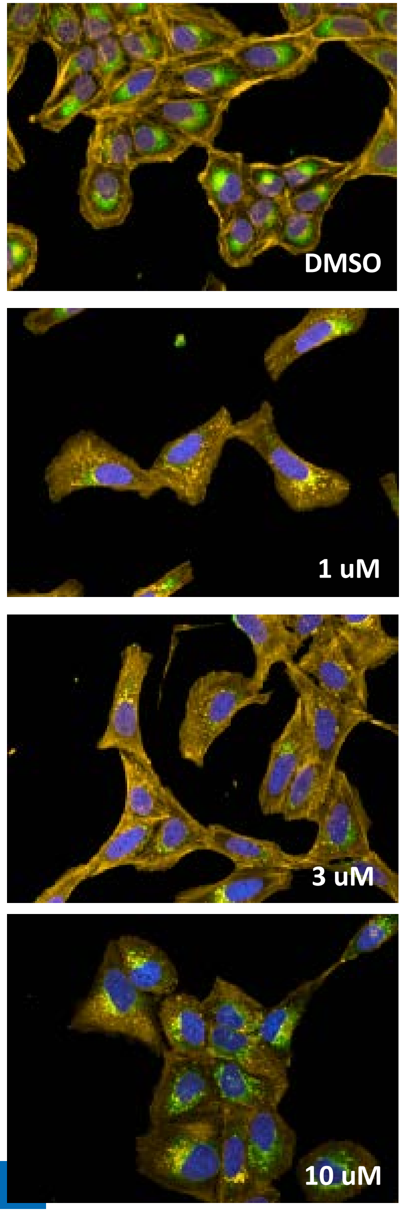


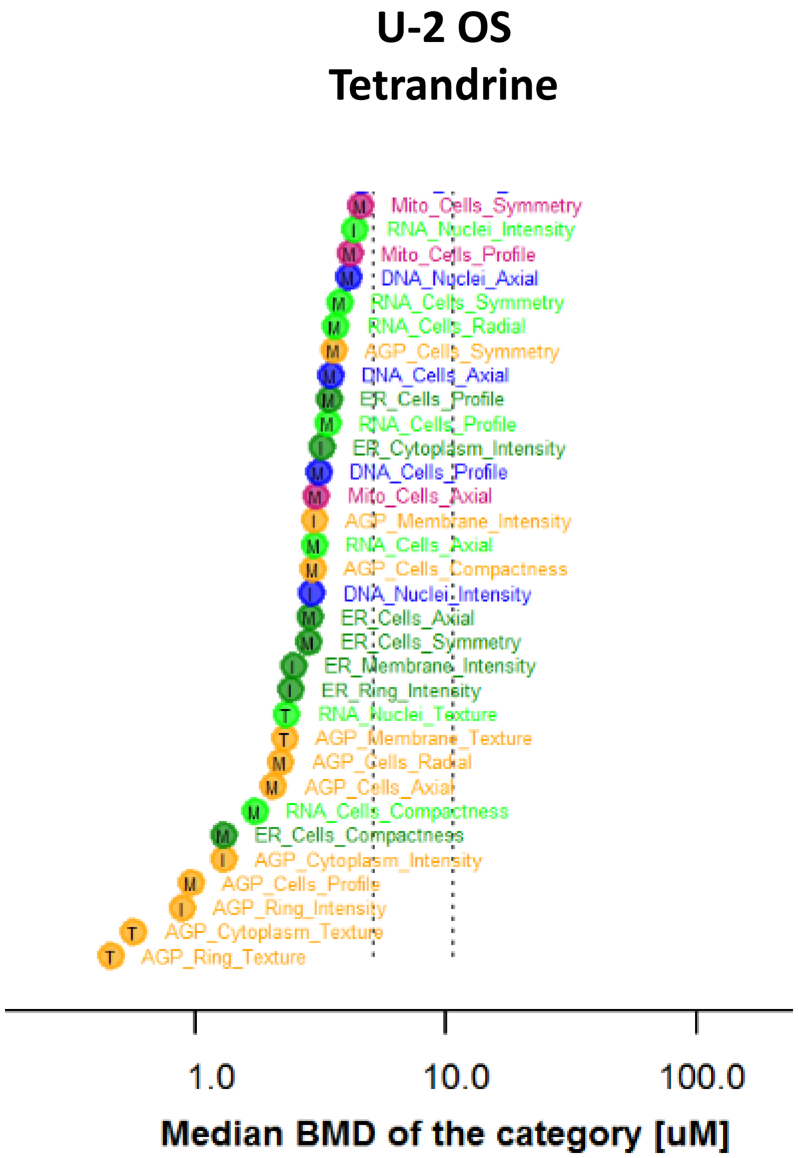


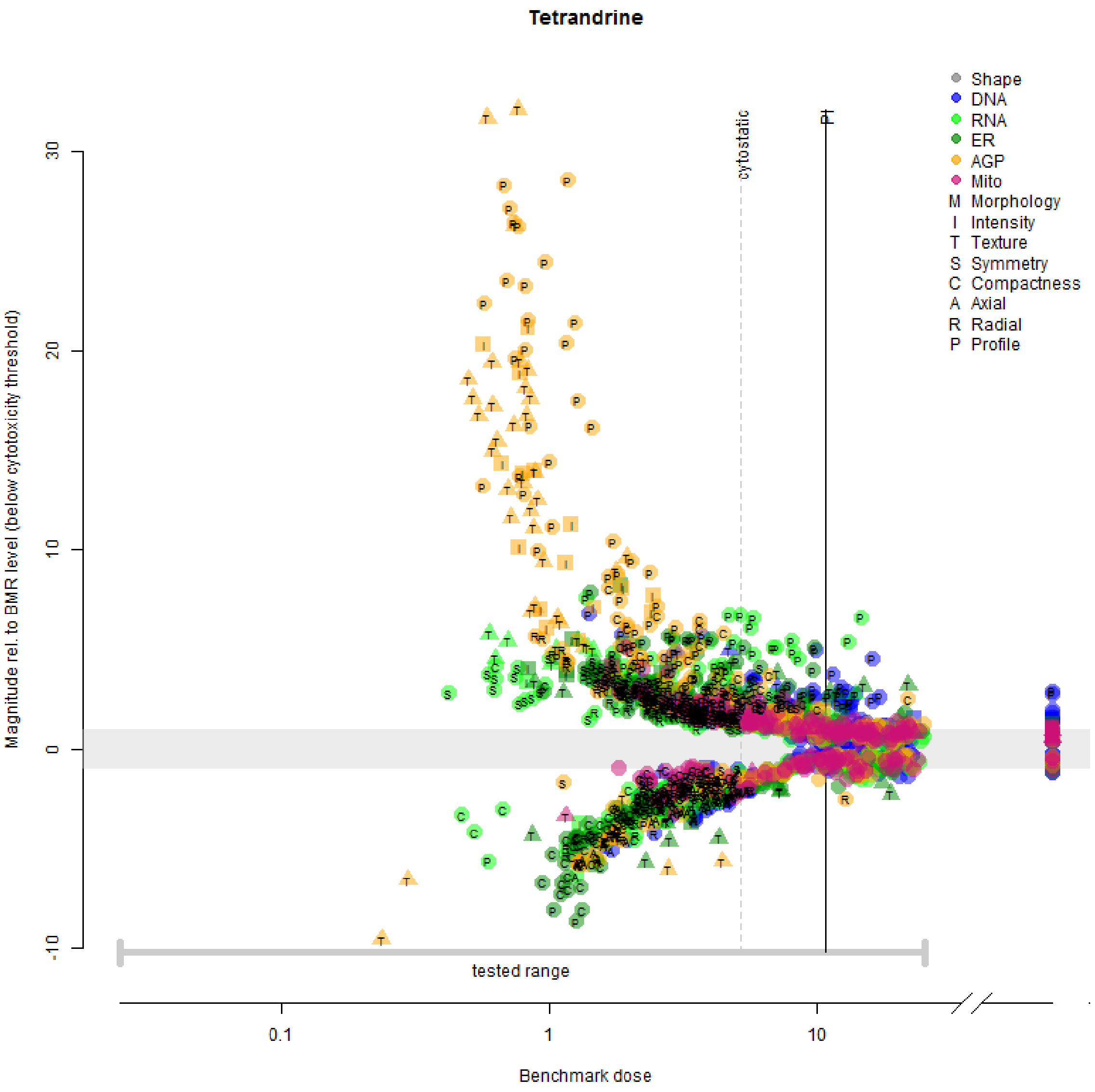


- Phenotype ontology analysis illustrates differences in chemical response profiles.

Phenotype Ontogeny Analysis







Acknowledgments

HTTr Project Team:
Imran Shah
Woody Setzer
Derik Haggard
Matt Martin*
Richard Judson
Rusty Thomas

NCCT Labs:
Clinton Willis
Danielle Suarez

Unilever:
Paul Carmichael
Andy White
Sophie Malcomber
Richard Stark*

National Toxicology Program:
Scott Auerbach

Sciome:
Jason Phillips

BioSpyder
Joel McComb
Bruce Seligmann
Jo Yeakley
Milos Babic
Pete Shepard
Kyle LeBlanc
Jason Downing

National Center for Computational Toxicology