United States Environmental Protection Agency

High-Throughput Transcriptomics (HTTr) in *In Vitro* Chemical Screening, A Sensitive Tool for Benchmark Dose Assessment

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

Office of Research and Development

American Society of Pharmacology and Experimental Therapeutics
Annual Meeting, San Diego, CA
April 24th, 2018

United States Environmental Protection Agency

Disclaimer

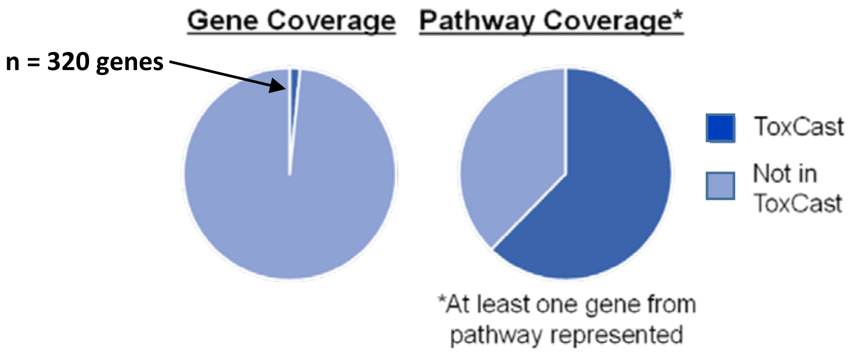
The views expressed in this [article/presentation/poster] are those of the author(s) and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency, nor does mention of trade names or products represent endorsement for use.

Outline

- **Background**
  - Why high-throughput transcriptomics (HTTr)?
  - TempO-Seq Technology Overveiw
- **Experimental Design & Workflows**
  - Screening format
  - Treatment Randomization
  - Quality Control Samples
- **Technical Reproducibility**
- **Applications**
  - Concentration-Response Modeling
  - Bioactivity Thresholds
  - Comparison of Structurally Related Chemicals
  - Gene Set Analysis
  - Mechanism of Action (MoA) Prediction

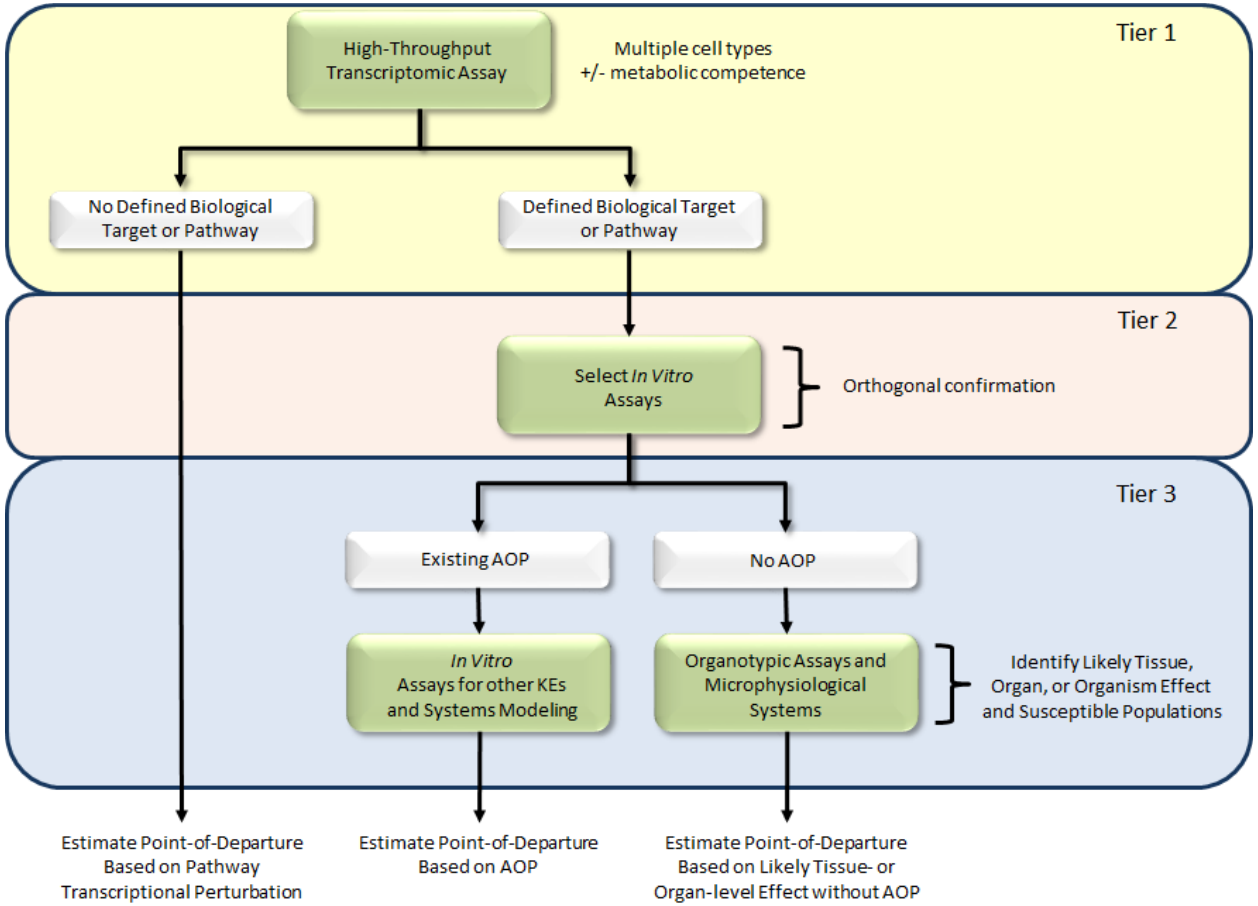
Background

- 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

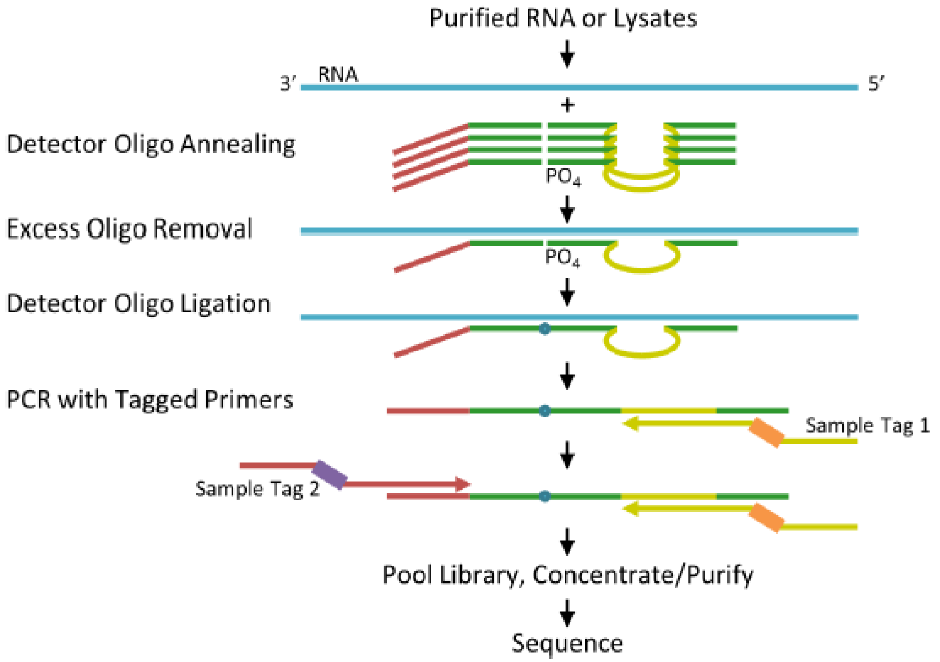


Background

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



United States Environmental Protection Agency

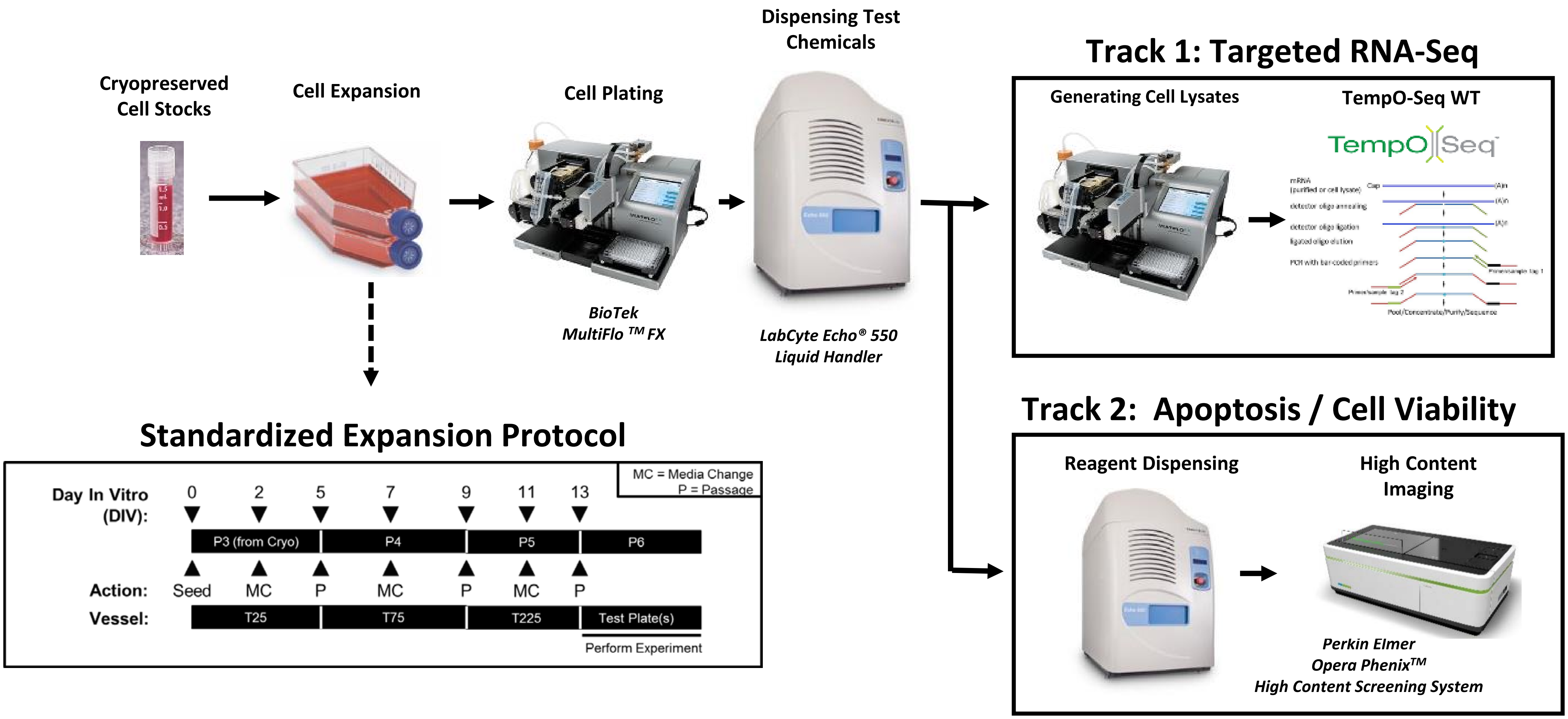
HTTr MCF Screen: Experimental Design

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/).

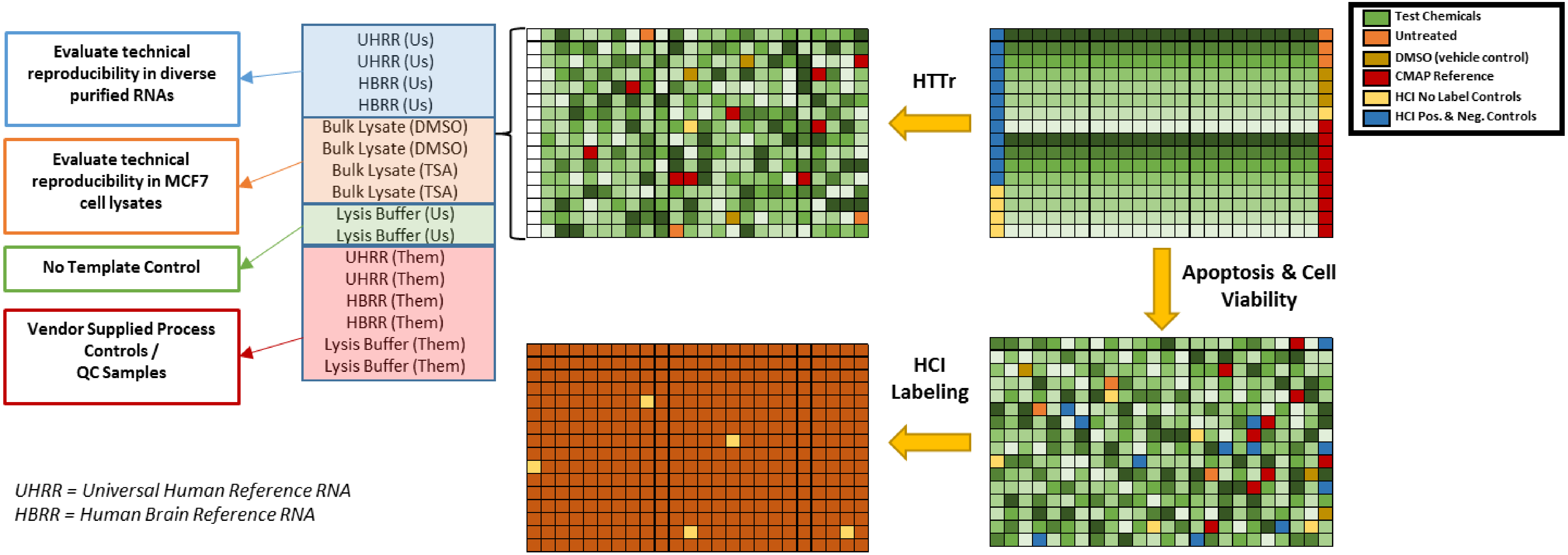
Experimental Workflow



Treatment Randomization & Quality Control Samples

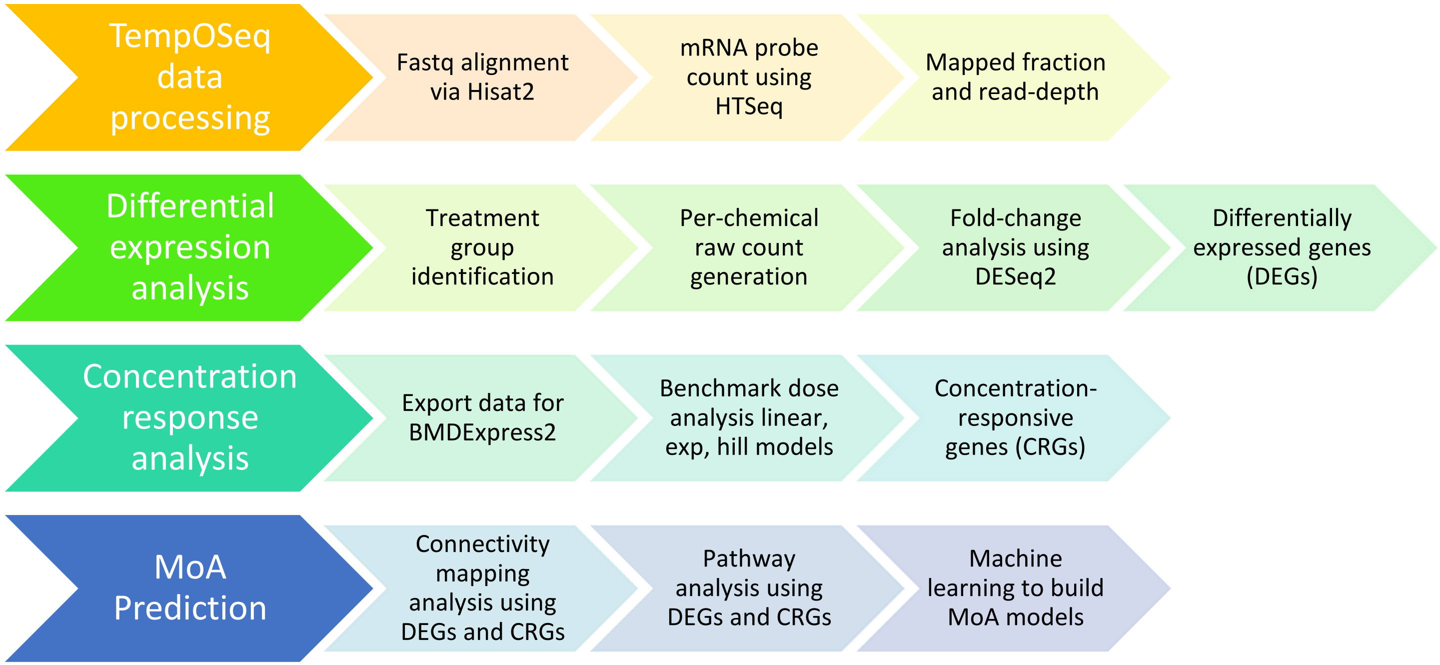
Treatment Randomization: *Each test plate uniquely randomized with respect to treatment.*

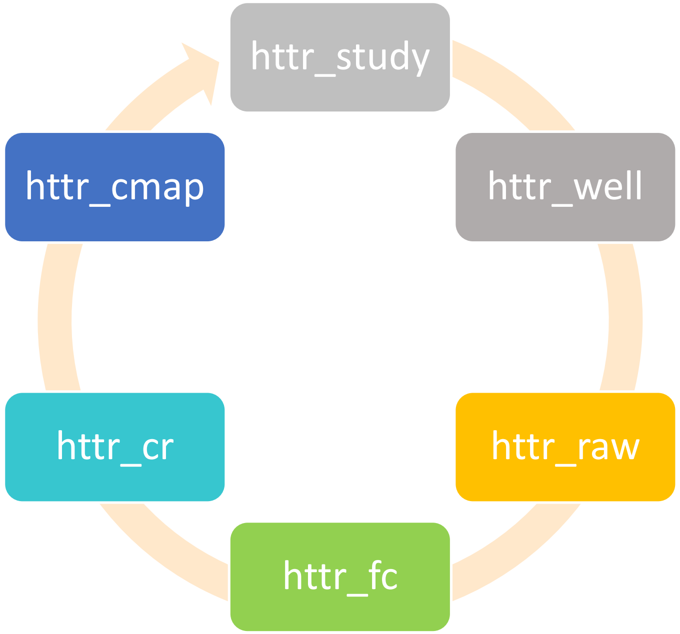
QC Samples: *Quality Control samples included on each plate*



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

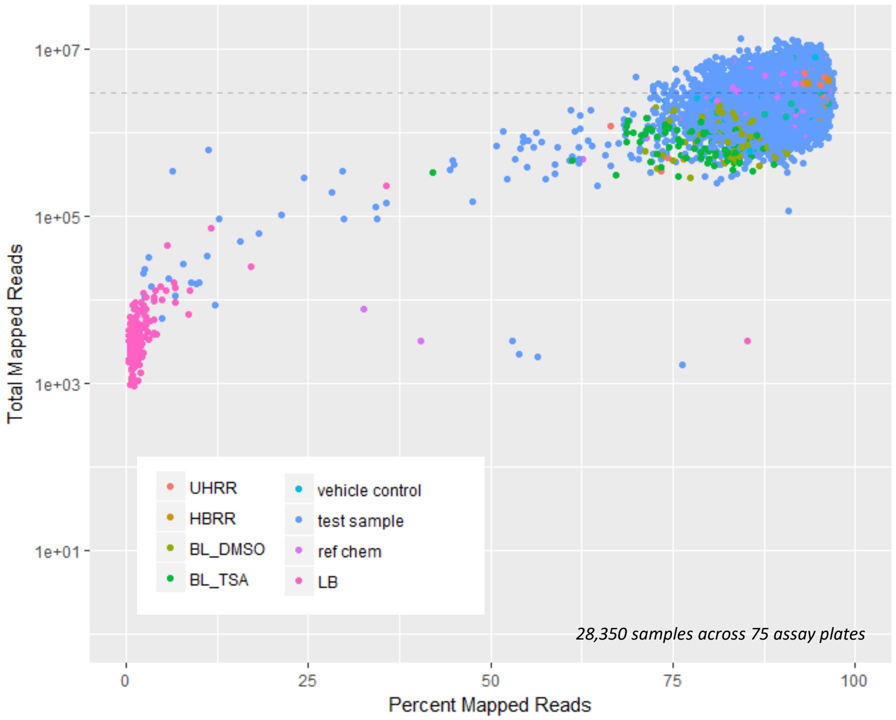
MongoDB

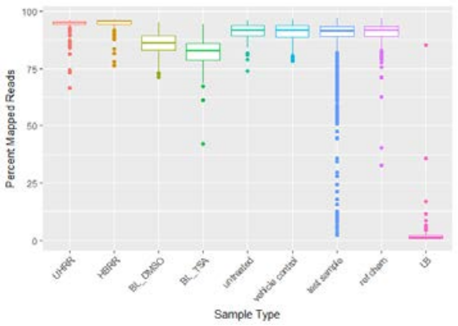
mongodb://pb.epa.gov/httr_v1
mongodb://pb.epa.gov/cmap_v2

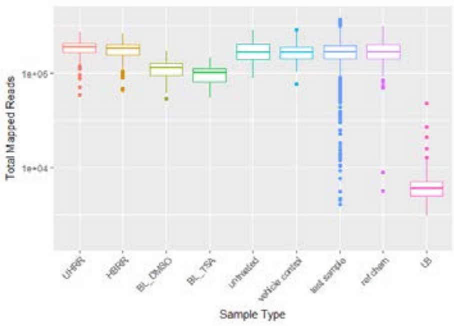
NCCT Investigators
NCCT IT

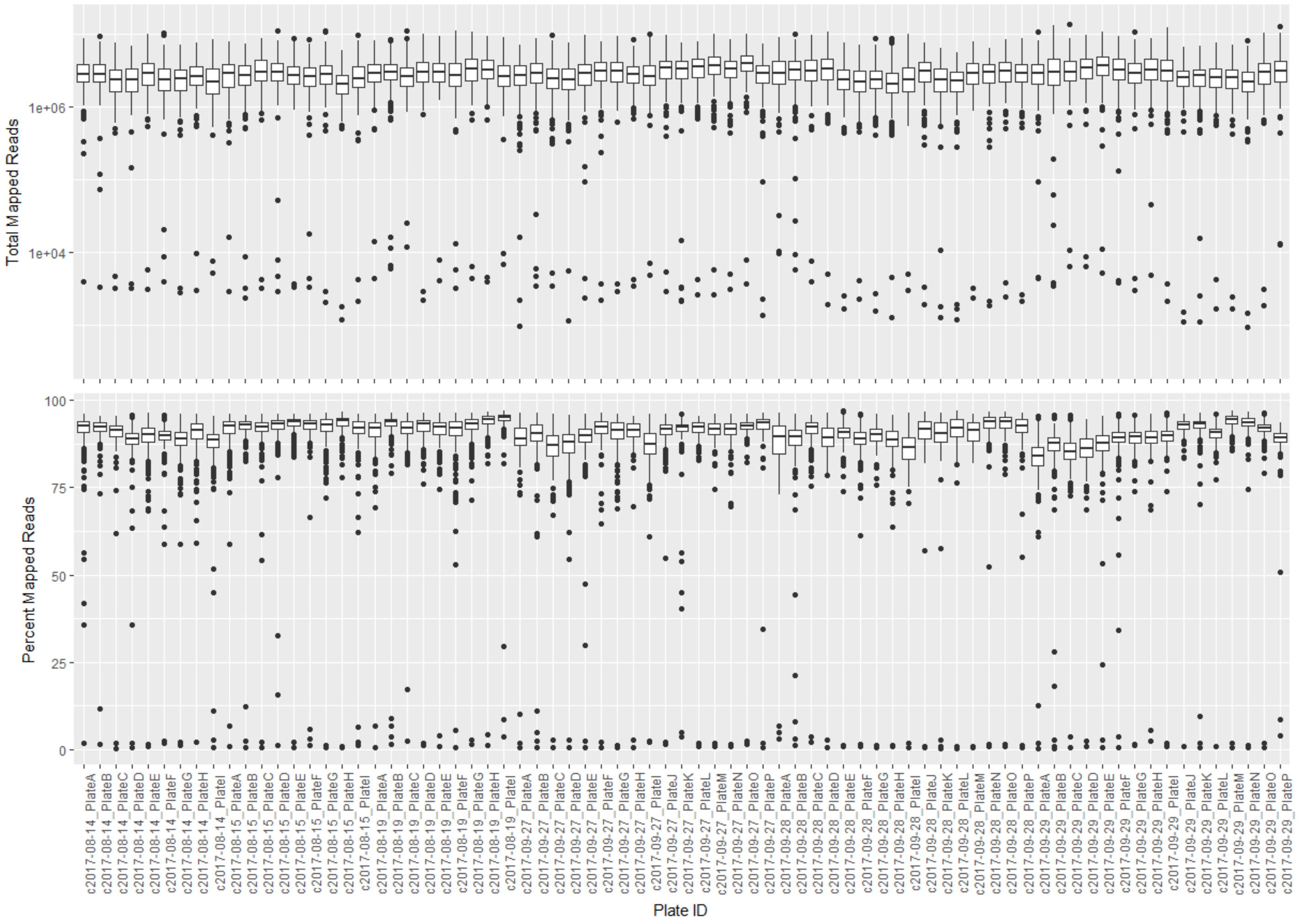
Technical Reproducibility

Reproducibility of Read Depths and Mapping Rates

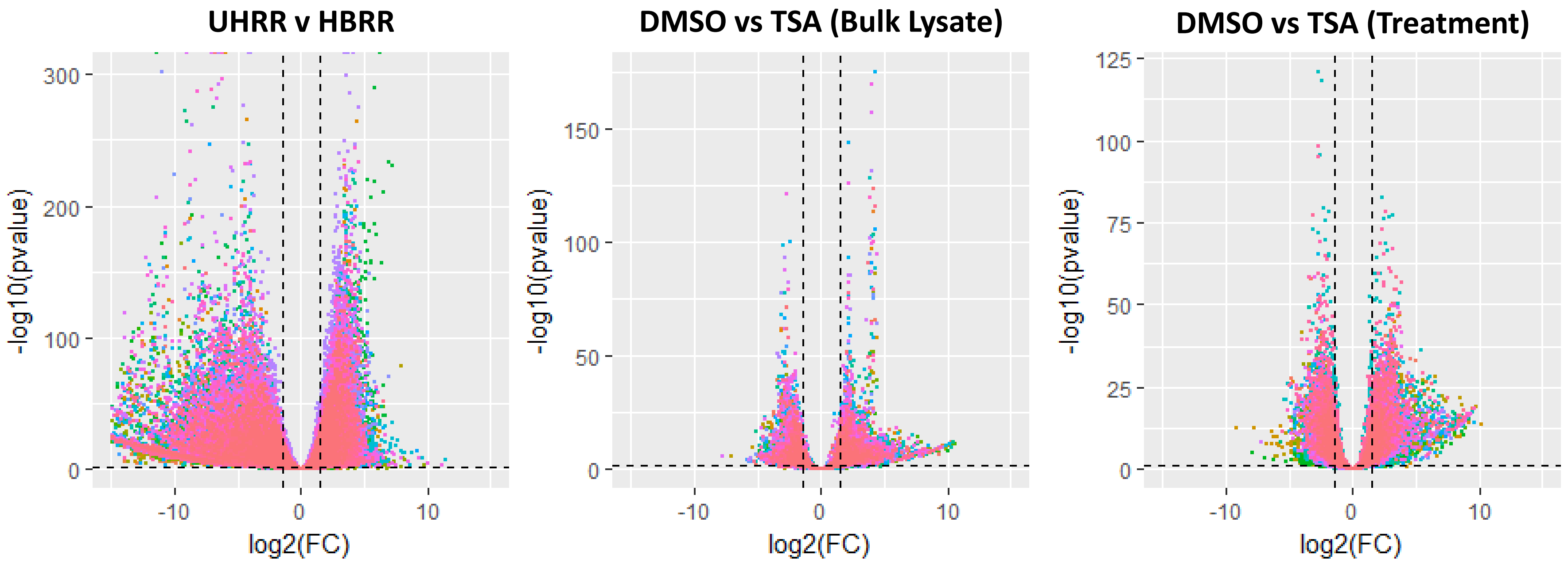


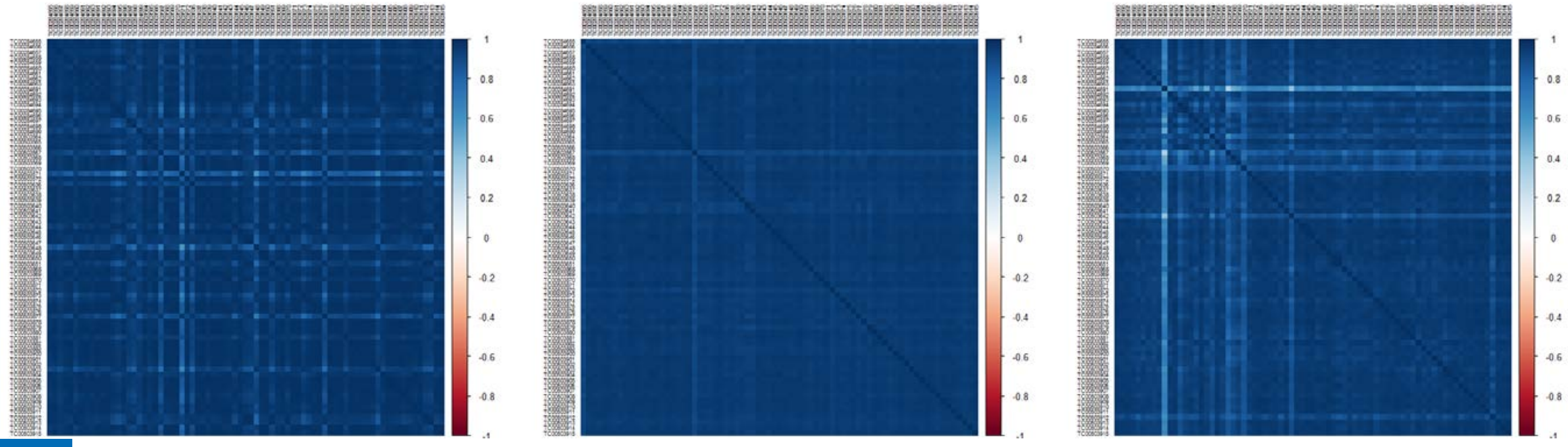


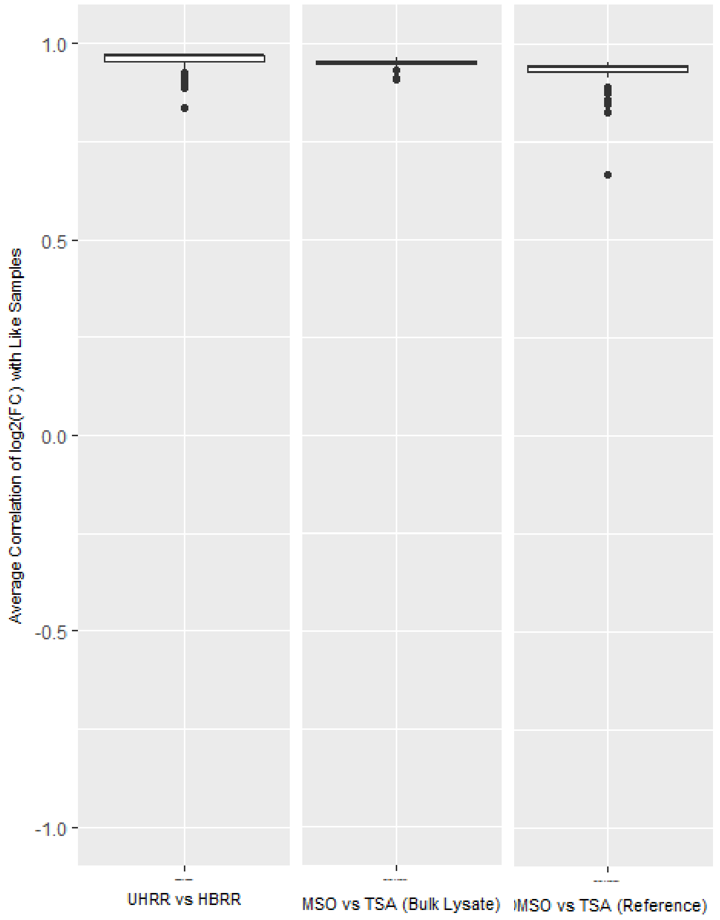




Reproducibility of Log_2(FC) Estimates







Concentration-Response Modeling / Bioactivity Thresholds

EPA United States Environmental Protection Agency

Benchmark Dose Modeling

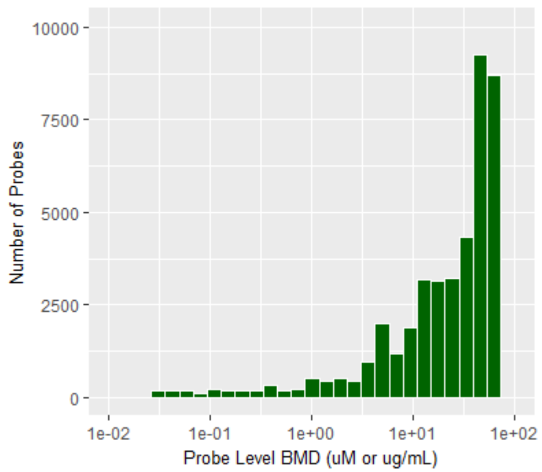
Parameter	Criteria [a]
Pre-filter:	ANOVA ($p_{raw} < 0.05$ & $\lvert FC \rvert \geq 2$)
Models	Hill, Exponential 2, *poly2, power, linear*
BMR Factor:	1.349 (10 %)
Best Model Selection:	Lowest AIC
Hill Model Flagging [b]:	'k' < 1/3 Lowest Positive Dose Retain Flagged Models
Pathway Analysis:	Genes with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

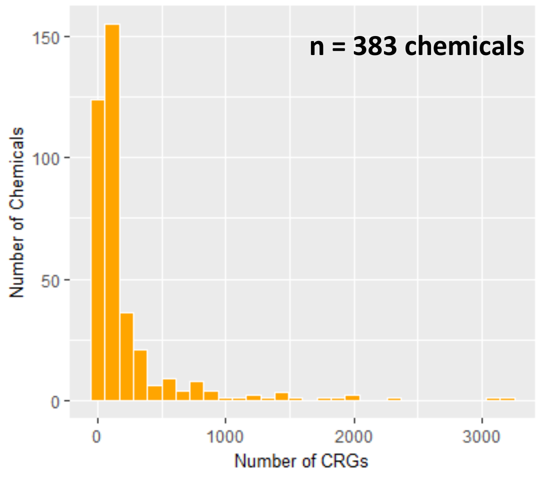
[a] *Exploratory analysis – modeling criteria not finalized*

[c] Gene Set Collections:

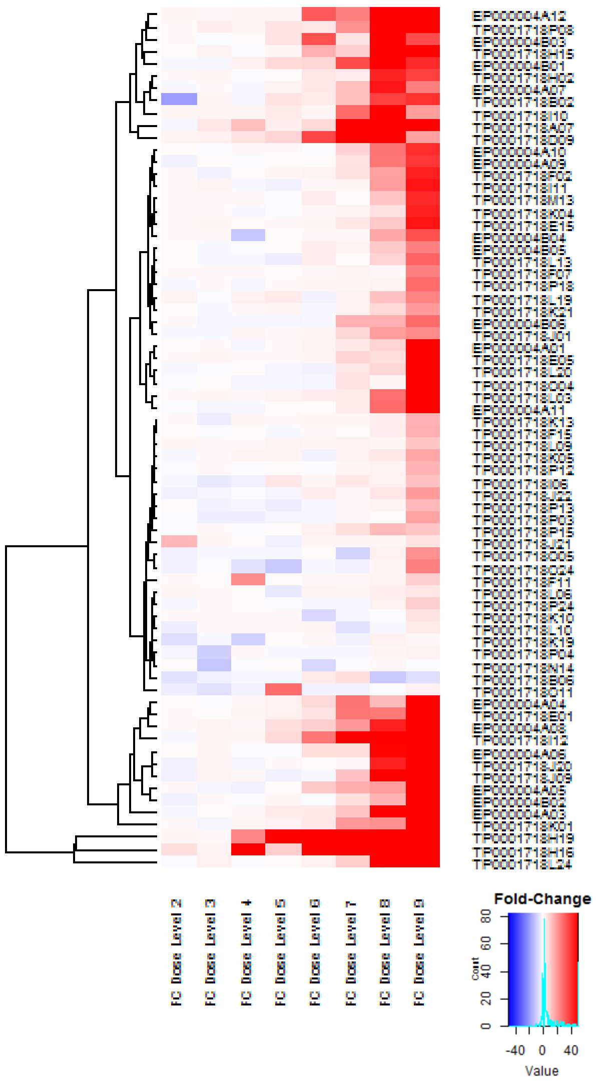
- **MSigDB_C2:** Curated gene sets from online pathway databases, publications and knowledge of domain experts (n = 4738).
- **MSigDB_H:** Coherently expressed signatures derived by aggregating many MSigDB gene sets to represent well-defined biological states or processes (n = 50).
- **Reactome:** Open-source, 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

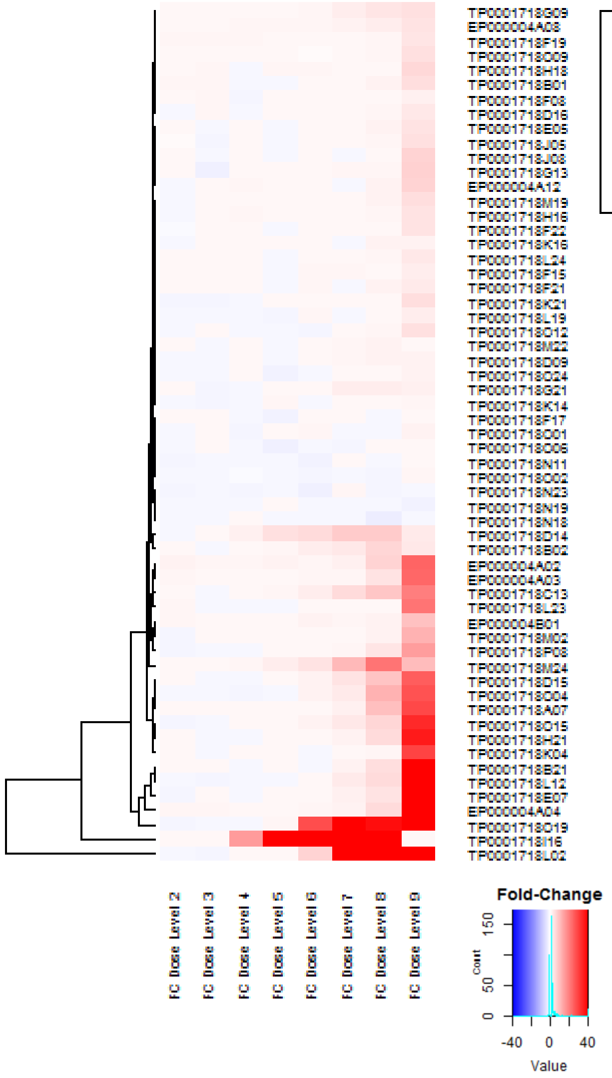




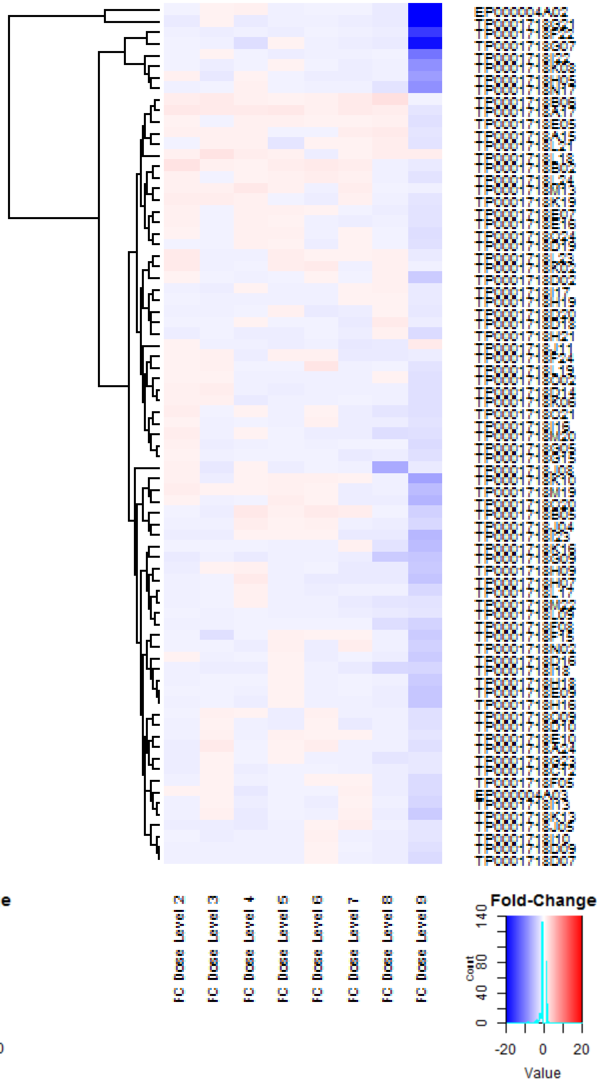
CYP1A1_10775



HMOX1_3041

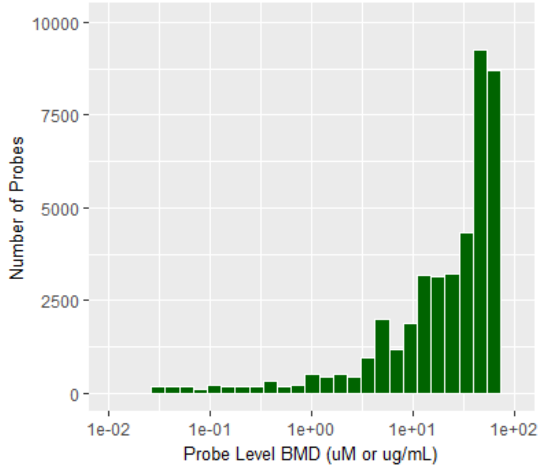


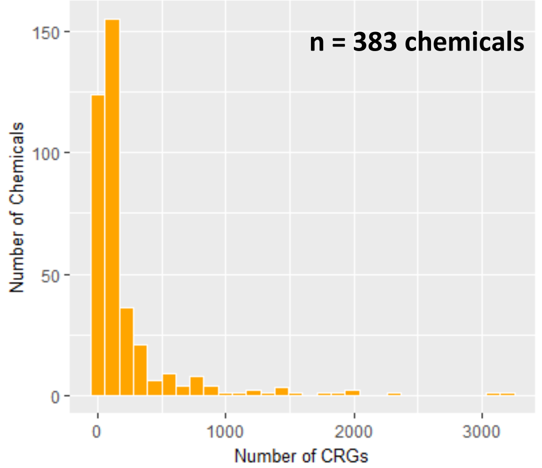
EIF4E2_16592



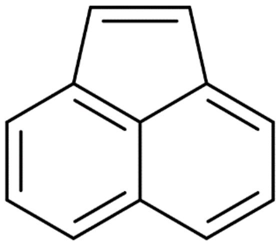
United States Environmental Protection Agency

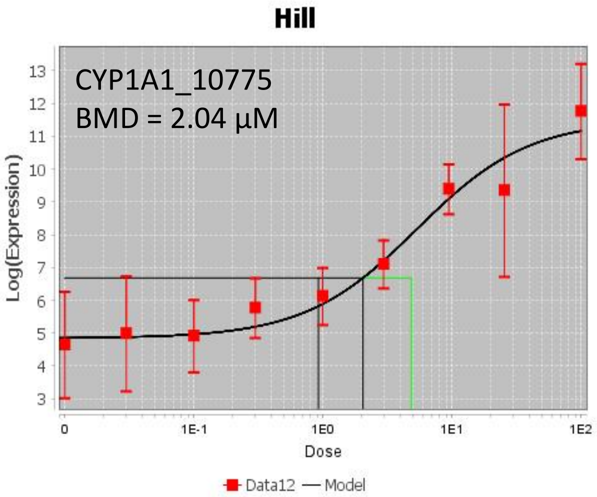
Benchmark Dose Modeling Summary



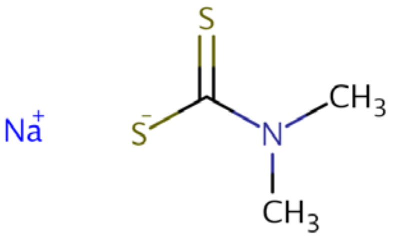


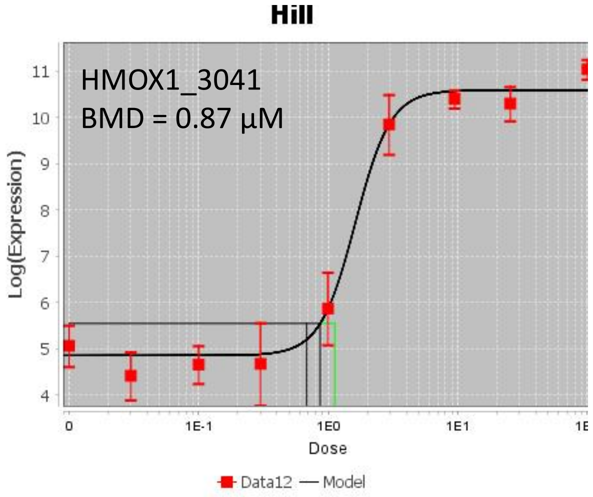
Acenaphthylene
208-96-8 | DTXSID3023845



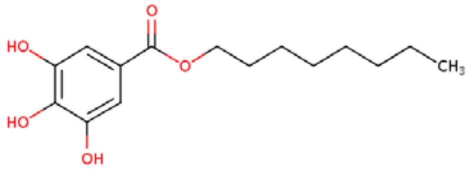


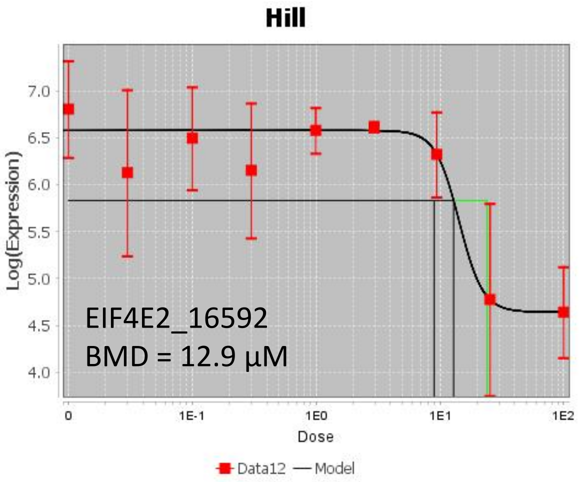
Sodium dimethyldithiocarbamate
128-04-1 | DTXSID6027050



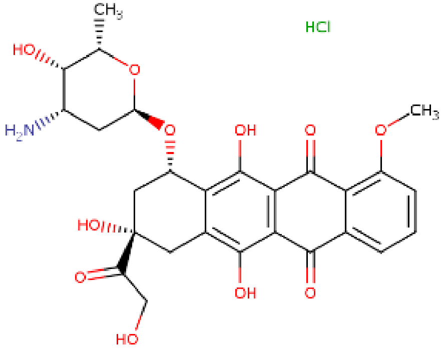


Octyl gallate
1034-01-1 | DTXSID4040713

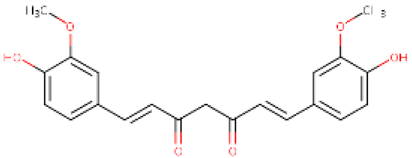




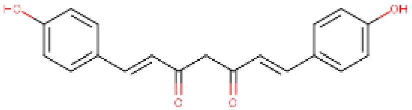
Case Study Chemicals



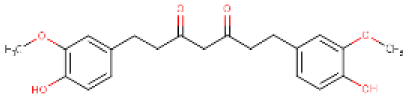
Adriamycin hydrochloride | DTXSID3030636



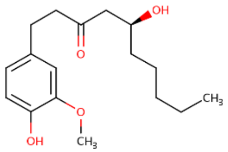
Curcumin | DTXSID8031077



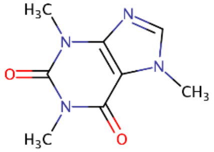
Bisdemethoxycurcumin | DTXSID00872663



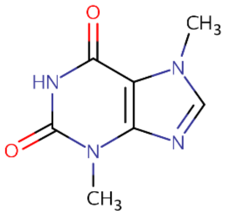
Tetrahydrocurcumin | DTXSID30865801



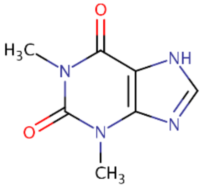
(6)-Gingerol | DTXSID3041035



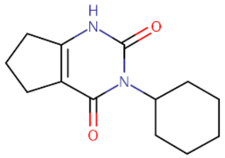
Caffeine | DTXSID0020232



Theobromine | DTXSID9026132

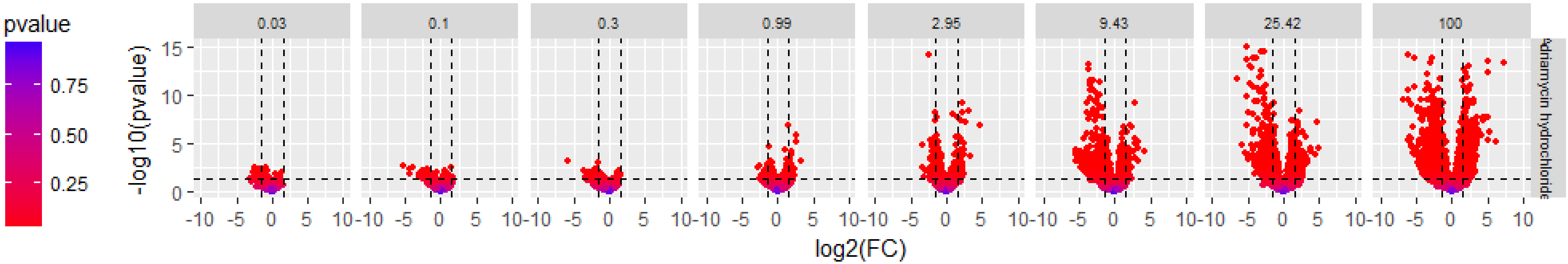


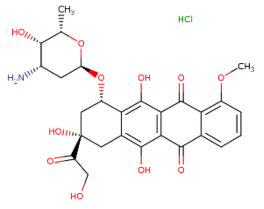
Theophylline | DTXSID5021336



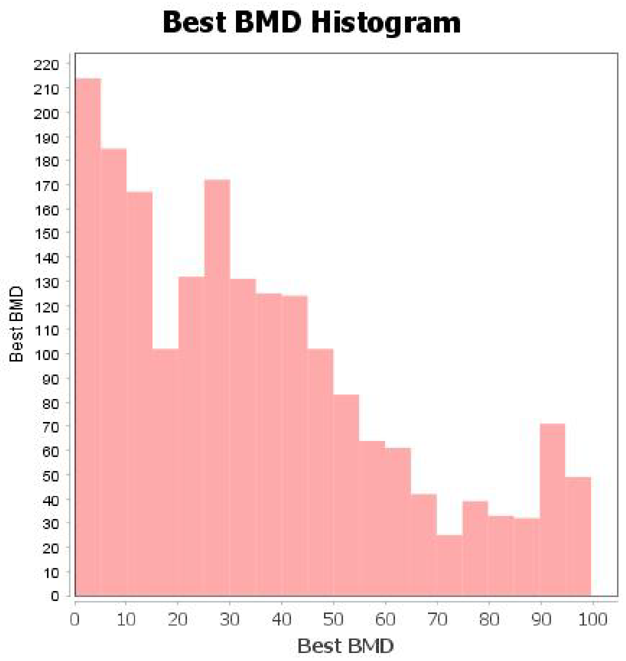
Lenacil | DTXSID9042093

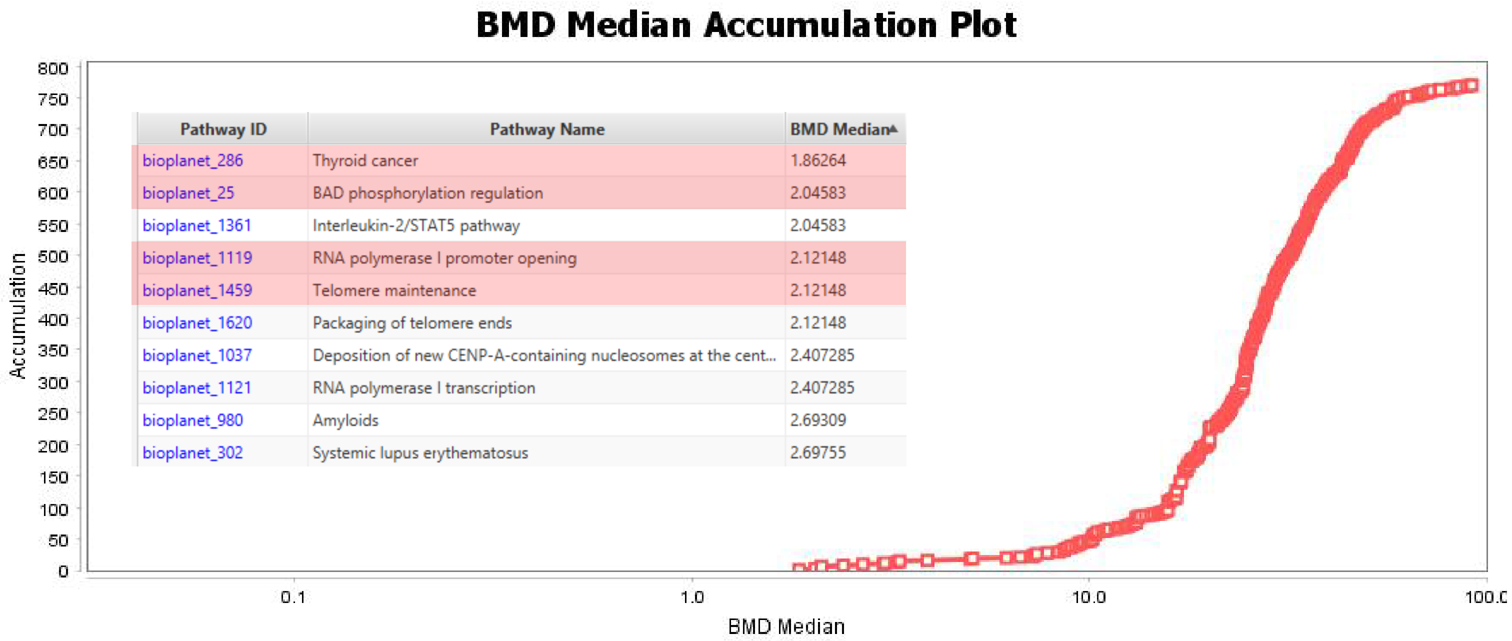
Doxorubicin (aka Adriamycin hydrochloride)





Adriamycin hydrochloride | DTXSID3030636

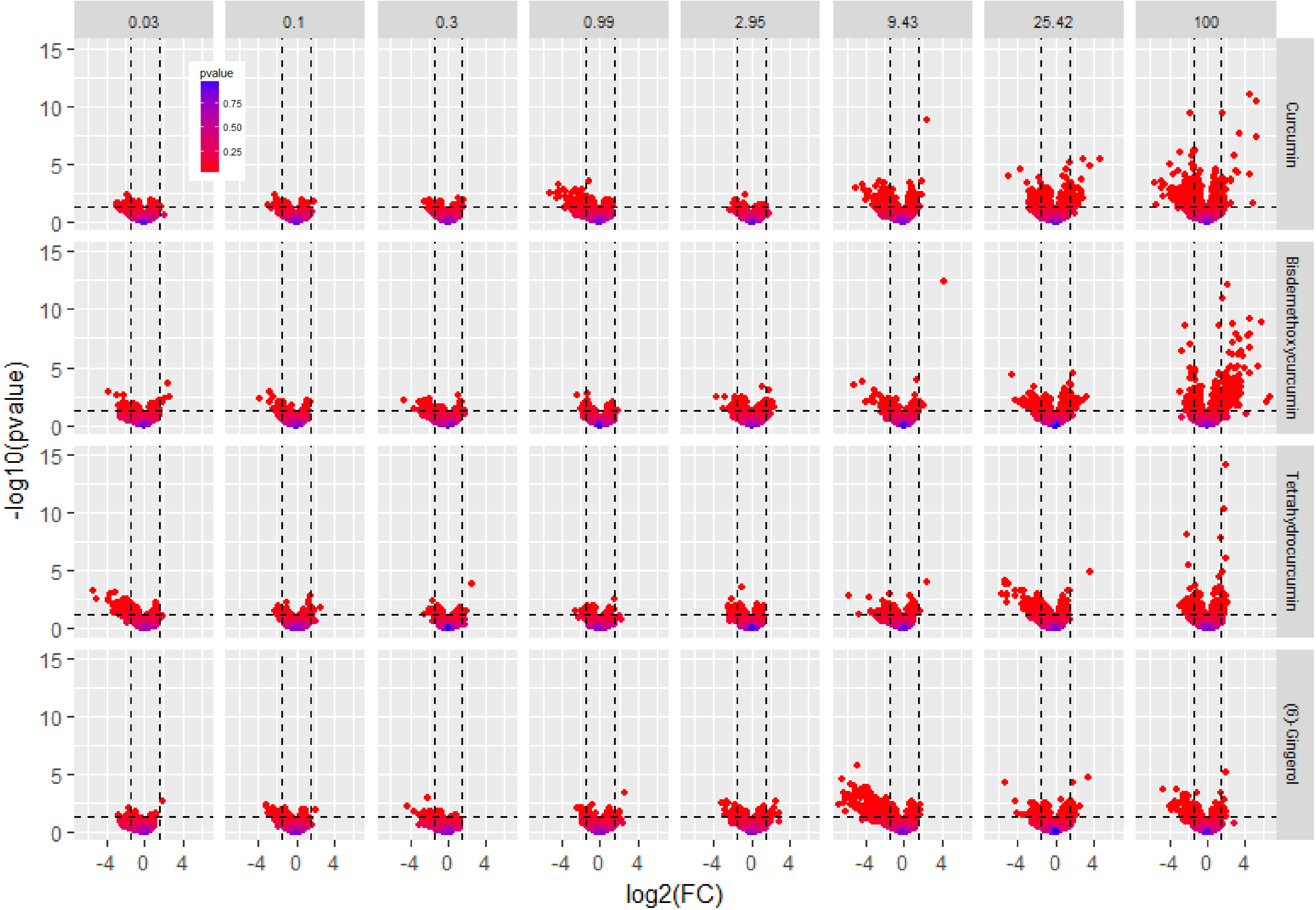


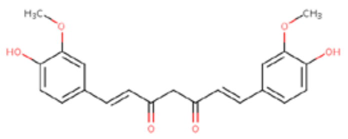


Pathway ID	Pathway Name	BMD Median
bioplanet_286	Thyroid cancer	1.86264
bioplanet_25	BAD phosphorylation regulation	2.04583
bioplanet_1361	Interleukin-2/STAT5 pathway	2.04583
bioplanet_1119	RNA polymerase I promoter opening	2.12148
bioplanet_1459	Telomere maintenance	2.12148
bioplanet_1620	Packaging 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

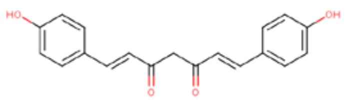
Int J Mol Cell Med. 2015 Spring;4(2):94-102.
Elevation of cAMP Levels Inhibits Doxorubicin-Induced Apoptosis in Pre- B ALL NALM- 6 Cells Through Induction of BAD Phosphorylation and Inhibition of P53 Accumulation.
Fatemi A[1], Kazemi A[1], Kashiri M[1], Safa M[2].

Evaluating Structurally Related Chemicals

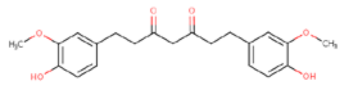




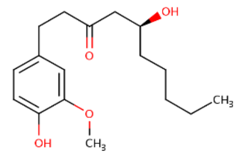
Curcumin | DTXSID8031077



Bisdemethoxycurcumin | DTXSID00872663

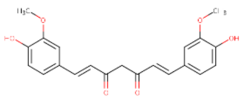


Tetrahydrocurcumin | DTXSID30865801

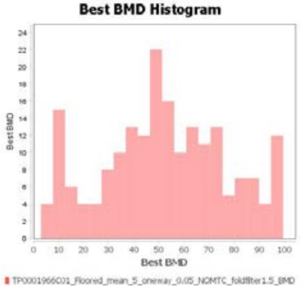


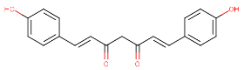
(6)-Gingerol | DTXSID3041035

Gene Set Analysis Using BMD Modeling Results

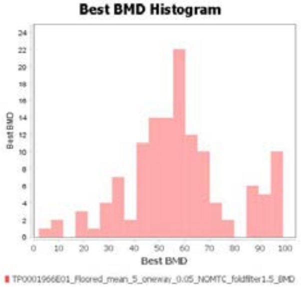


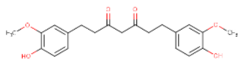
Curcumin | DTXSID8031077



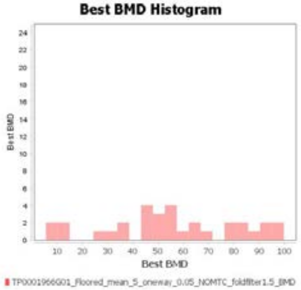


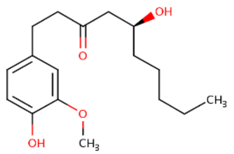
Bisdemethoxycurcumin | DTXSID00872663



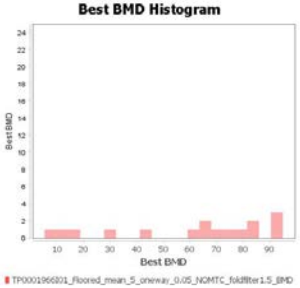


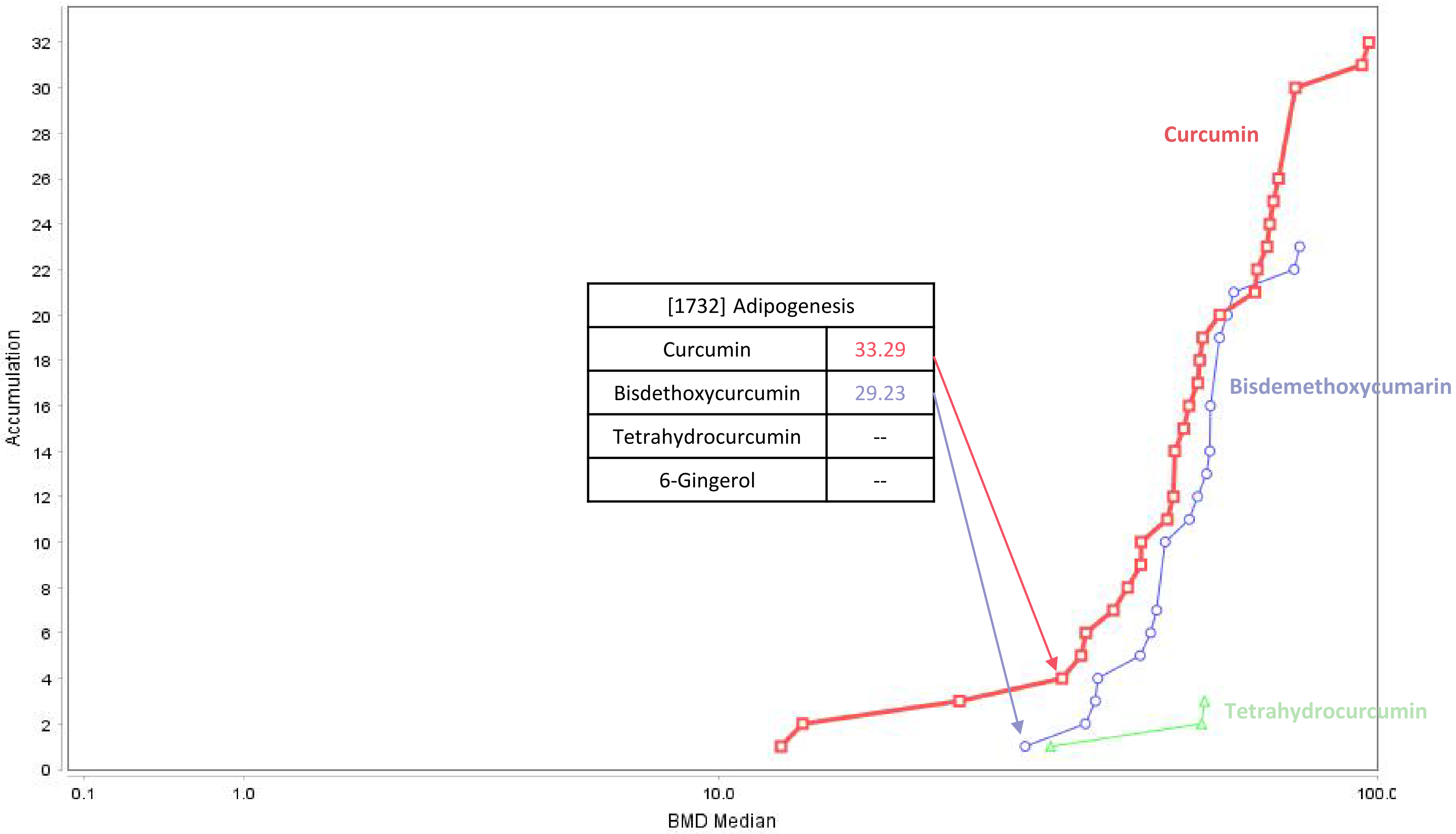
Tetrahydrocurcumin | DTXSID30865801





(6)-Gingerol | DTXSID3041035





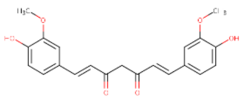
[1732] Adipogenesis	
Curcumin	33.29
Bisdethoxycurcumin	29.23
Tetrahydrocurcumin	--
6-Gingerol	--

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].

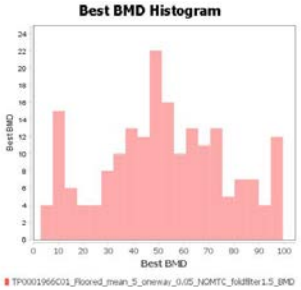
J Agric Food Chem. 2016 Feb 3;64(4):821-30. doi: 10.1021/acs.jafc.5b05577. Epub 2016 Jan 25.
Bisdemethoxycurcumin Inhibits Adipogenesis in 3T3-L1 Preadipocytes and Suppresses Obesity in High-Fat Diet-Fed C57BL/6 Mice.
Lai CS[1,2], Chen YY[1], Lee PS[1], Kalyanam N[3], Ho CT[4], Liou WS[5], Yu RC[1],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

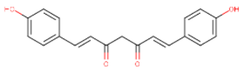
EPA United States Environmental Protection Agency

Gene Set Analysis Using BMD Modeling Results

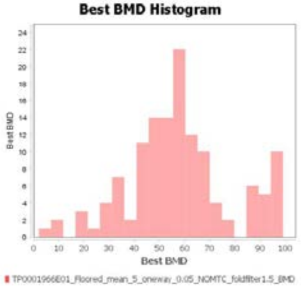


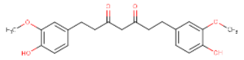
Curcumin | DTXSID8031077



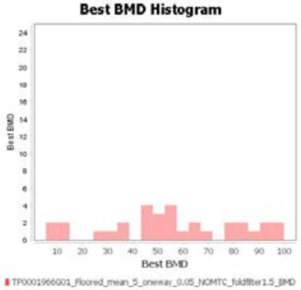


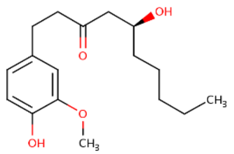
Bisdemethoxycurcumin | DTXSID00872663



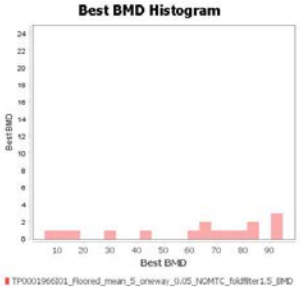


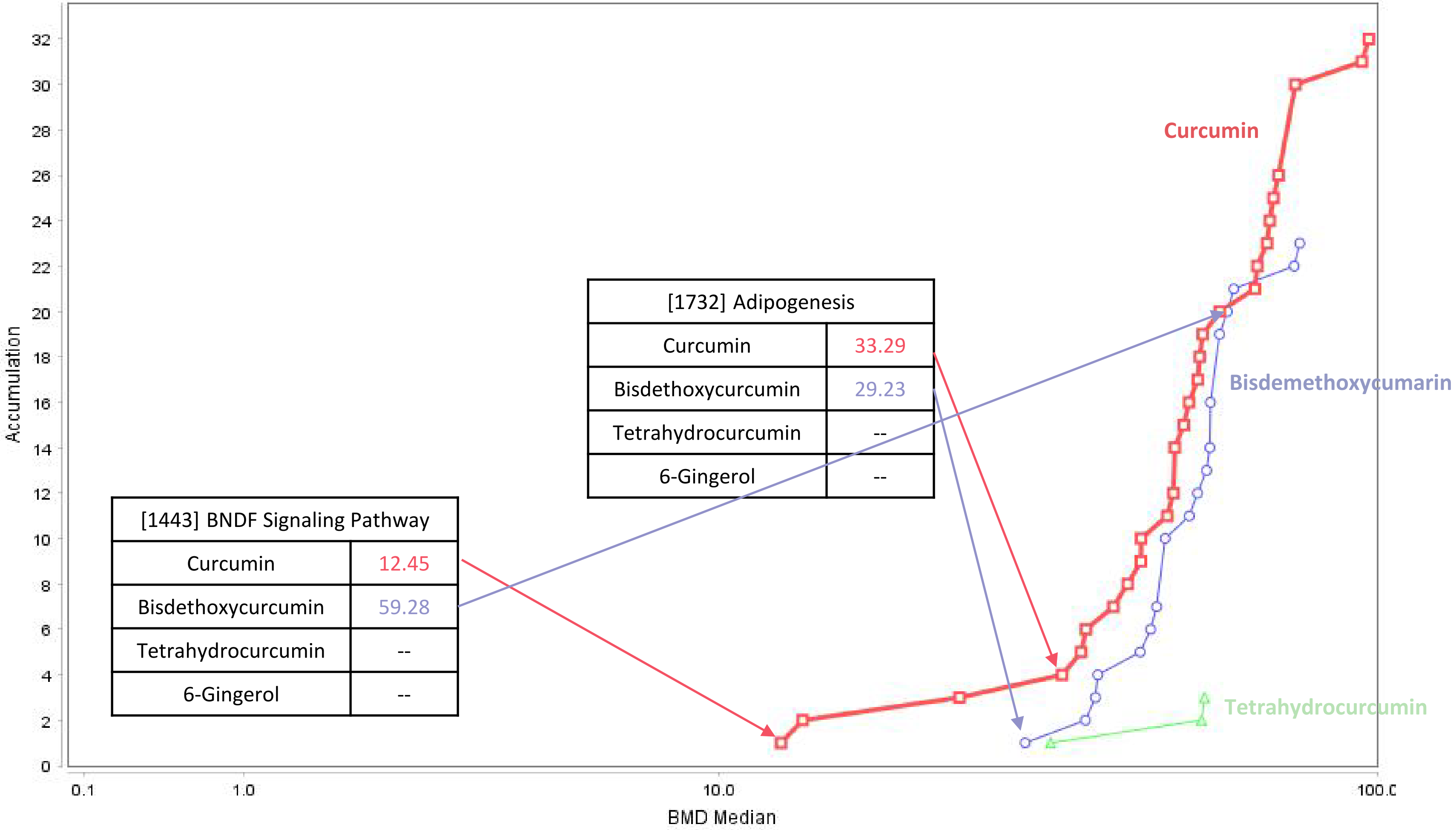
Tetrahydrocurcumin | DTXSID30865801





(6)-Gingerol | DTXSID3041035





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

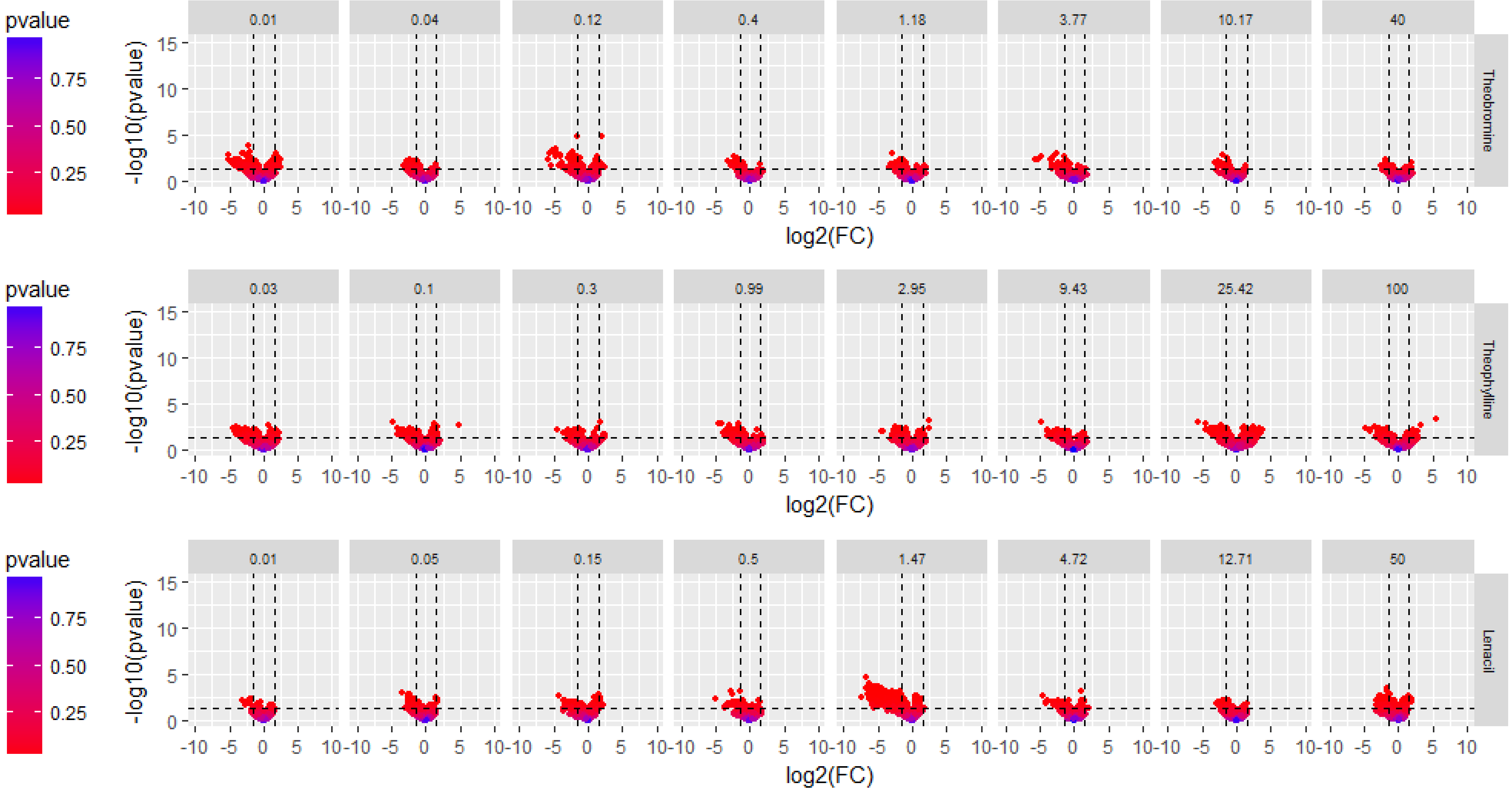
[1443] BNDF Signaling Pathway	
Curcumin	12.45
Bisdethoxycurcumin	59.28
Tetrahydrocurcumin	--
6-Gingerol	--

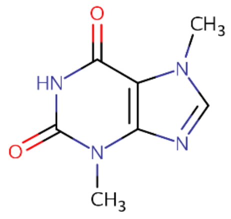
Neuropeptides. 2016 Apr;56:25-31. doi: 10.1016/j.npep.2015.11.003. Epub 2015 Nov 11.
Effect of curcumin on serum brain-derived neurotrophic factor levels in women with premenstrual syndrome: A randomized, double-blind, placebo-controlled trial. Fanaei H[1], Khayat S[2], Kasaeian A[3], Javadimehr M[4].

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].

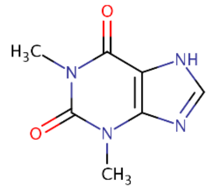
United States Environmental Protection Agency

Caffeine Analogs

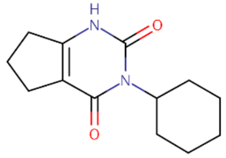




Theobromine | DTXSID9026132



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:

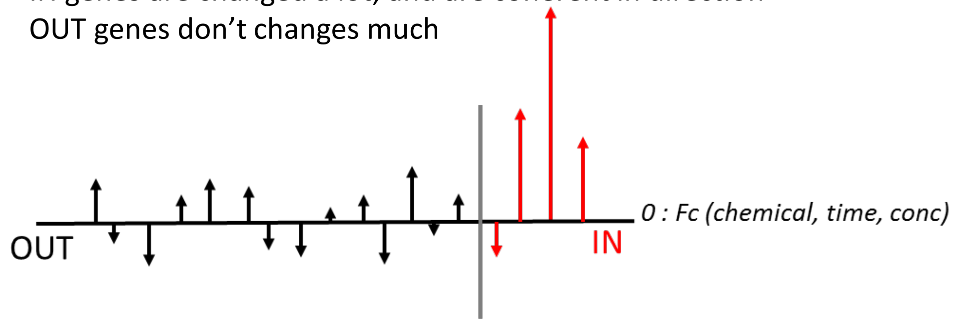
$$M = \sum_{i=1}^{ngene} \frac{\mathrm{fc}_i}{\mathrm{sc}_i^2} \Big/ \sum_{i=1}^{ngene} \frac{1}{\mathrm{sc}_i^2}$$

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

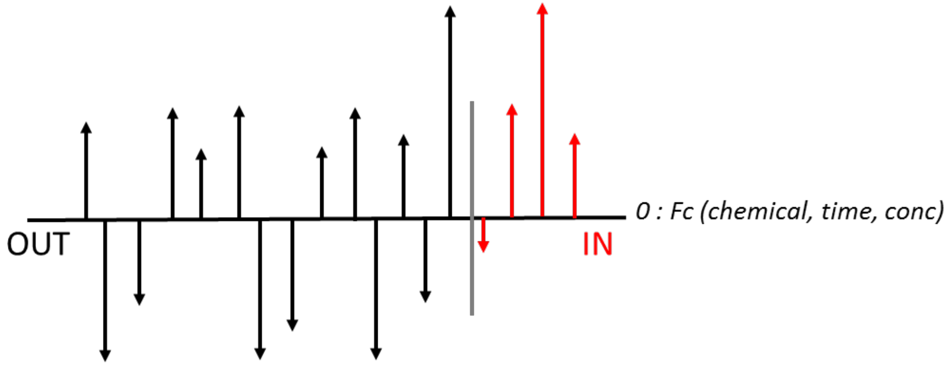
Step 2: CR Modeling

- 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

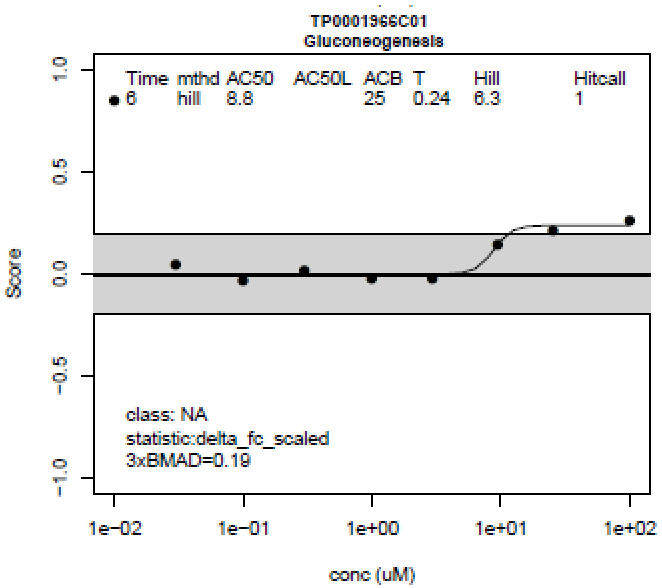
- IN genes are changed a lot, and are coherent in direction
- OUT genes don't changes much

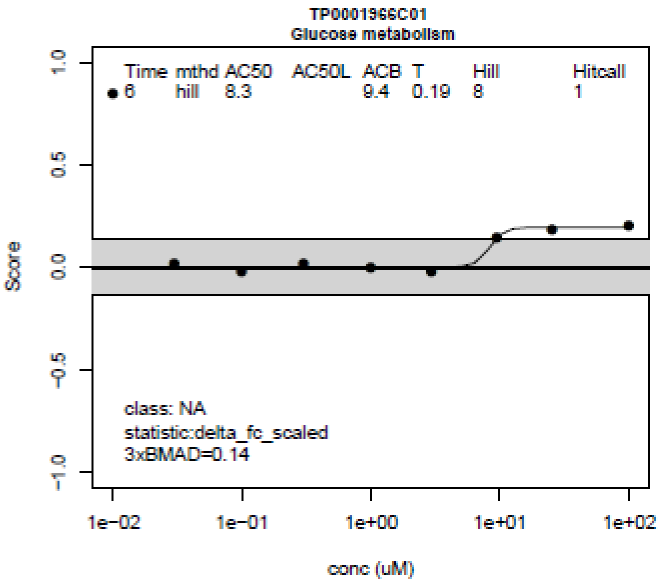


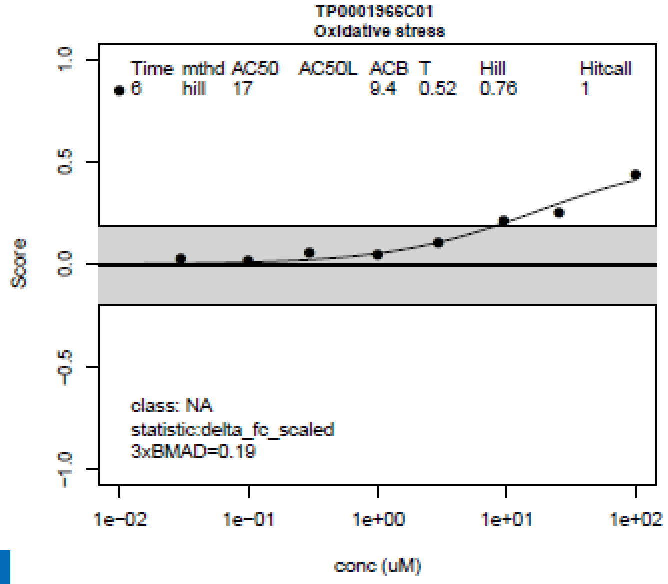
- IN genes are changed a lot, and are coherent in direction
- OUT genes change a lot but are not coherent (mean ~ 0)

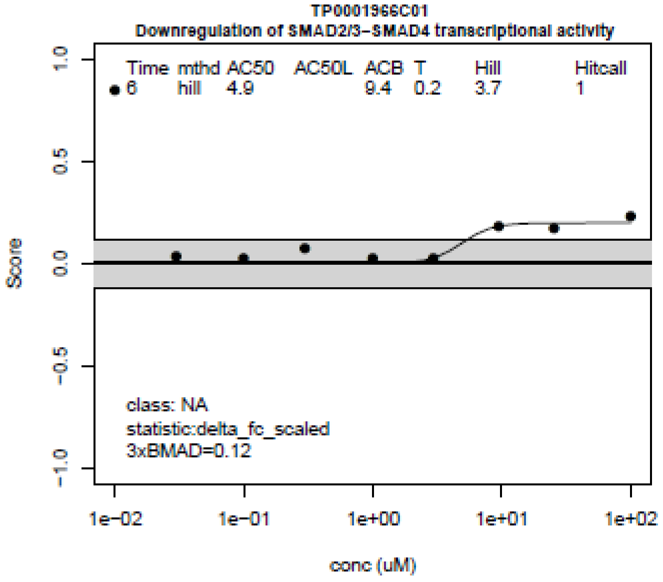


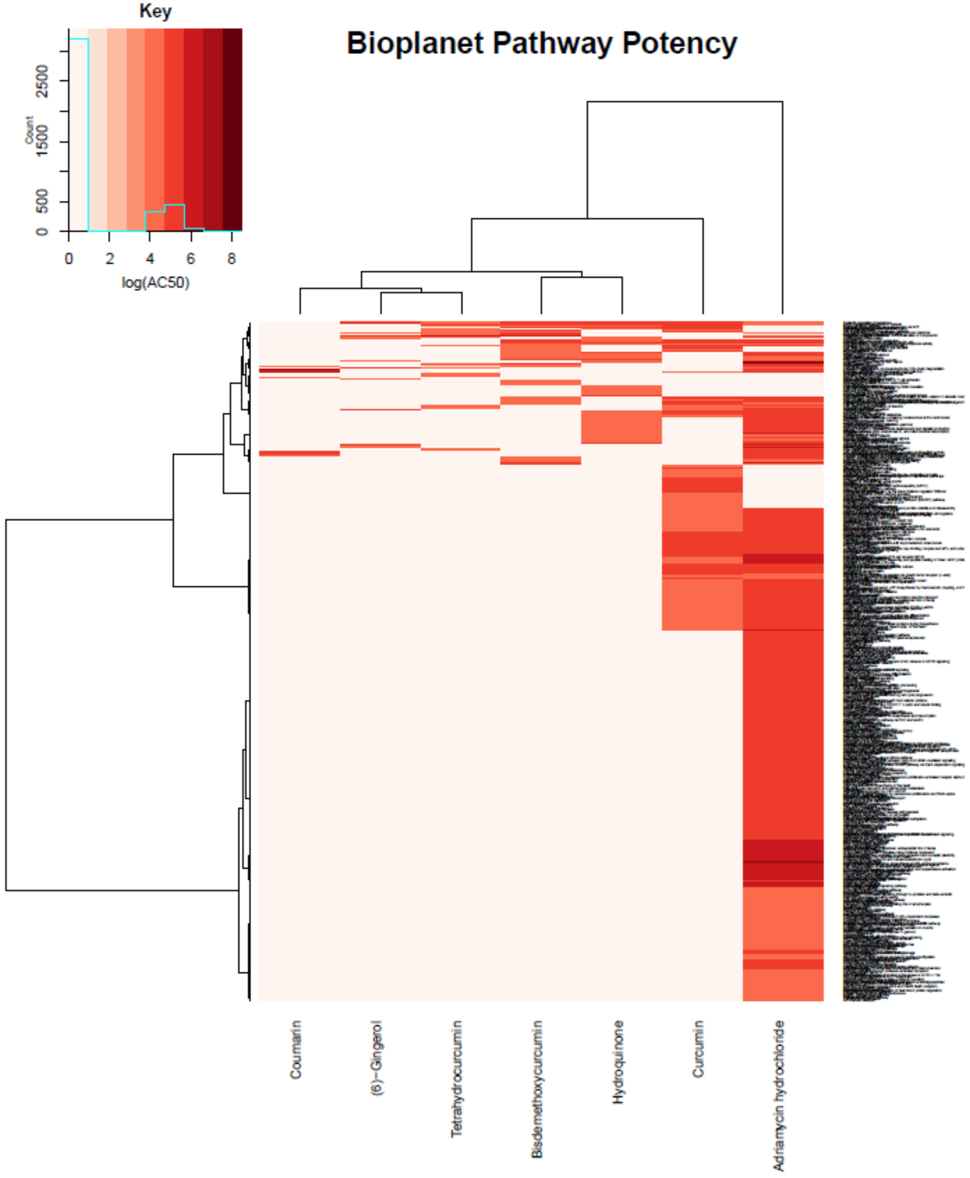
Comparing Activities Across Chemicals







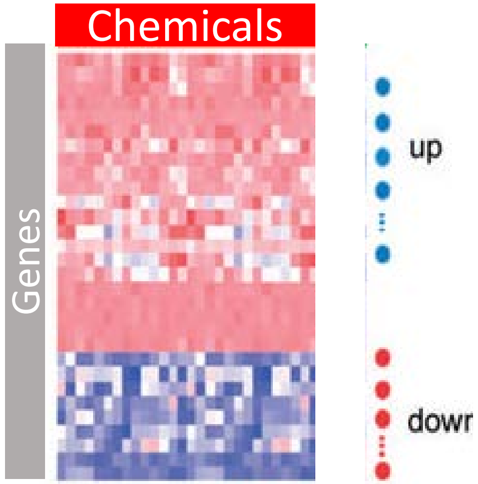




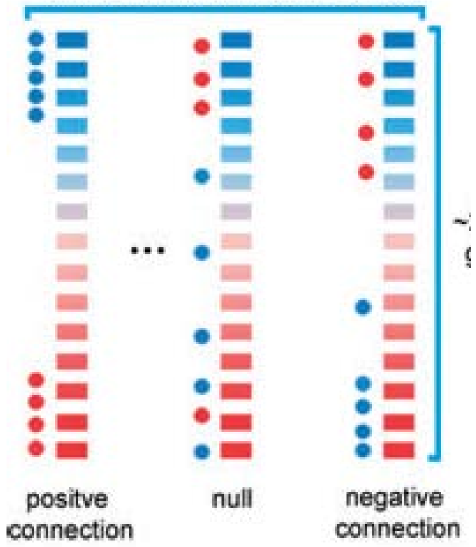
EPA United States Environmental Protection Agency

MOA by Connectivity Mapping

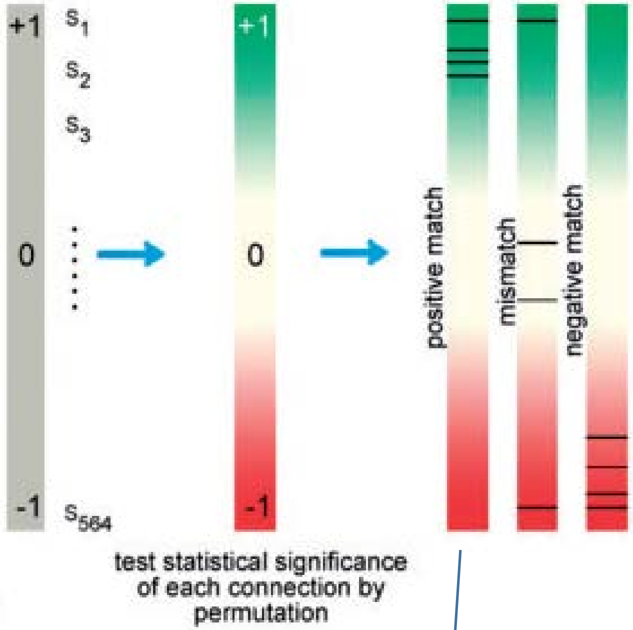
Input Profile



Query Signature DB
CMap or BSP



Find best positive matches



Infer MIE/Target by best match

Issues

- Translating DEG/CRG to signature
- Many measures of similarity
- Only as good as reference chemical MoA annotation
- Highly sensitive but not very specific
- Chemicals that cause global perturbations “hit” all MoAs – how do we distinguish signal from noise ?

Lamb *et al* (2006)
Musa *el al* (2017)

United States Environmental Protection Agency

Various Approaches to Connectivity Mapping

DRG: Differentially Expressed Gene → Based on single-conc. comparison to control.
CRG: Concentration-Responsive Gene → Based on ANOVA or conc.-response modeling.

CMap vs CMap

DEGs
- **Query:** Affymetrix CMap DEGs
- **DB:** Affymetrix CMap

HTTr vs HTTr

DEGs
- **Query:** HTTr DEGs
- **DB:** HTTr DEGs

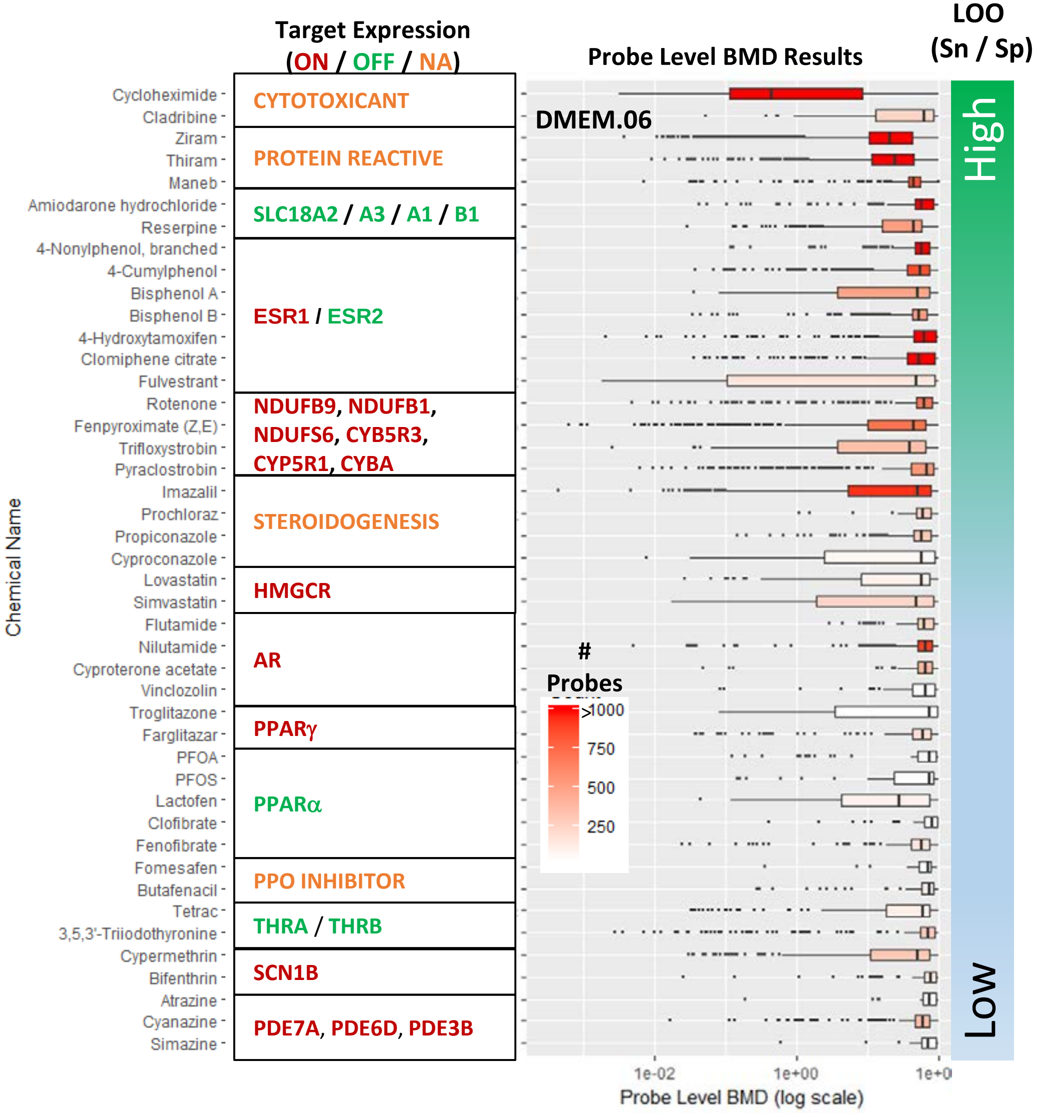
CRGs
- **Query:** HTTr CRGs
- **DB:** HTTr CRGs

HTTr vs CMap

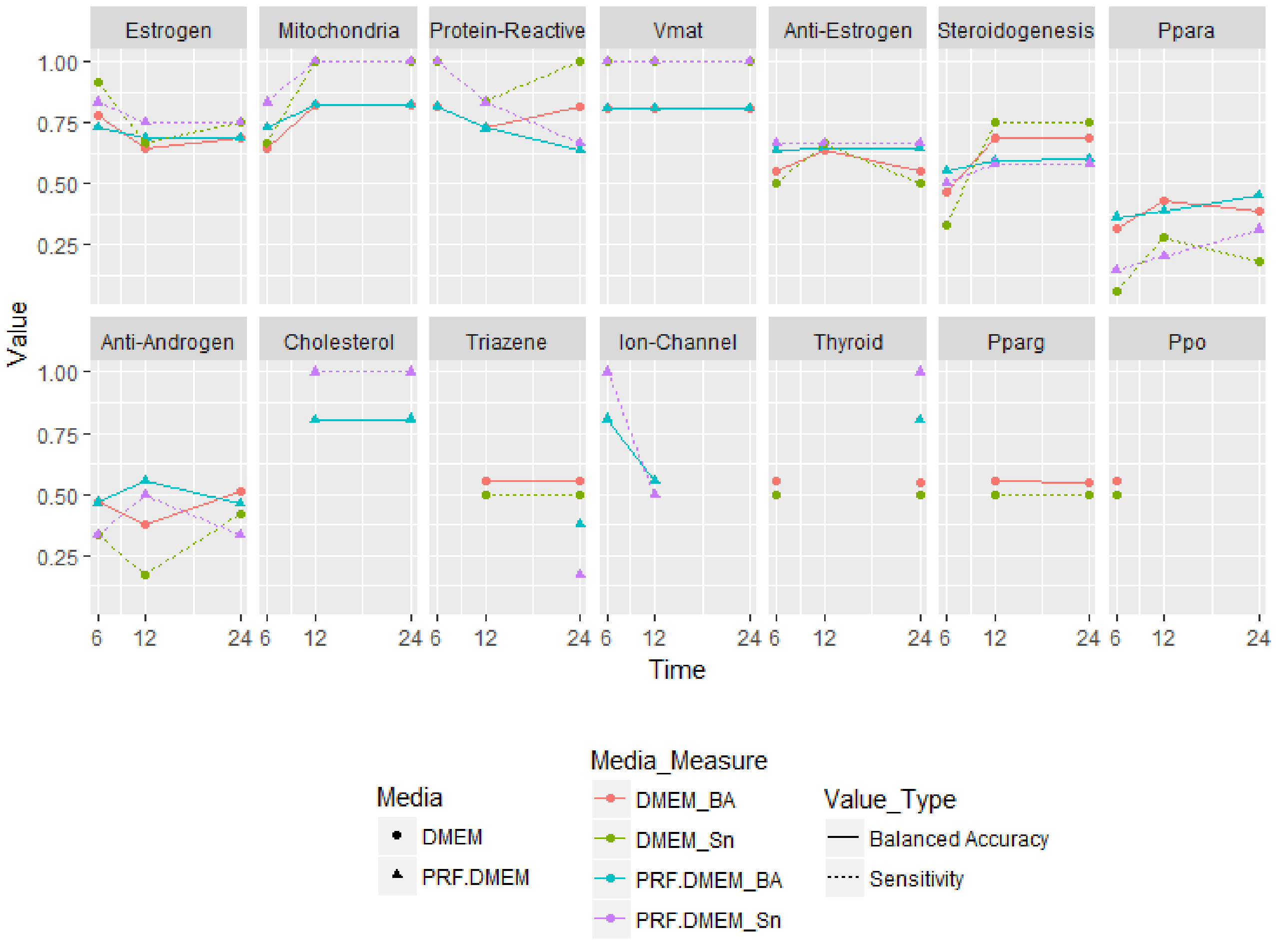
DEGs
- **Query:** TempO-Seq DEGs
- **DB:** Affymetrix CMap

CRGs
- **Query:** TempO-Seq CRGs
- **DB:** Affymetrix CMap

Connectivity Mapping for MoA Prediction



Sensitivity and Balanced Accuracy of MoA Prediction Using CRGs and Leave-One-Out Cross-Validation



United States Environmental Protection Agency

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