EPA
United States
Environmental Protection
Agency

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)

Unilever Safety & Environment Assurance Centre
Bedford, UK
April 18th, 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.

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

USEPA NCCT / Unilever CRADA:

- Paul Carmichael
- Andy White
- Sophie Malcomber
- Richard Stark*

"…further research and development of the ToxCast technology, high-throughput transcriptomics, integration of metabolic competence in high-throughput in vitro assays, and its translation of the results into risk assessment for use by private and public entities."

EPA United States Environmental Protection Agency

Outline

- **Background**
  - Computational Toxicology Strategic Vision
  - Tier 1 Screening Assays

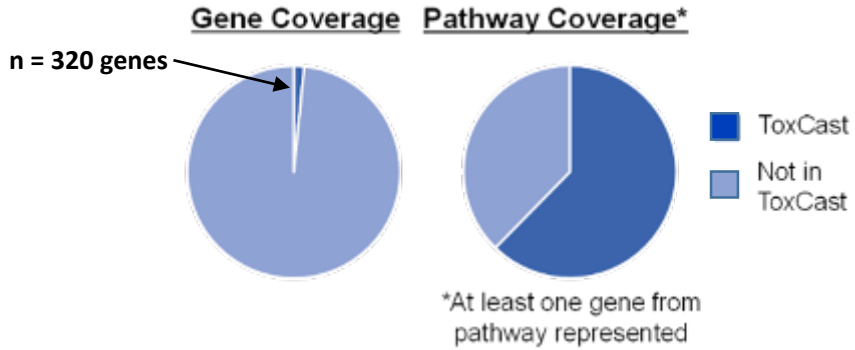
- **TempO-Seq**
  - TempO-Seq Technology Overview
  - Screening Format
  - Technical Reproducibility
  - Applications

- **HCI Phenotypic Profiling Assay**
  - Assay Description
  - Laboratory & Analysis Workflows
  - Concentration-response Modeling
  - Applications

- **Computational Modeling of Biological Diversity in Cell Lines**
  - Cell Line Selection
  - NCCT/Unilever CRADA Addendum

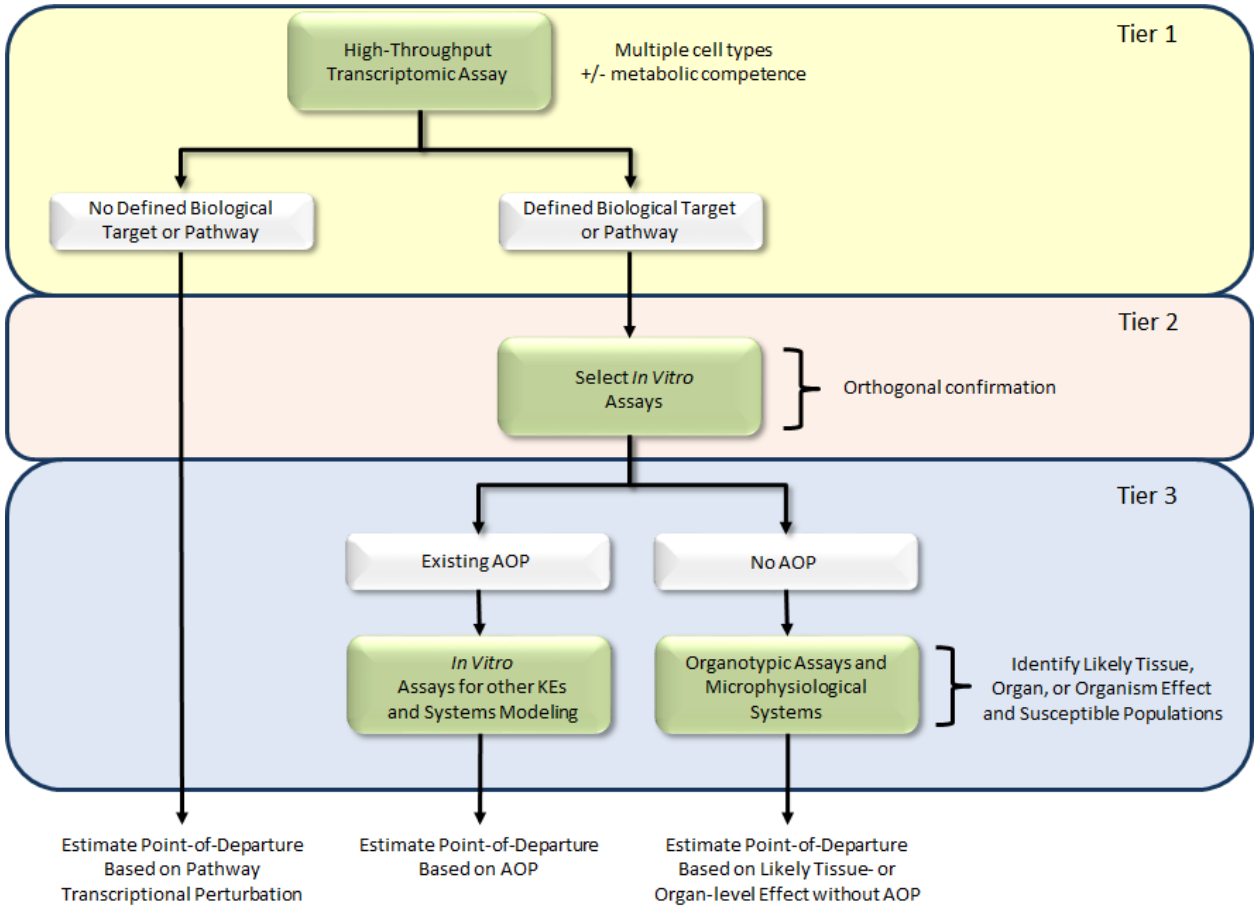
Background

- 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
  - High content → yield many measurements per sample.
  - Increasing efficiency and declining cost of generating whole transcriptome profiles has made **high-throughput transcriptomics (HTTr)** a practical option for chemical screening
  - Certain types of **high content imaging (HCI)** are also cost effective mean for profiling the effects of chemicals and screening.

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

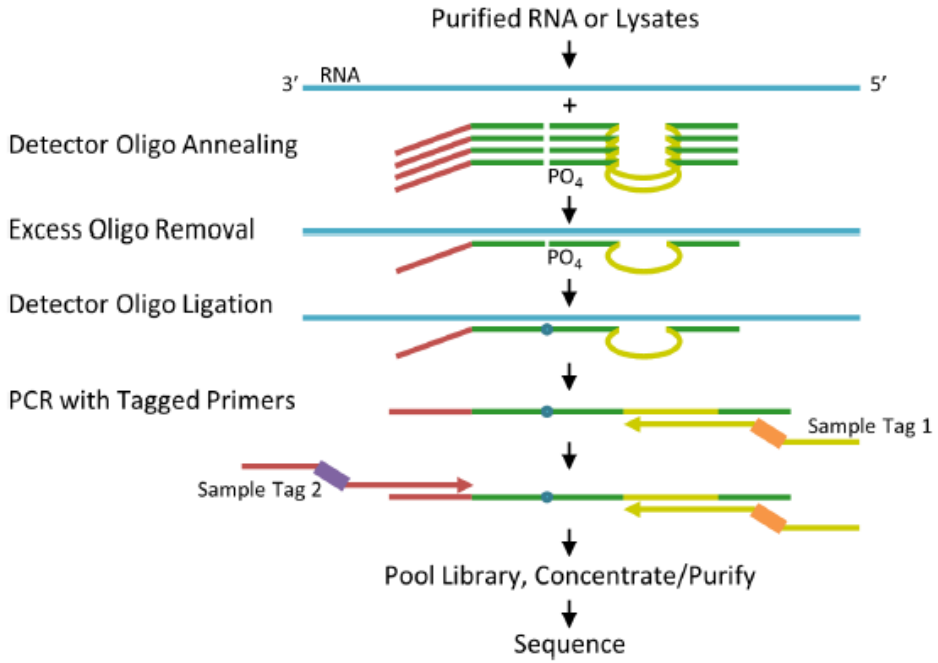


Background

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) no detection of intronic sequences or ribosomal RNA (*common in RNA-Seq*)
  - 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

Cryopreserved Cell Stocks

Cell Expansion

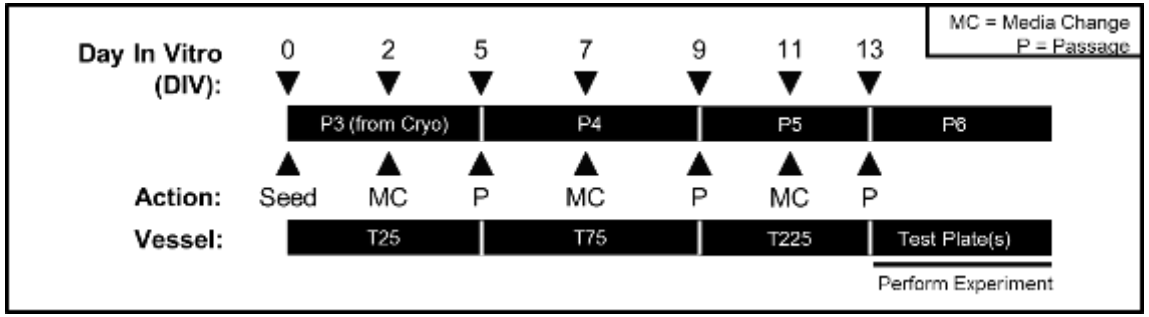
Cell Plating

BioTek MultiFlo ™ FX

Dispensing Test Chemicals

LabCyte Echo® 550 Liquid Handler

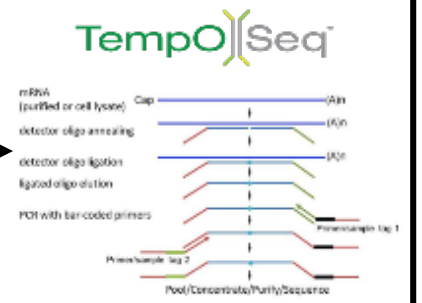
Standardized Expansion Protocol



Track 1: Targeted RNA-Seq

Generating Cell Lysates

TempO-Seq WT



Track 2: Apoptosis / Cell Viability

Reagent Dispensing

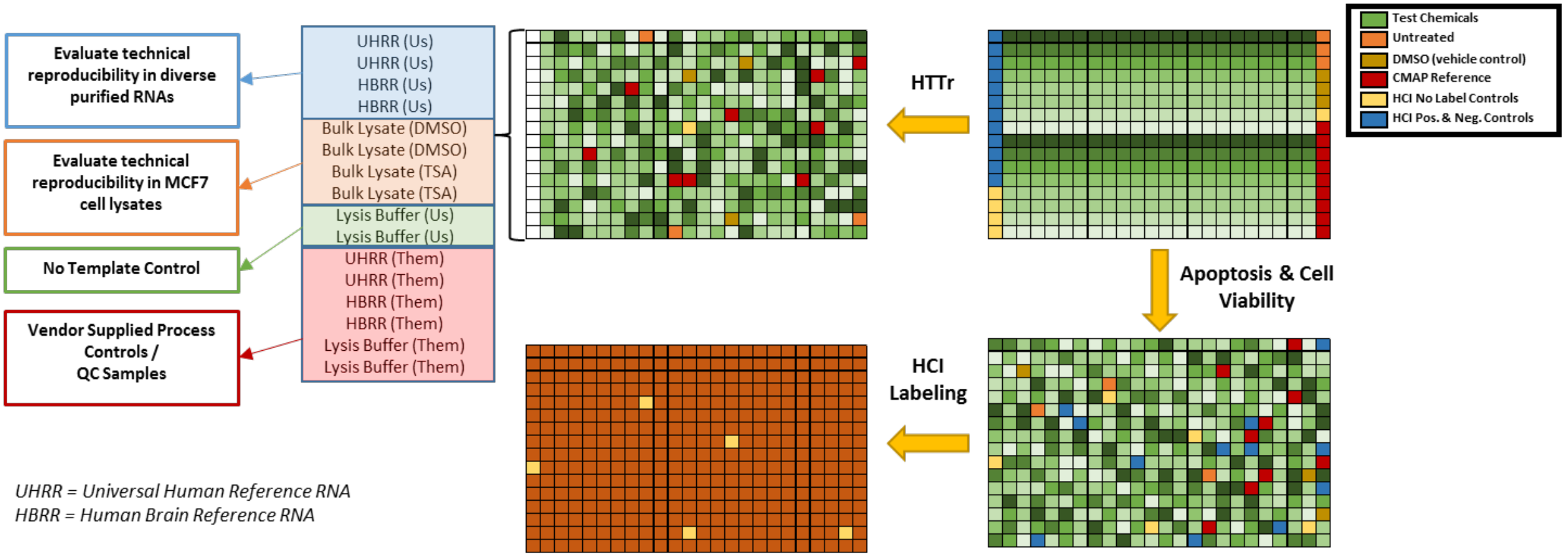
High Content Imaging

Perkin Elmer Opera Phenix™ High Content Screening System

Treatment Randomization & Quality Control Samples

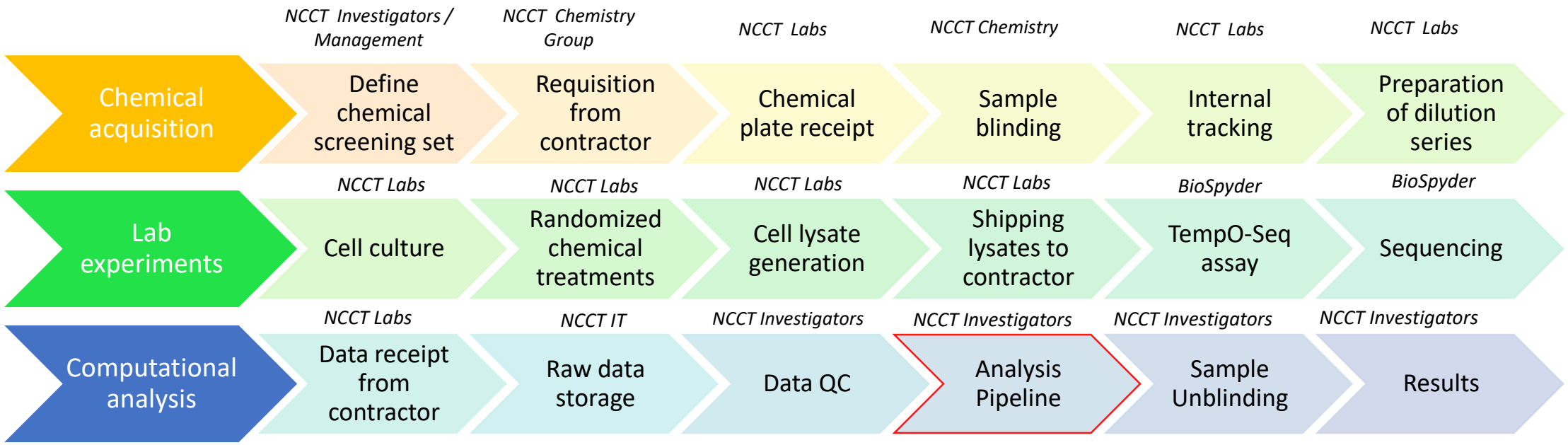
Treatment Randomization: *Each test plate uniquely randomized with respect to treatment.*

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

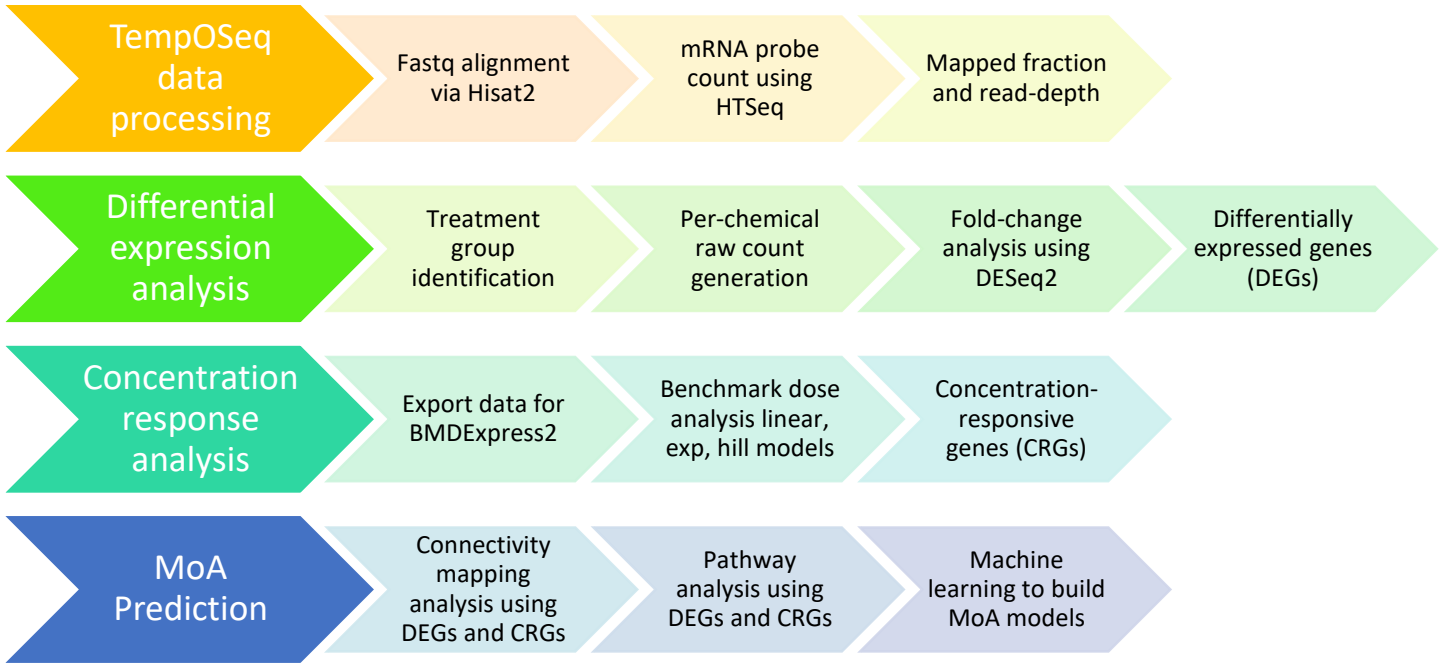


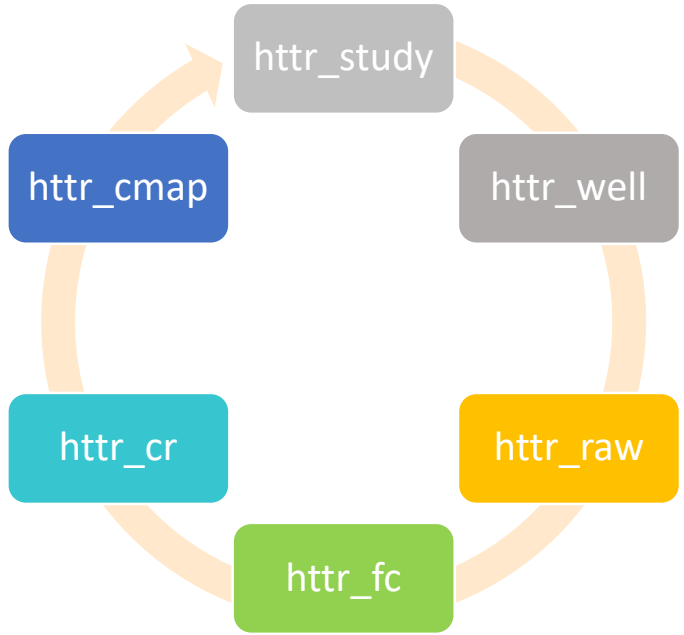
UHRR = Universal Human Reference RNA
HBRR = Human Brain Reference RNA

HTTr Screening, An "NCCT-Wide" Effort



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

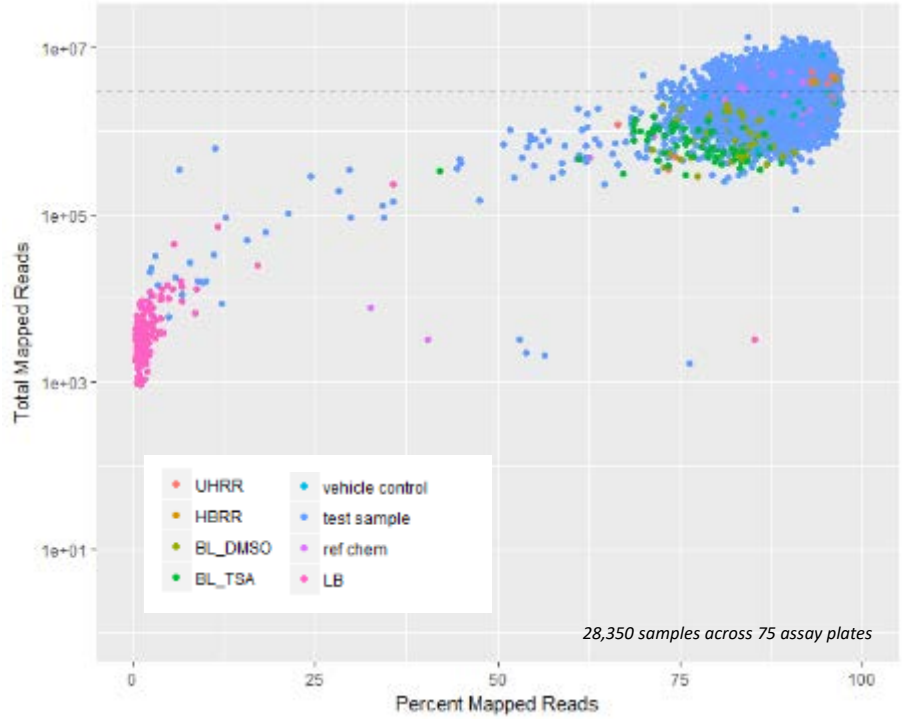
mongodb://pb.epa.gov/httr_v1
mongodb://pb.epa.gov/cmap_v2

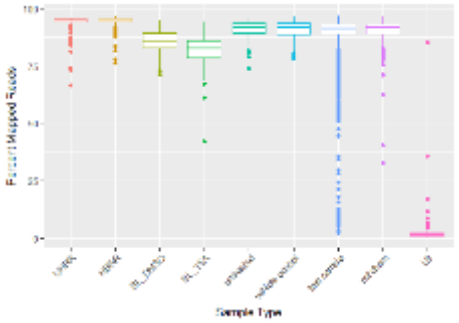
NCCT Investigators
NCCT IT

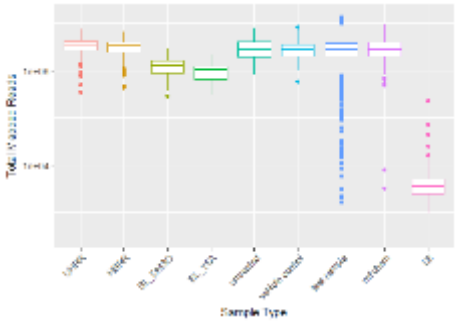
EPA
United States
Environmental Protection
Agency

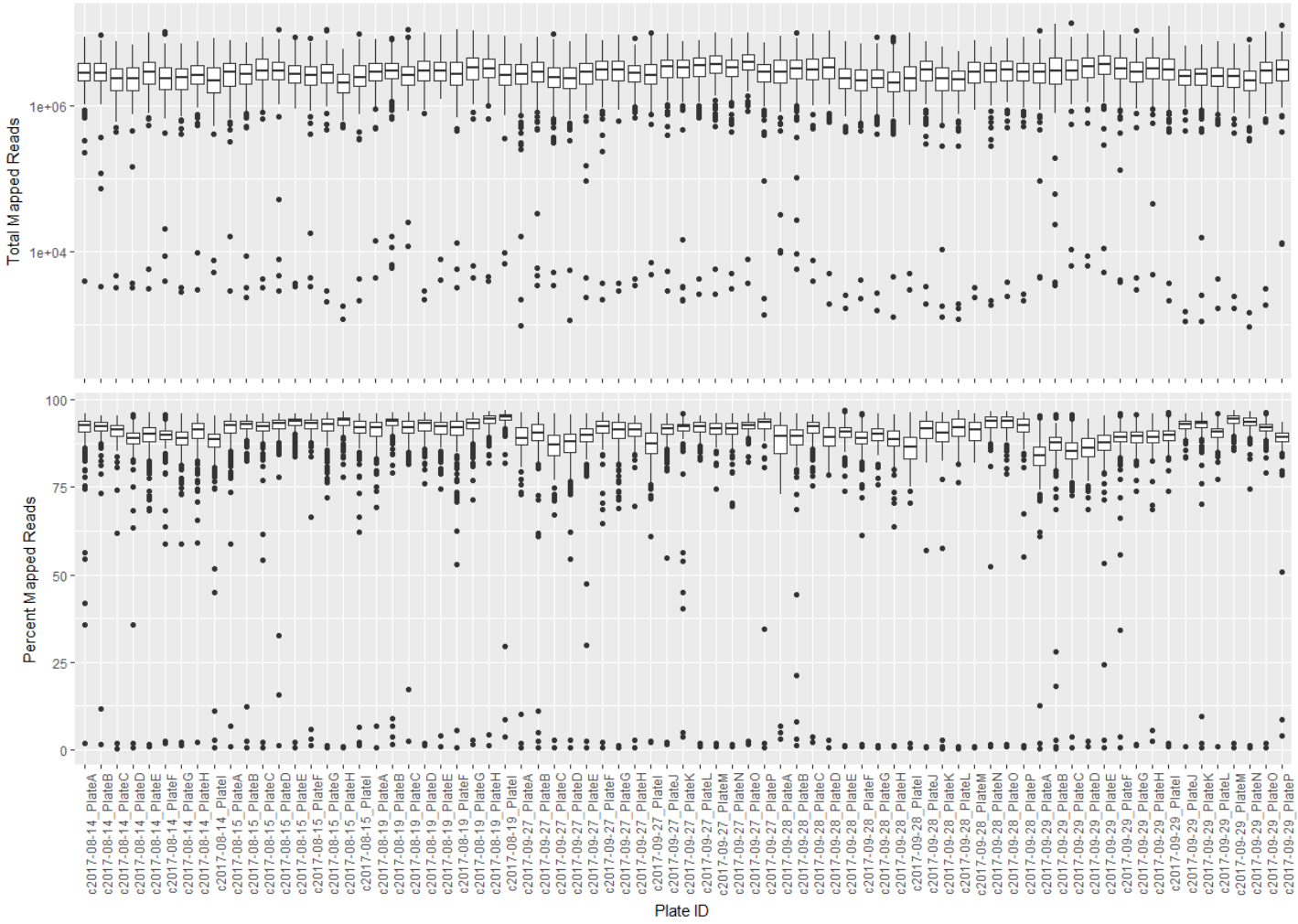
Technical Reproducibility

Reproducibility of Read Depths and Mapping Rates



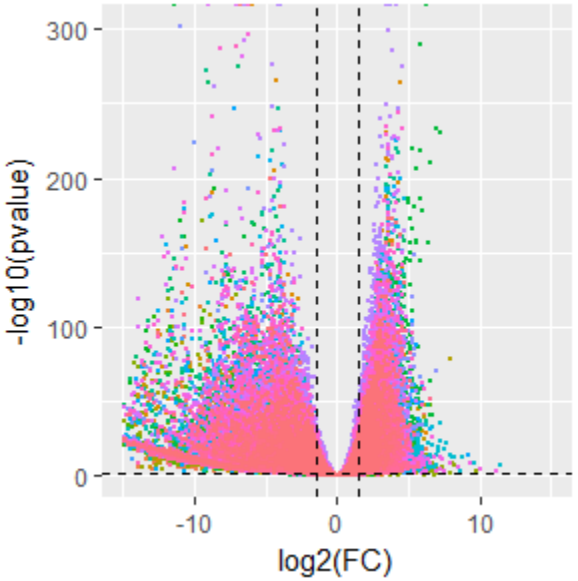




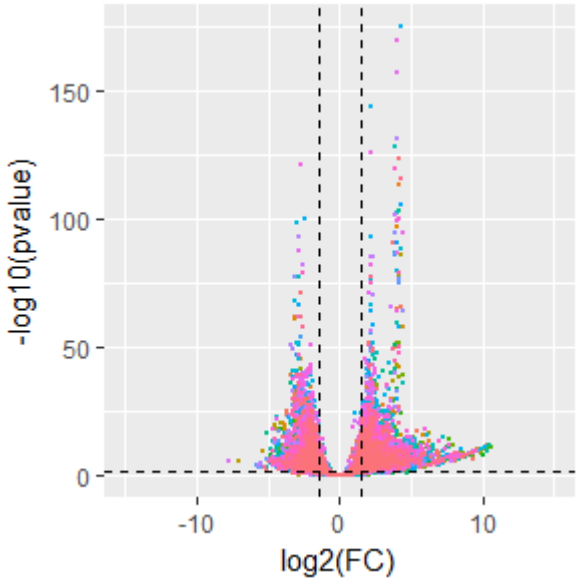


Reproducibility of Log_2(FC) Estimates

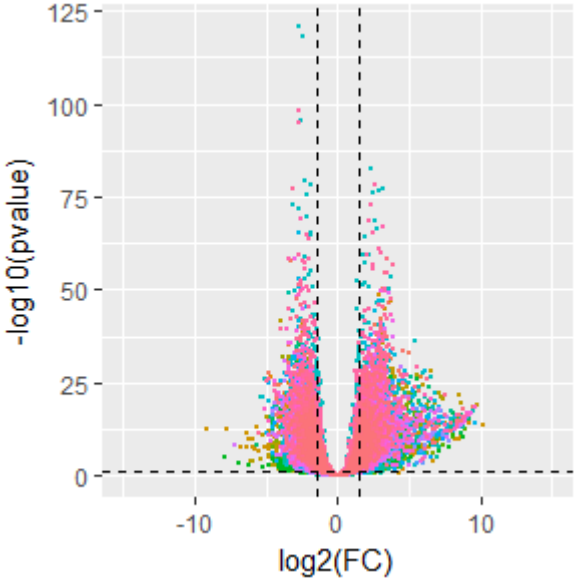
UHRR v HBRR

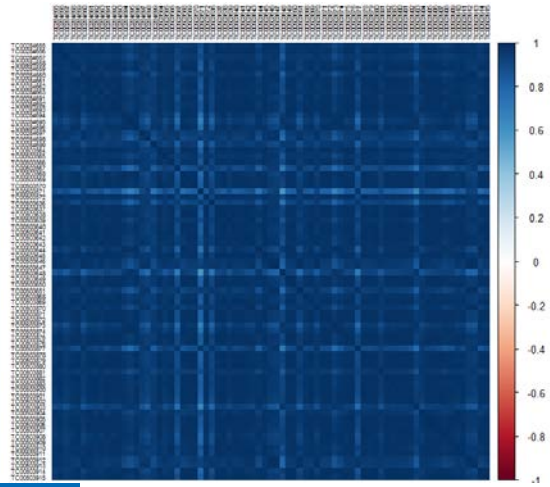


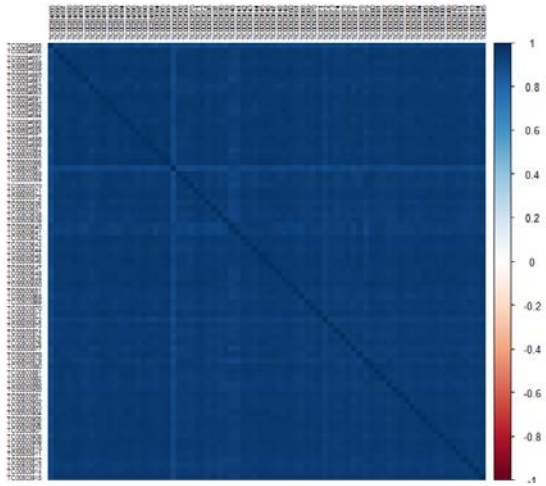
DMSO vs TSA (Bulk Lysate)

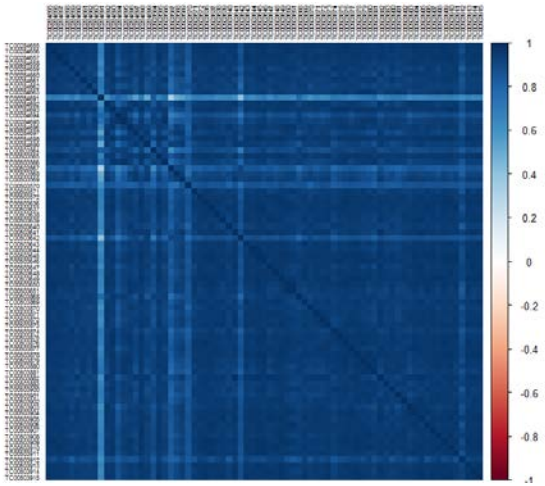


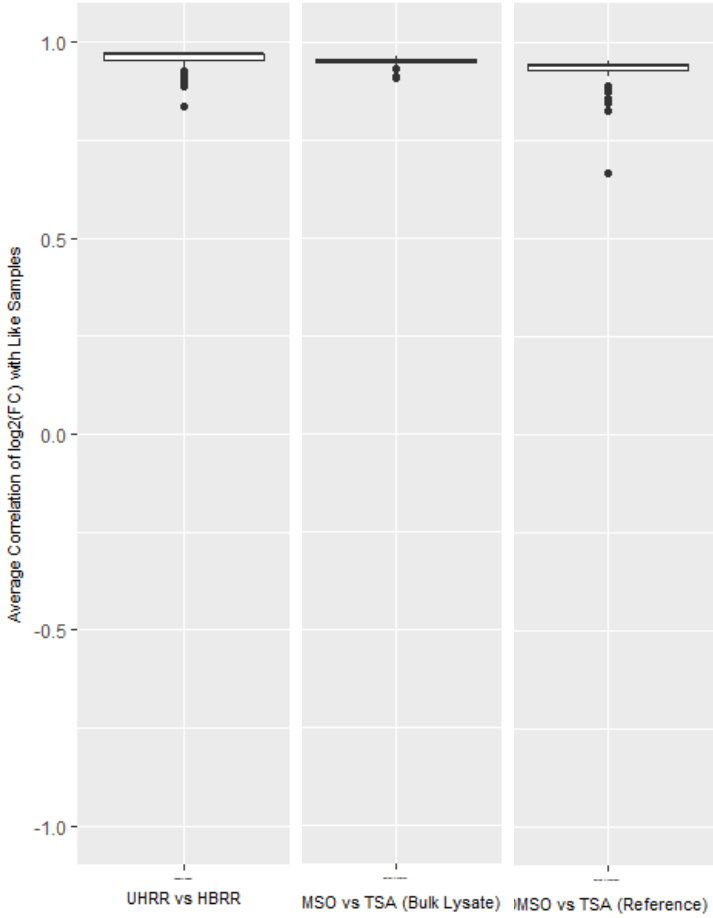
DMSO vs TSA (Treatment)











EPA United States Environmental Protection Agency

Concentration-Response Modeling / Bioactivity Thresholds

Benchmark Dose Modeling

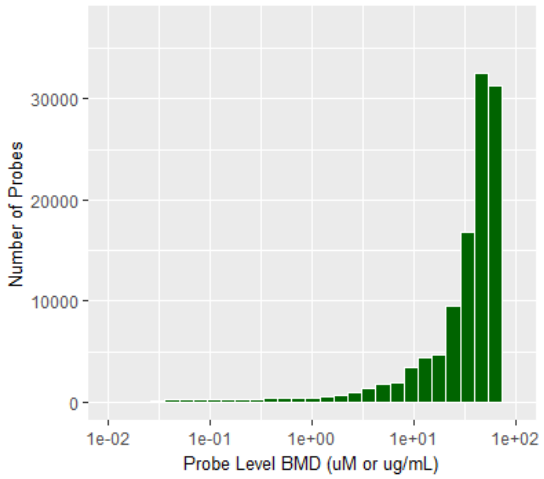
Parameter	Criteria [a]
Pre-filter:	ANOVA ($p_{raw} < 0.05$ & $\|FC\| \geq 1.5$)
Models	Hill, Exponential 2, *poly2, power, linear*
BMR Factor:	1.349 (10 %)
Best Model Selection:	Lowest AIC
Hill Model Flagging [b]:	'k' < 1/3 Lowest Positive Dose Retain Flagged Models
Pathway Analysis:	Genes with BMD <= Highest Dose ≥ 3 ≥ 1% Gene Set Coverage
Gene Set Collections [c]:	MSigDB_C2 MSigDB_H Reactome BioPlanet KEGG

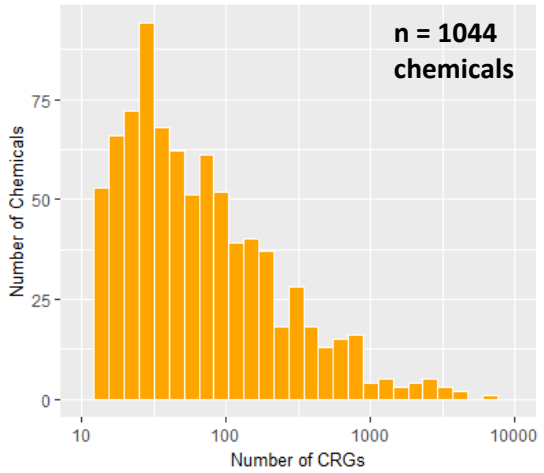
[a] Exploratory analysis – modeling criteria not finalized

[c] Gene Set Collections:

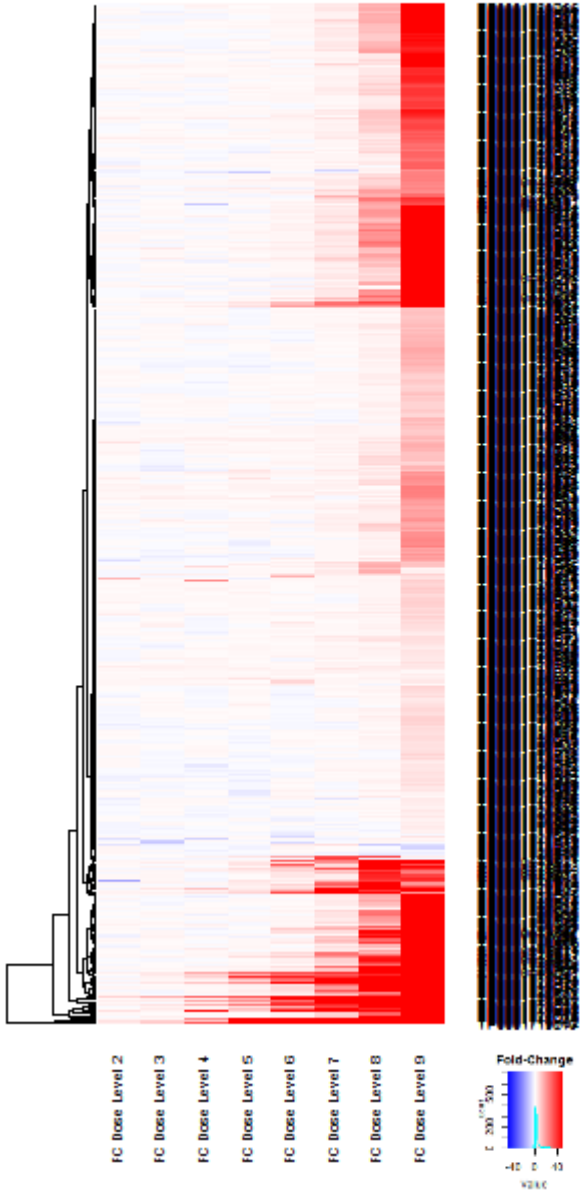
- **MSigDB_C2:** Curated gene sets from online pathway databases, publications and knowledge of domain experts (n = 4738).
- **MSigDB_H:** Coherently expressed signatures derived by aggregating many MSigDB gene sets to represent well-defined biological states or processes (n = 50).
- **Reactome:** Open-source,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

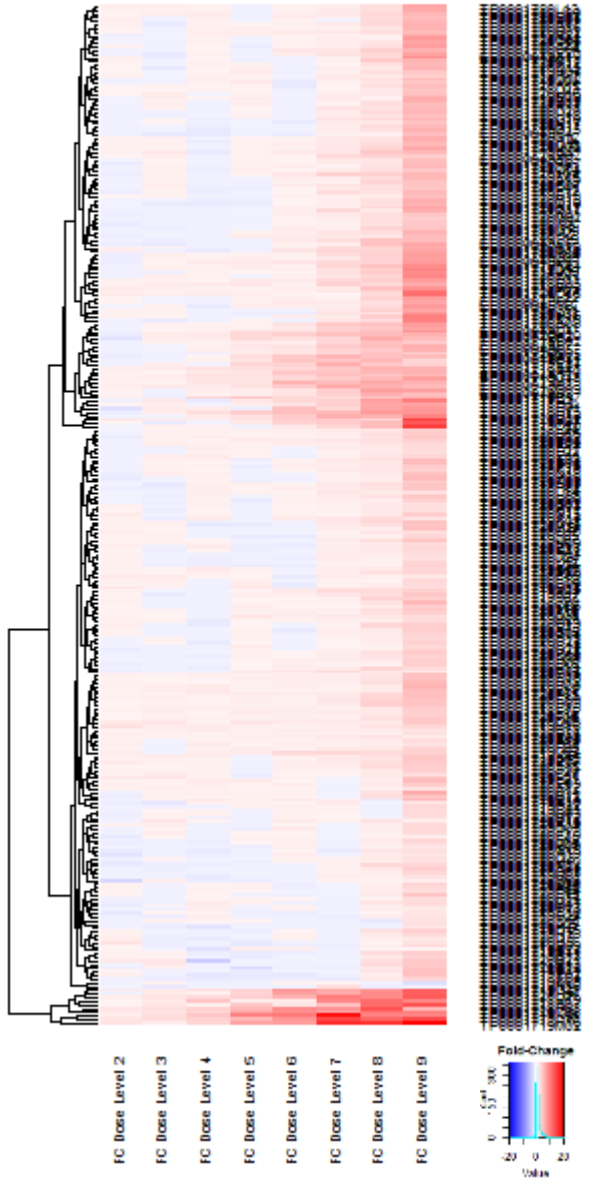




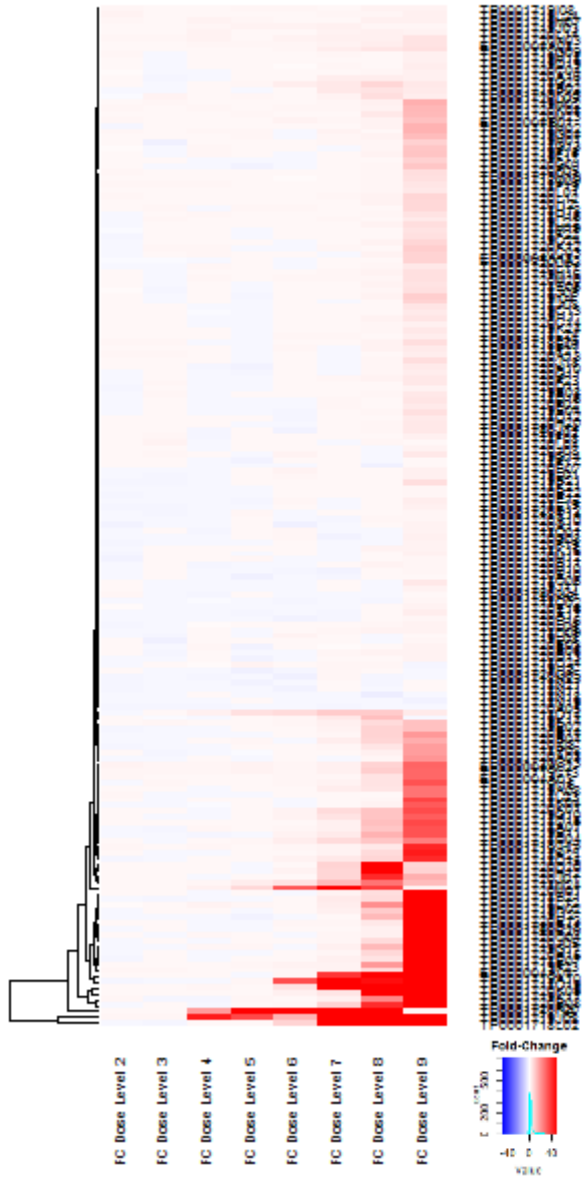
CYP1A1_10775
(n = 473)



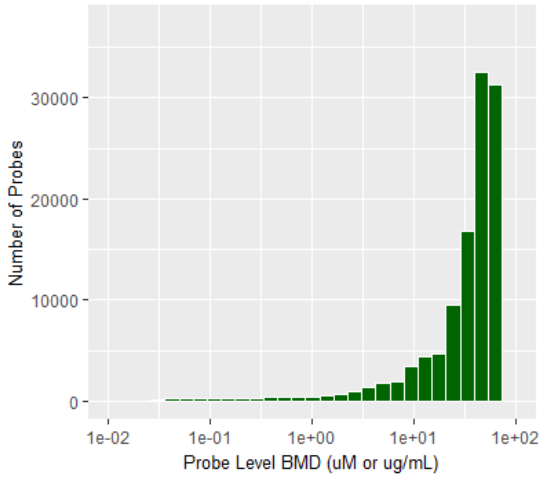
CYP1B1_17315
(n = 279)

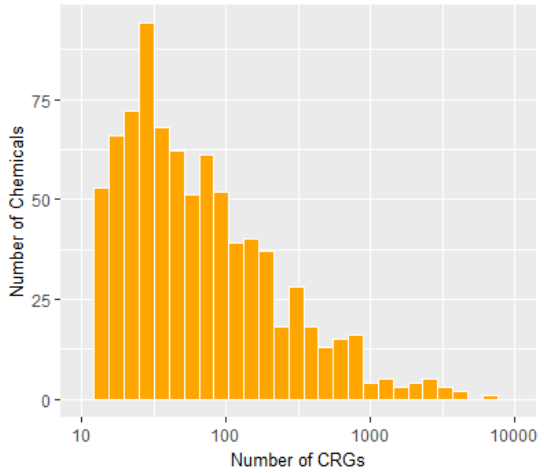


HMOX1_3041
(n = 174)



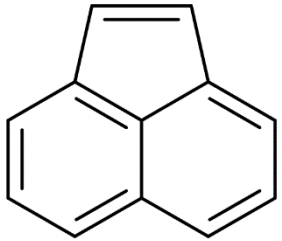
Benchmark Dose Modeling Summary & Inducible Genes

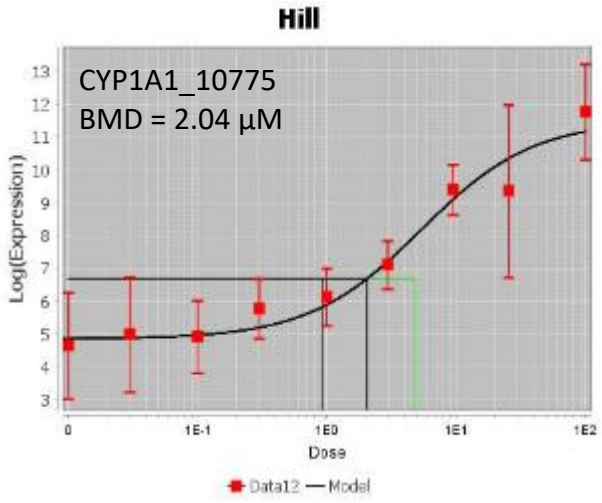




CYP1A1_10775
(n = 473)

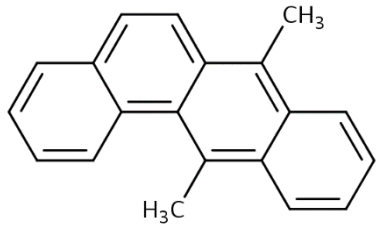
Acenaphthylene
208-96-8 | DTXSID3023845

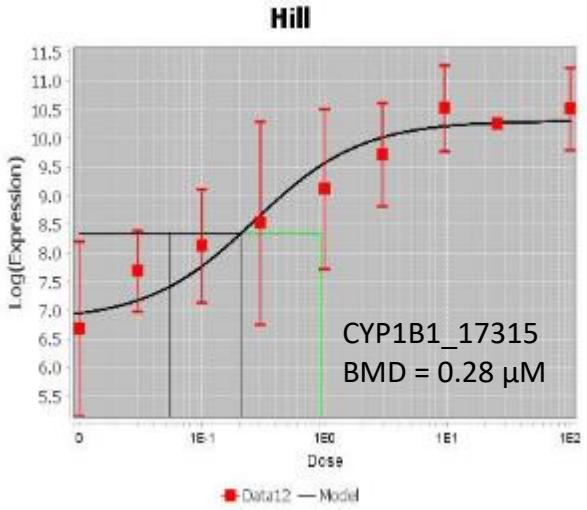




CYP1B1_17315
(n = 279)

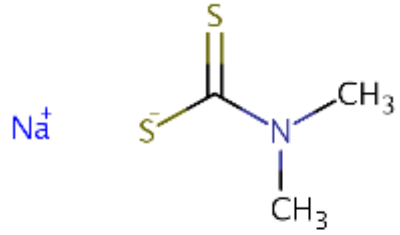
7,12-Dimethylbenz(a)anthracene
57-97-6 | DTXSID1020510

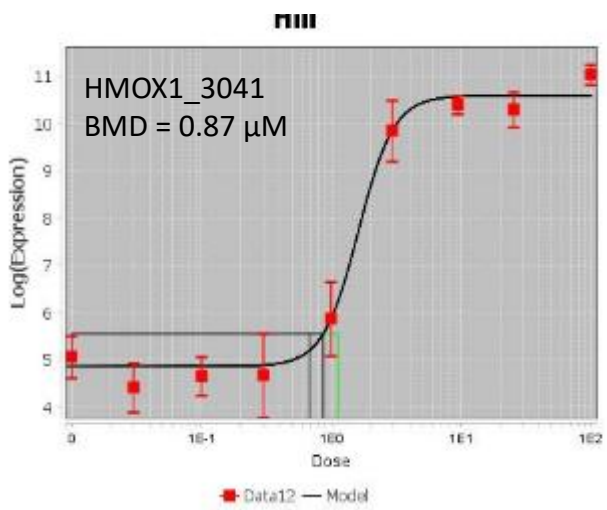




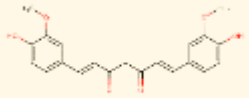
HMOX1_3041
(n = 174)

Sodium
dimethyldithiocarbamate
128-04-1 | DTXSID6027050



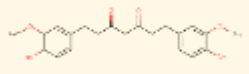


Unilever Case Study Chemicals

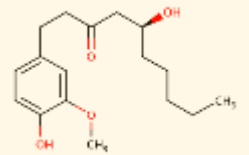


Curcumin | DTXSID8031077

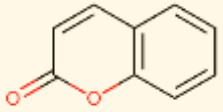
Bisdemethoxycurcumin | DTXSID00872663



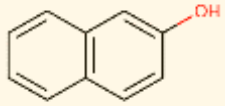
Tetrahydrocurcumin | DTXSID30865801



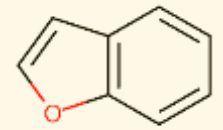
(6)-Gingerol | DTXSID3041035



Coumarin | DTXSID7020348

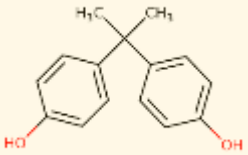


2-Naphthalenol | DTXSID5027061

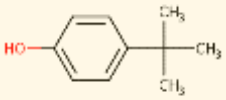


2,3-Benzofuran | DTXSID6020141

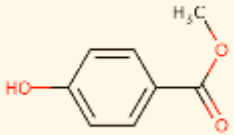
Hydroquinone | DTXSID7020716



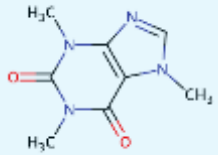
Bisphenol A | DTXSID7020182



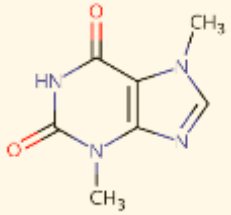
4-tert-Butylphenol | DTXSID1020221



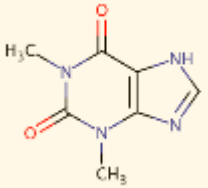
Methylparaben | DTXSID4022529



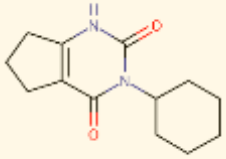
Caffeine | DTXSID0020232



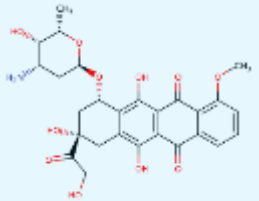
Theobromine | DTXSID9026132



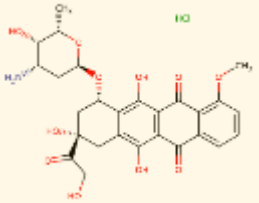
Theophylline | DTXSID5021336



Lenacil | DTXSID9042093



Doxorubicin | DTXSID8021480

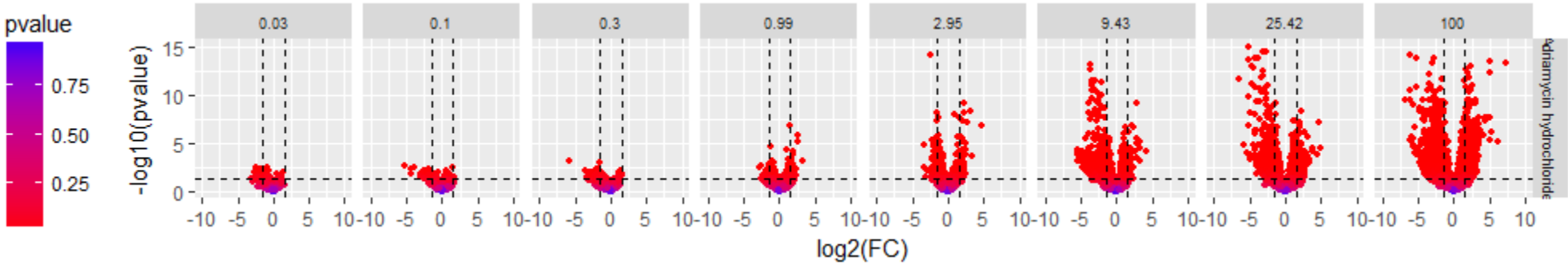


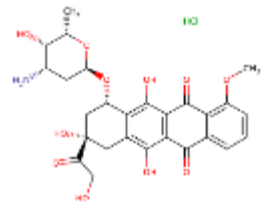
Adriamycin hydrochloride | DTXSID3030636

Present in Block 1, unblinded

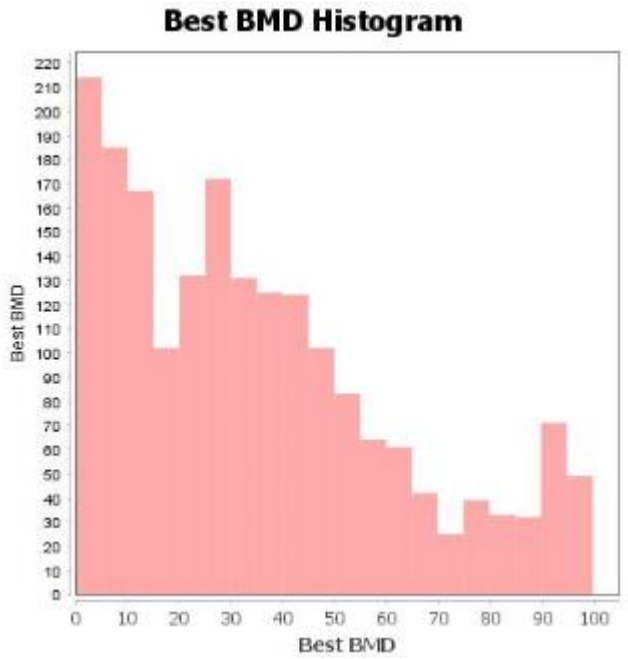
Not present in Block 1

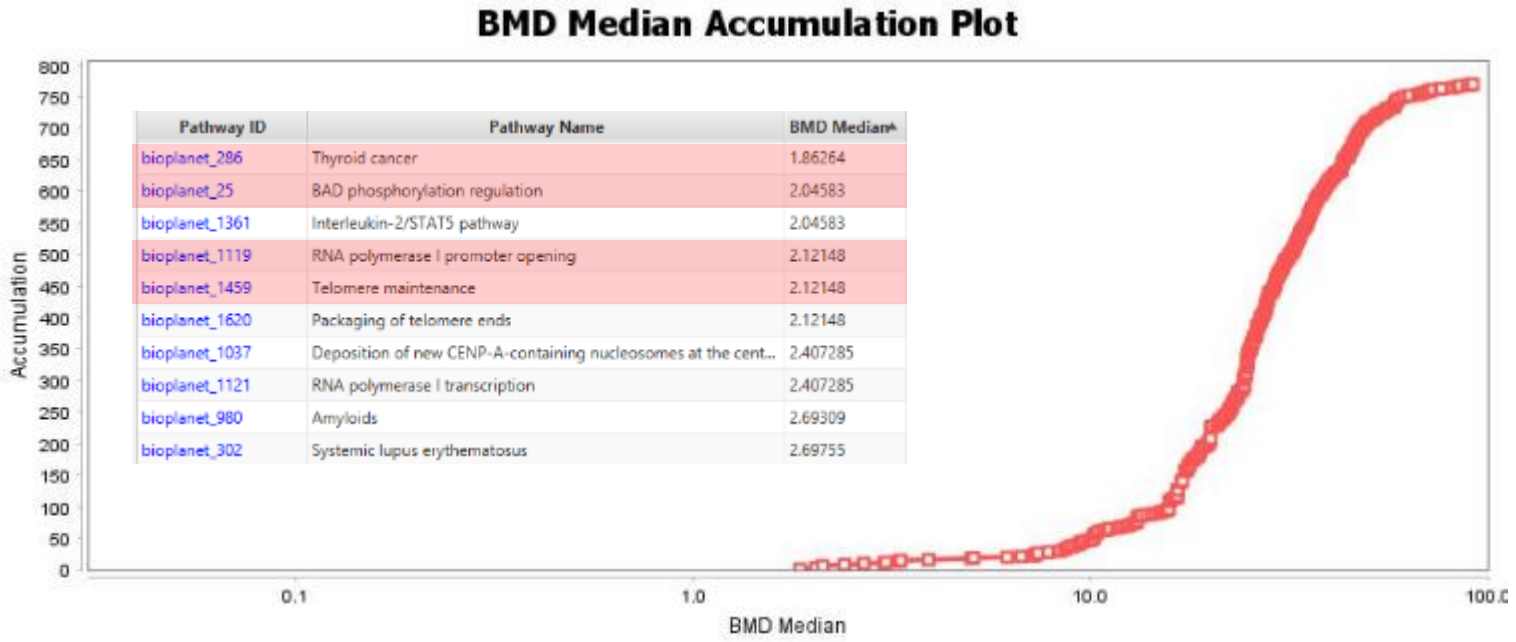
Doxorubicin (aka Adriamycin hydrochloride)





Adriamycin hydrochloride | DTXSID3030636



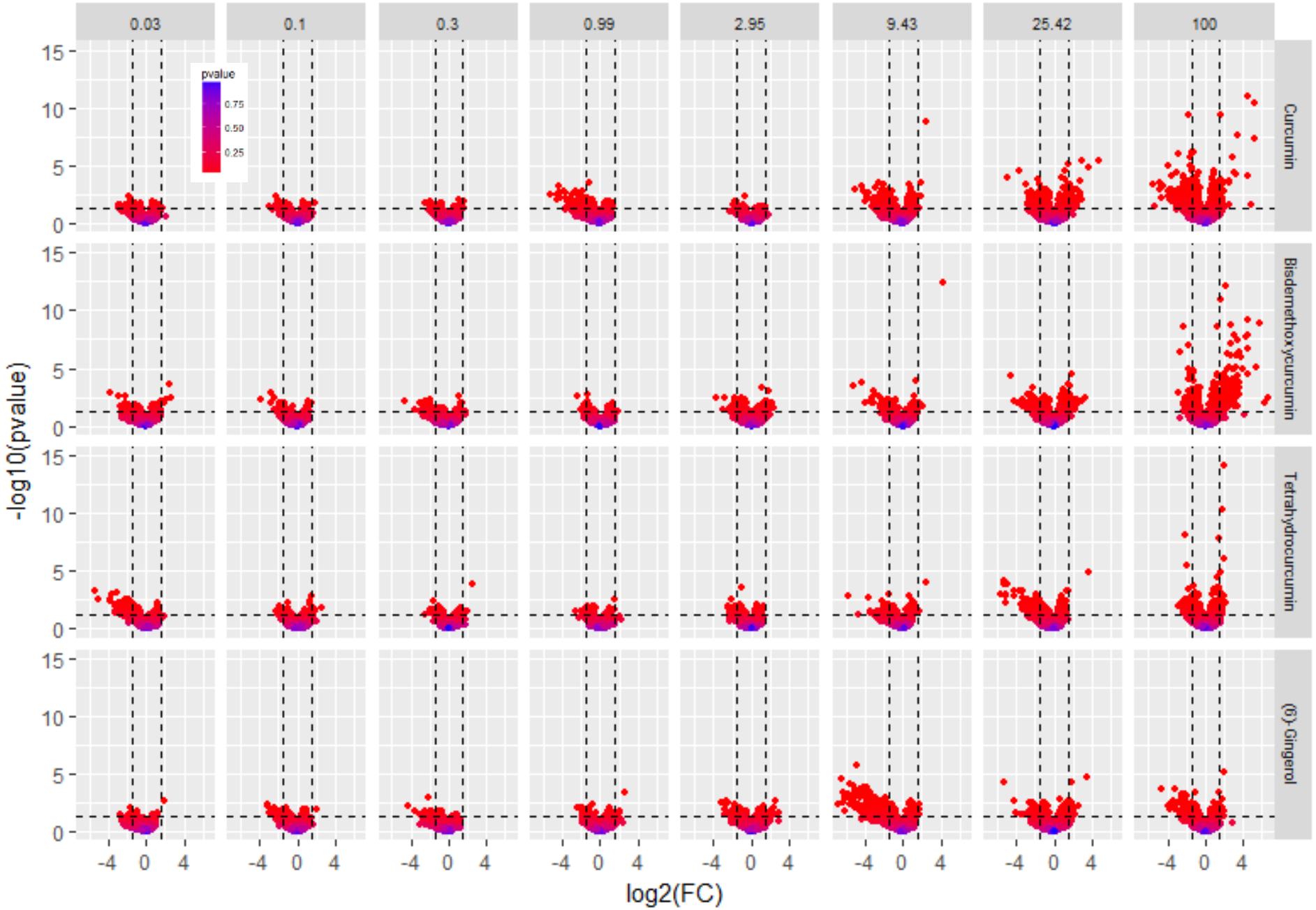


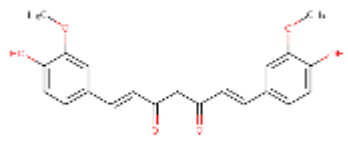
Int J Mol Cell Med. 2015 Spring;4(2):94-102.

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.

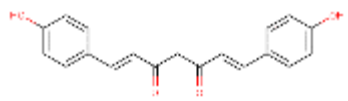
Fatemi A[1], Kazemi A[1], Kashiri M[1], Safa M[2].

Evaluating Structurally Related Chemicals

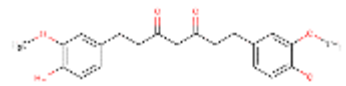




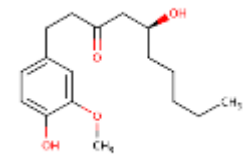
Curcumin | DTXSID8031077



Bisdemethoxycurcumin | DTXSID00872663

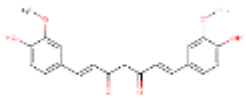


Tetrahydrocurcumin | DTXSID30865801

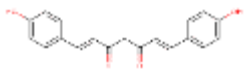


(6)-Gingerol | DTXSID3041035

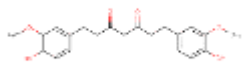
Gene Set Analysis Using BMD Modeling Results



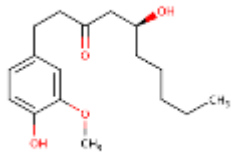
Curcumin | DTXSID8031077



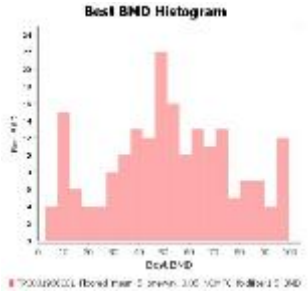
Bisdemethoxycurcumin | DTXSID00872663

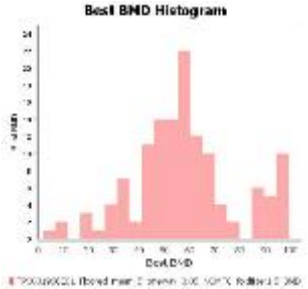


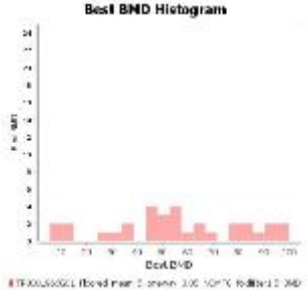
Tetrahydrocurcumin | DTXSID30865801

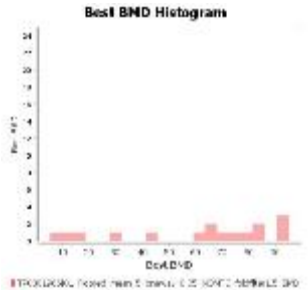


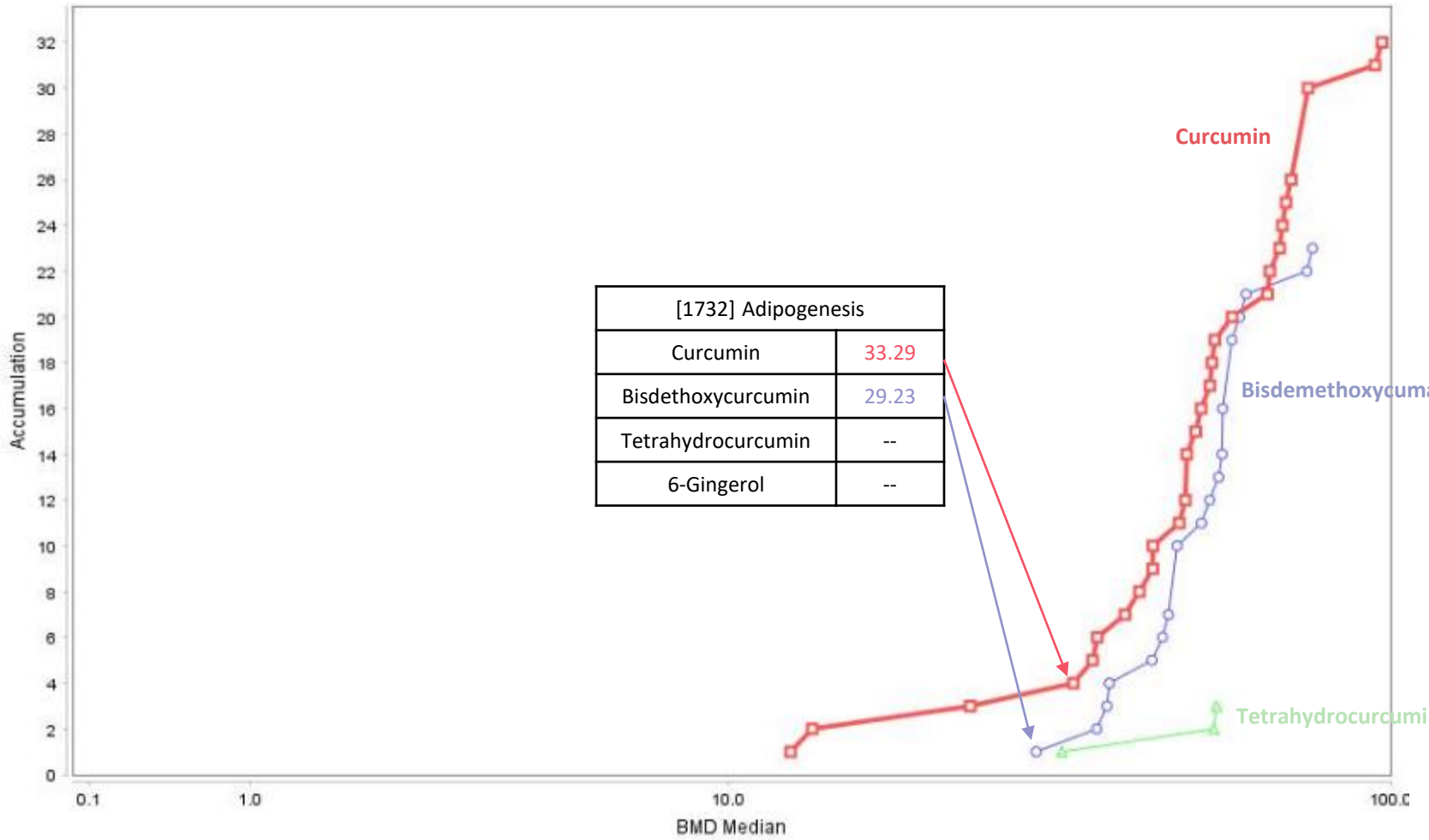
(6)-Gingerol | DTXSID3041035









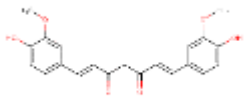


[1732] Adipogenesis	
Curcumin	33.29
Bisdethoxycurcumin	29.23
Tetrahydrocurcumin	--
6-Gingerol	--

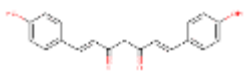
Toxicol Appl Pharmacol. 2017 Aug 15;329:158-164. doi: 10.1016/j.taap.2017.05.036.
Curcumin inhibits adipogenesis induced by benzyl butyl phthalate in 3T3-L1 cells.
Sakuma S[1], Sumida M[2], Endoh Y[2], Kurita A[2], Yamaguchi A[2], Watanabe T[2], Kohda T[2], Tsukiyama Y[2], Fujimoto Y[3].

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].

Gene Set Analysis Using BMD Modeling Results

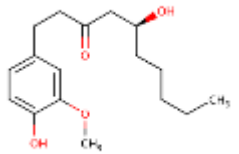


Curcumin | DTXSID8031077

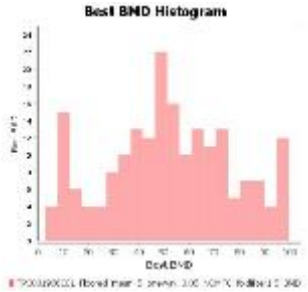


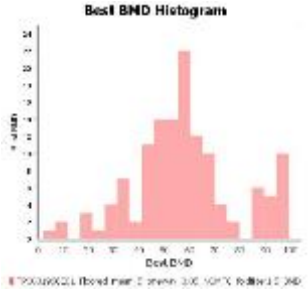
Bisdemethoxycurcumin | DTXSID00872663

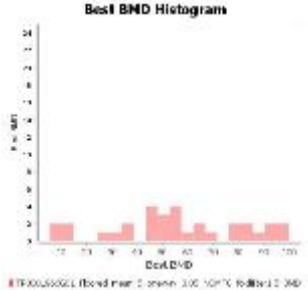
Tetrahydrocurcumin | DTXSID30865801

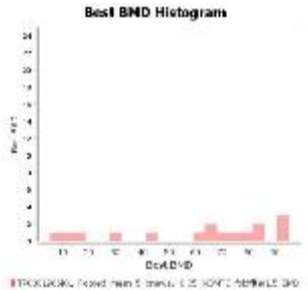


(6)-Gingerol | DTXSID3041035



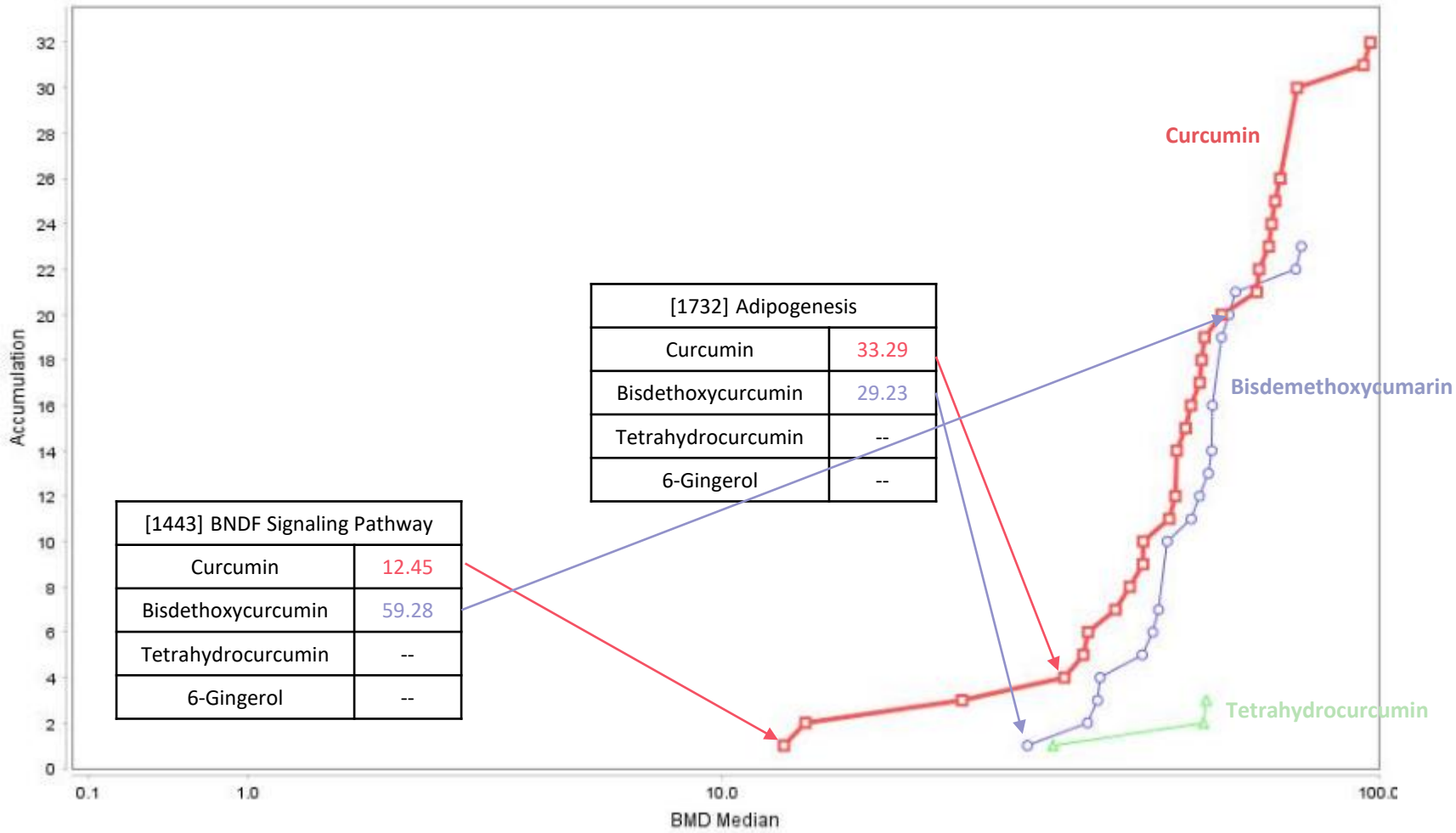






[1443] BNDF Signaling Pathway	
Curcumin	12.45
Bisdethoxycurcumin	59.28
Tetrahydrocurcumin	--
6-Gingerol	--

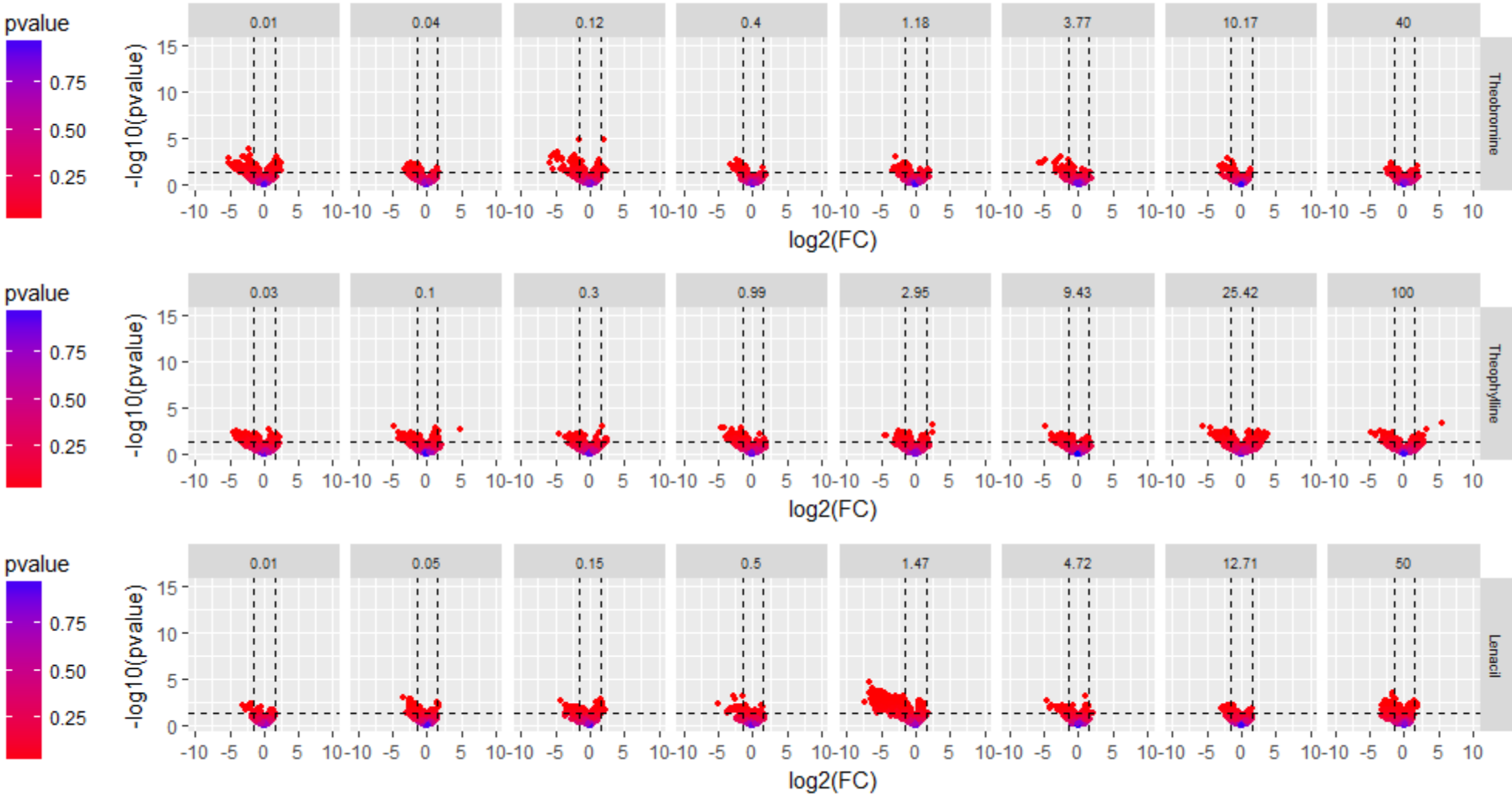
[1732] Adipogenesis	
Curcumin	33.29
Bisdethoxycurcumin	29.23
Tetrahydrocurcumin	--
6-Gingerol	--

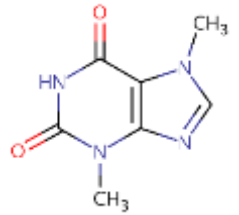


Neuropeptides. 2016 Apr;56:25-31. doi: 10.1016/j.npep.2015.11.003. Epub 2015 Nov 11.
Effect of curcumin on serum brain-derived neurotrophic factor levels in women with premenstrual syndrome: A randomized, double-blind, placebo-controlled trial. Fanaei H[1], Khayat S[2], Kasaeian A[3], Javadimehr M[4].

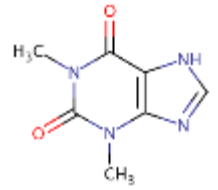
Biomed Pharmacother. 2017 Mar;87:721-740. doi: 10.1016/j.biopha.2016.12.020. Epub 2017 Jan 14.
Curcumin confers neuroprotection against alcohol-induced hippocampal neurodegeneration via CREB-BDNF pathway in rats.
Motaghinejad M[1], Motevalian M[2], Fatima S[3], Hashemi H[1], Gholami M[4].

Caffeine Analogs

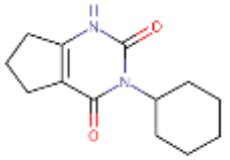




Theobromine | DTXSID9026132



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:

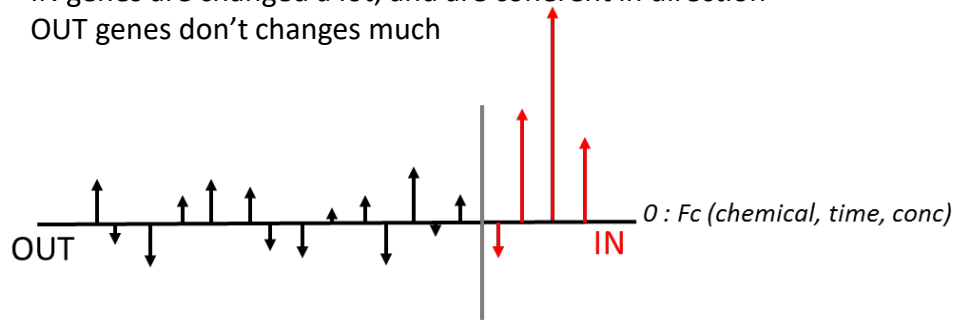
$$M = \sum_{i=1}^{ngene} \frac{\mathrm{fc}_i}{\mathrm{sc}_i^2} \Big/ \sum_{i=1}^{ngene} \frac{1}{\mathrm{sc}_i^2}$$

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

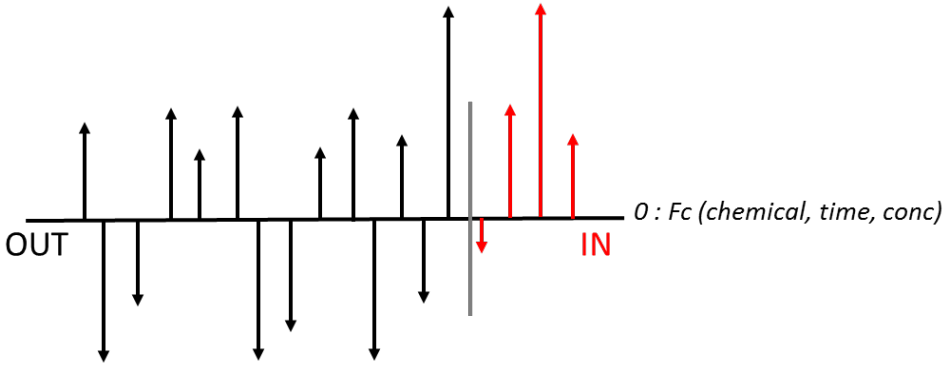
Step 2: CR Modeling

- 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

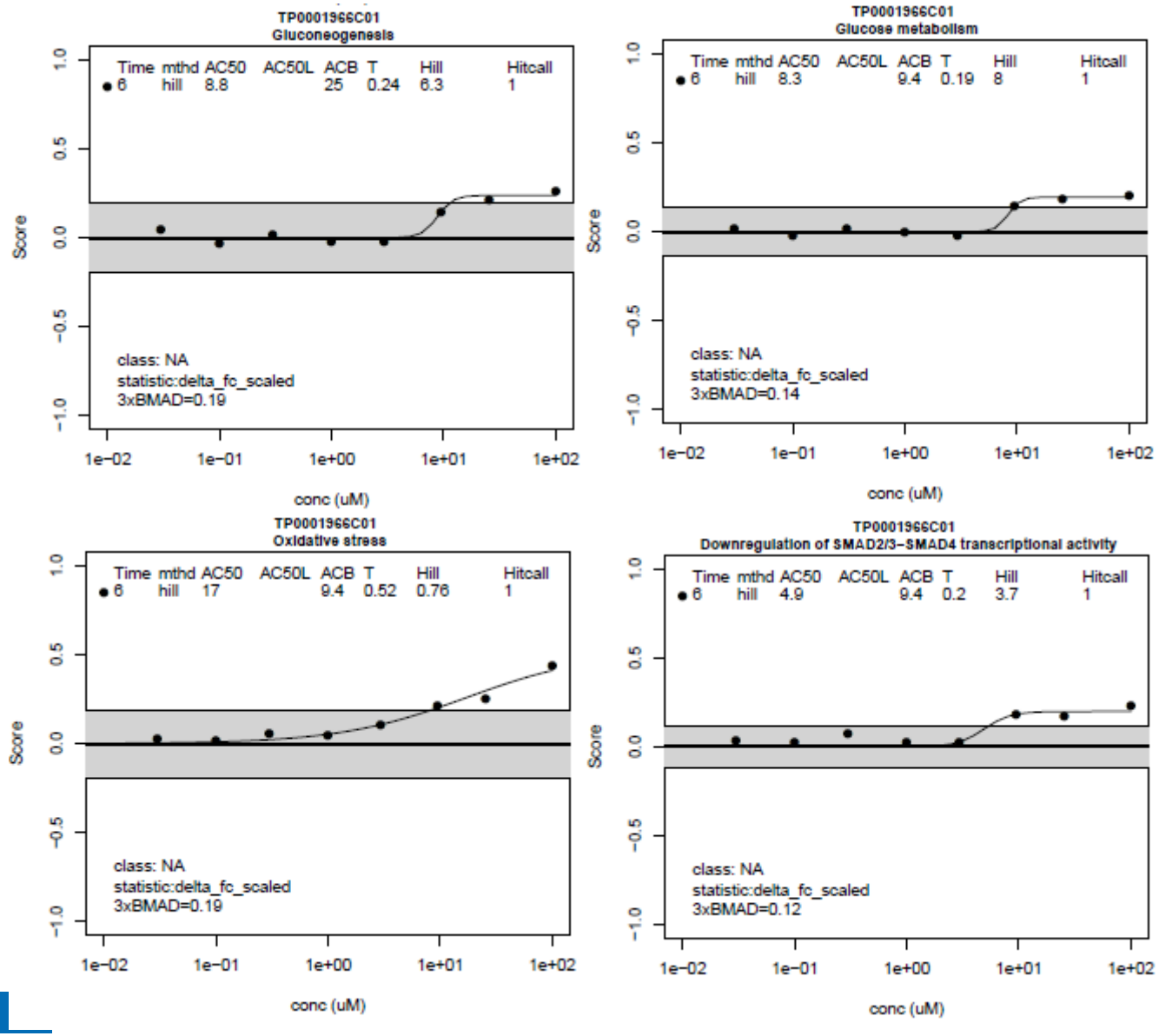
- IN genes are changed a lot, and are coherent in direction
- OUT genes don't changes much

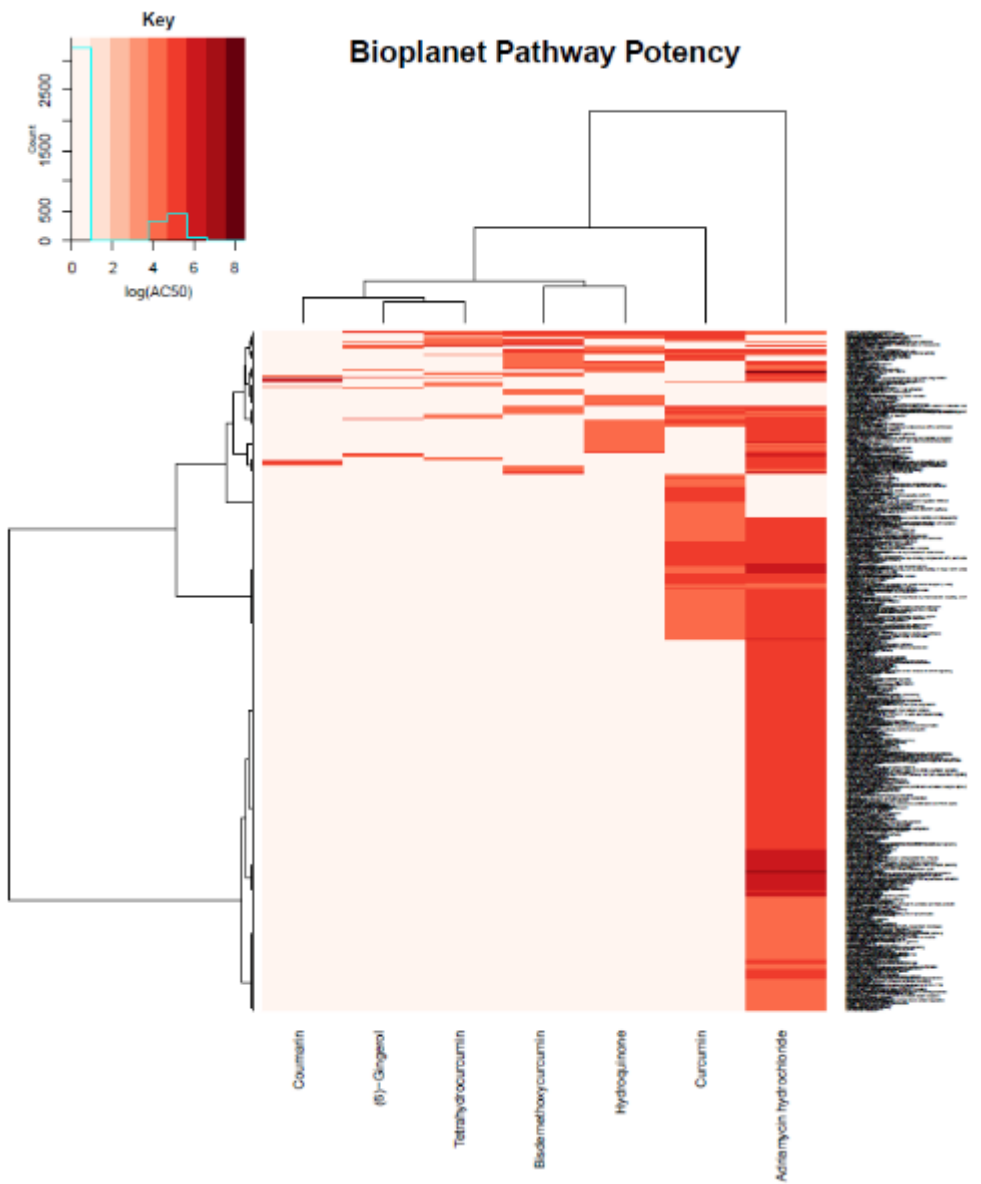


- IN genes are changed a lot, and are coherent in direction
- OUT genes change a lot but are not coherent (mean ~ 0)



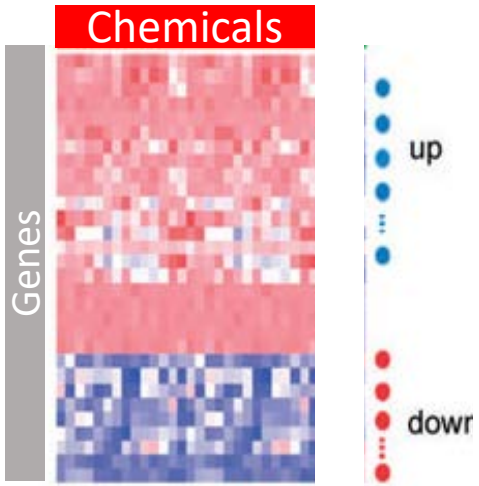
Comparing Activities Across Chemicals



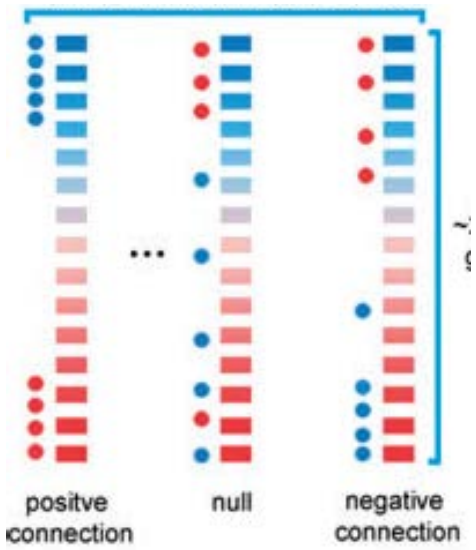


MOA by Connectivity Mapping

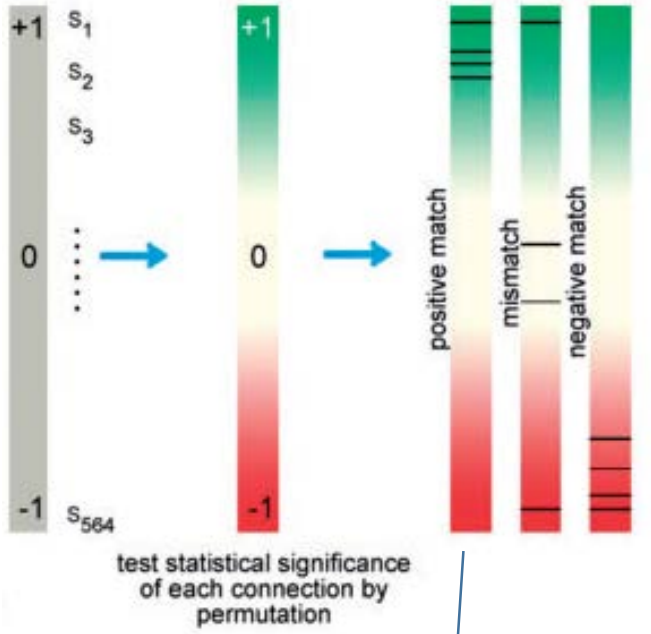
Input Profile



Query Signature DB
CMap or BSP



Find best positive matches



Infer MIE/Target
by best match

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

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 ?

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

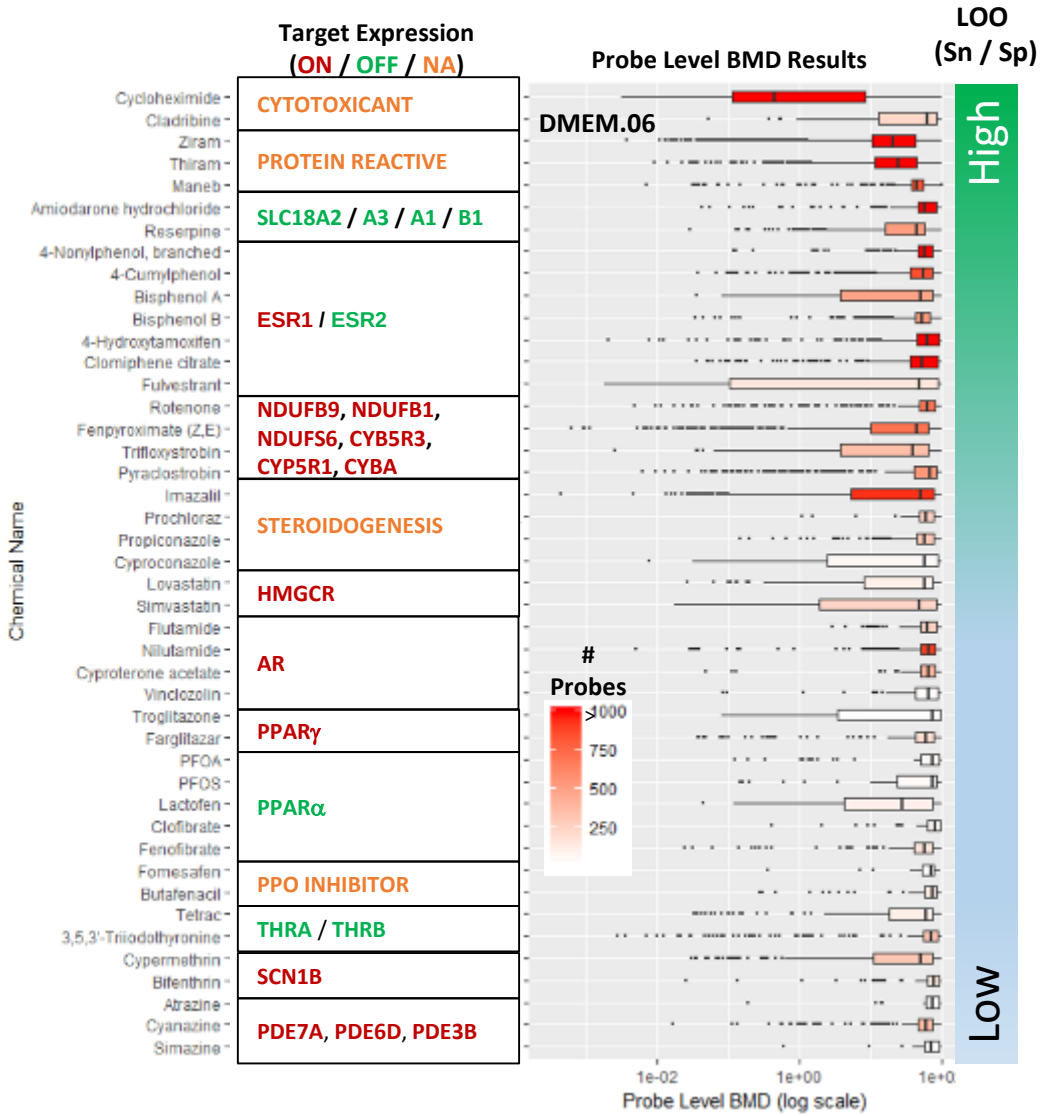
CRGs
- **Query:** HTTr CRGs
- **DB:** HTTr CRGs

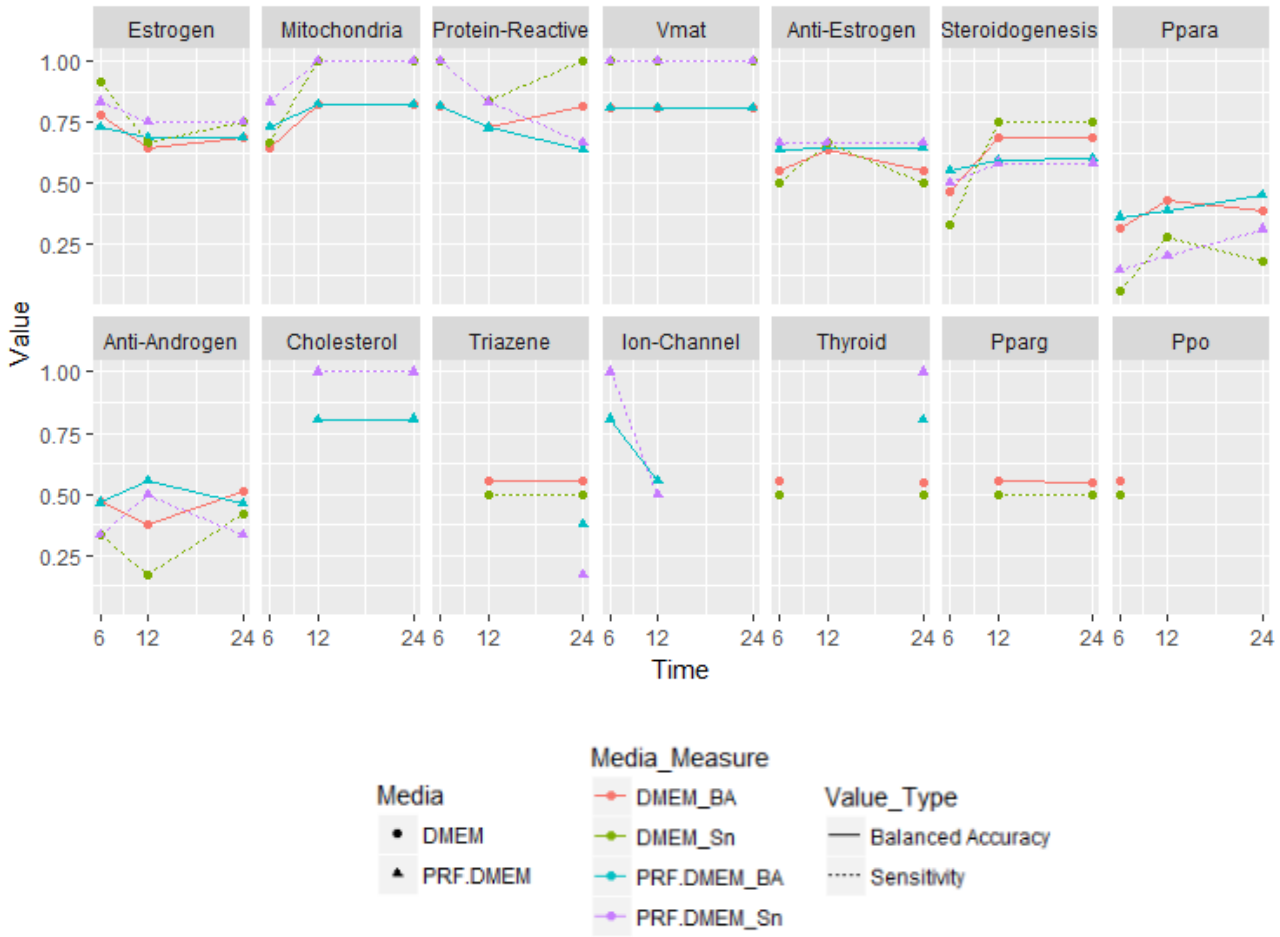
HTTr vs CMap

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

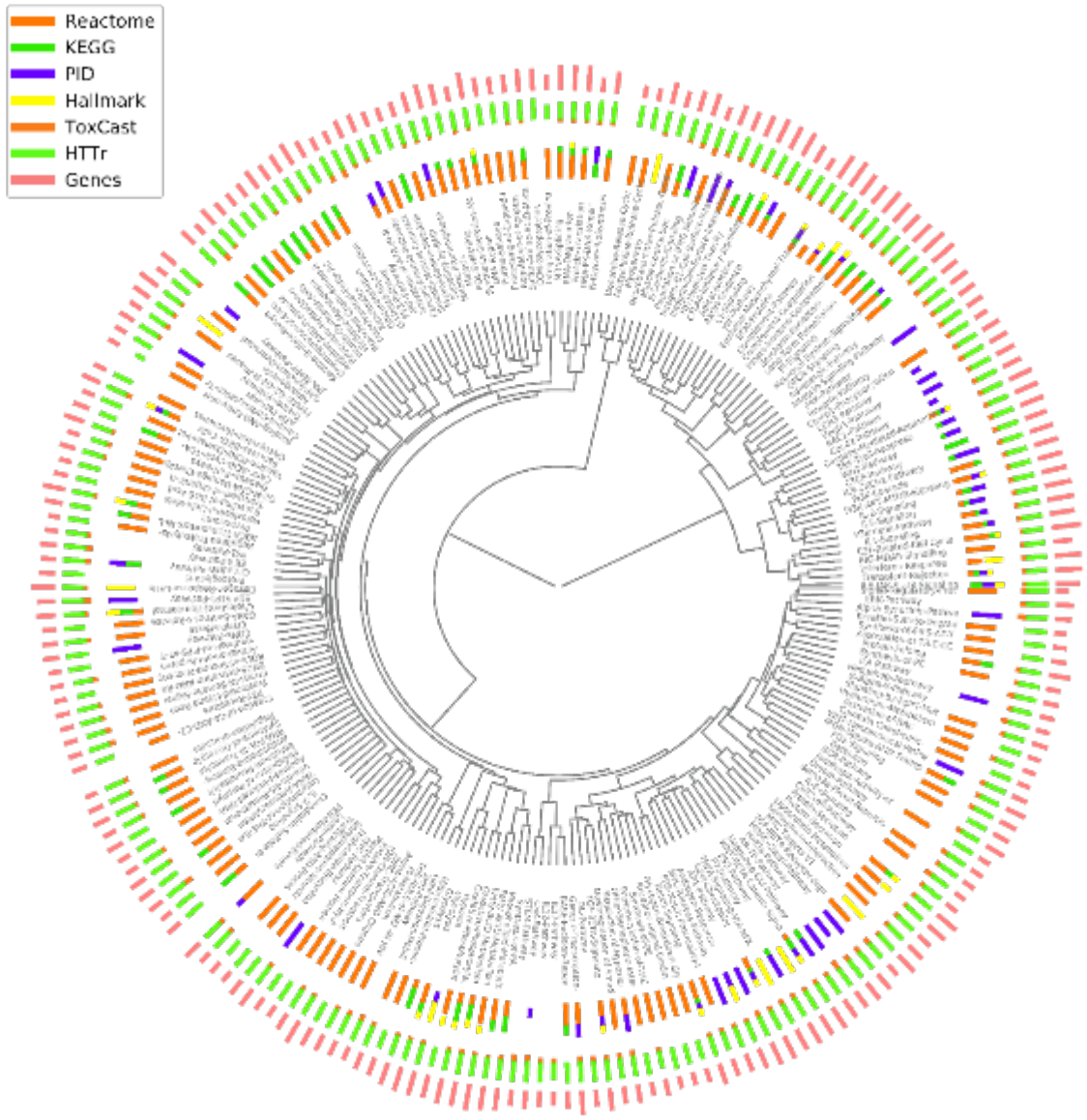
CRGs
- **Query:** TempO-Seq CRGs
- **DB:** Affymetrix CMap

Connectivity Mapping for MoA Prediction

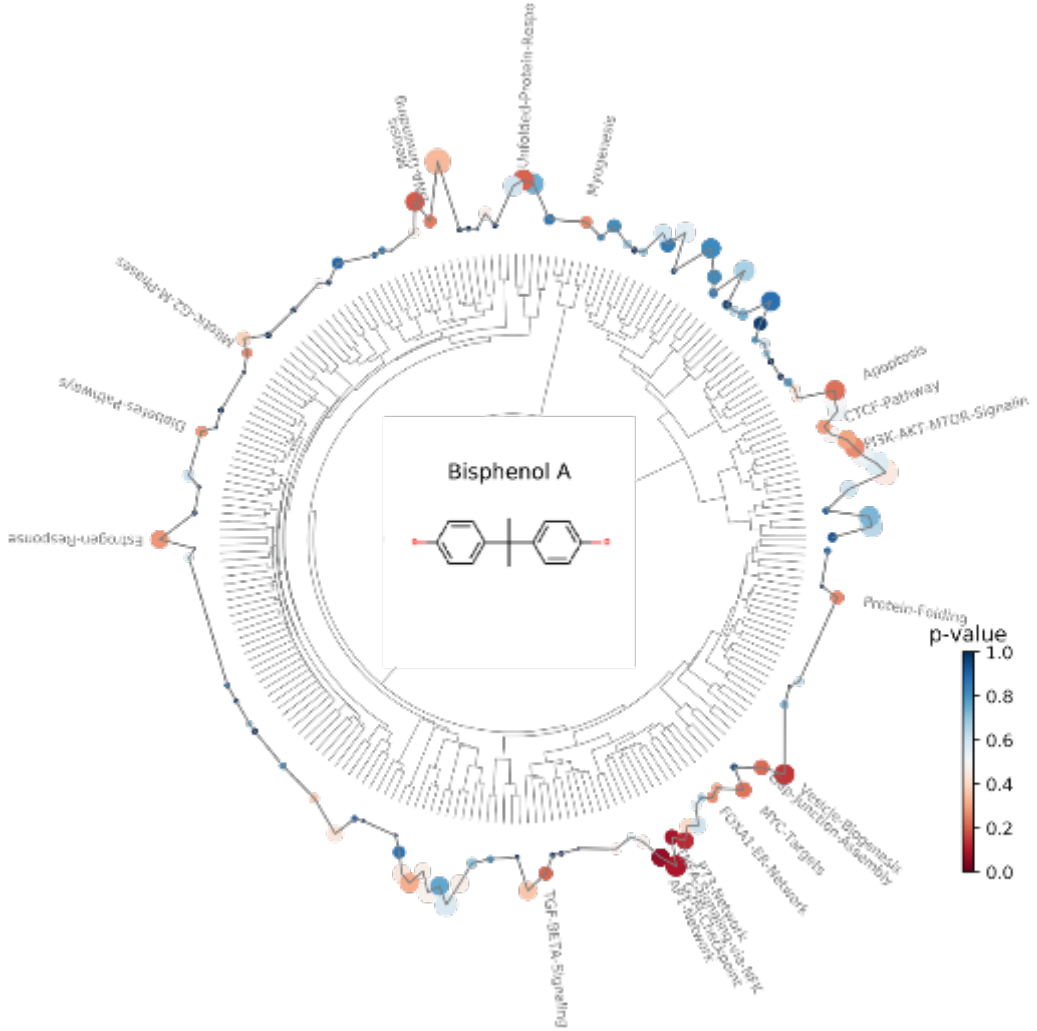




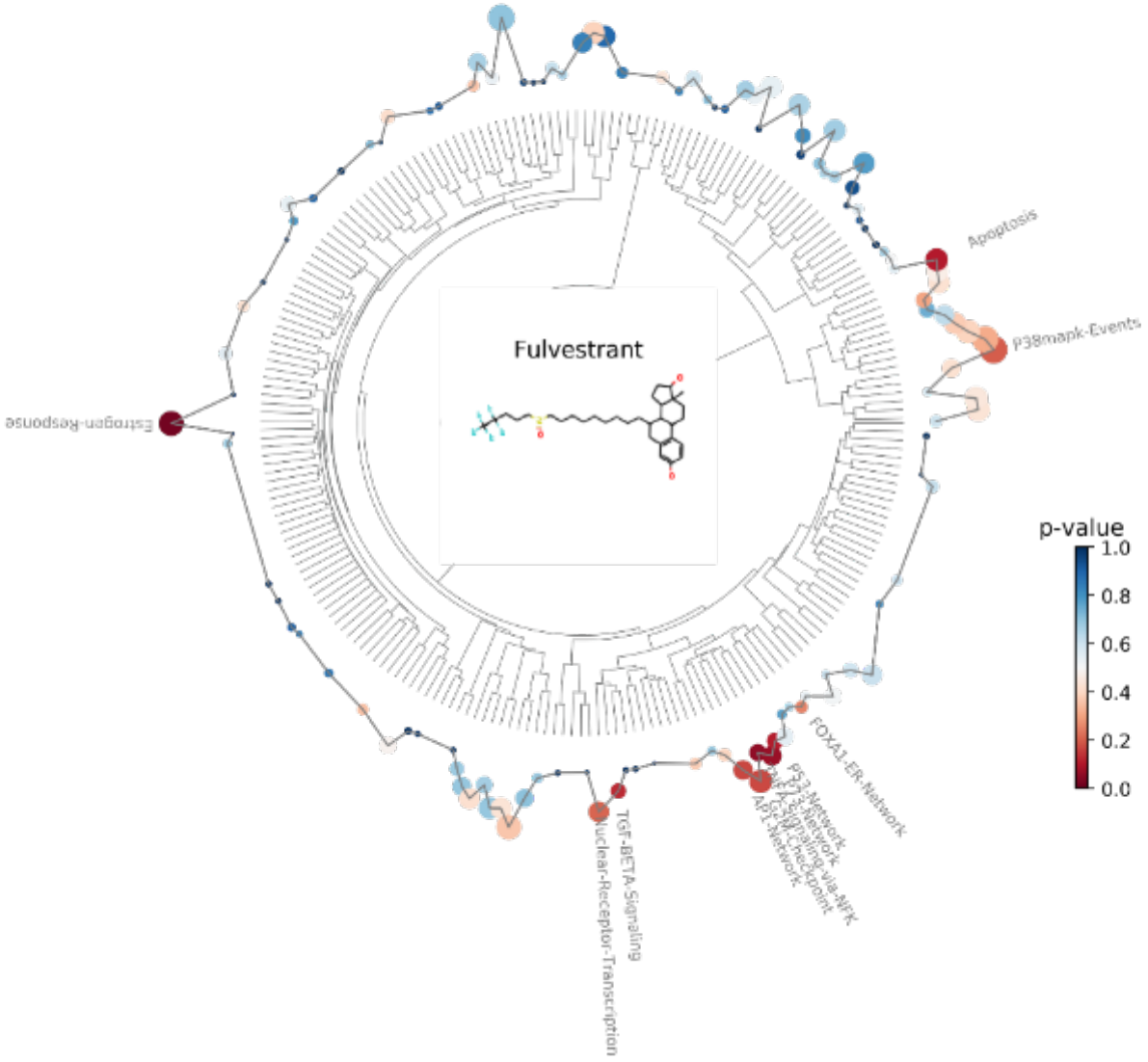
MoA Prediction: Super-Pathway Analysis

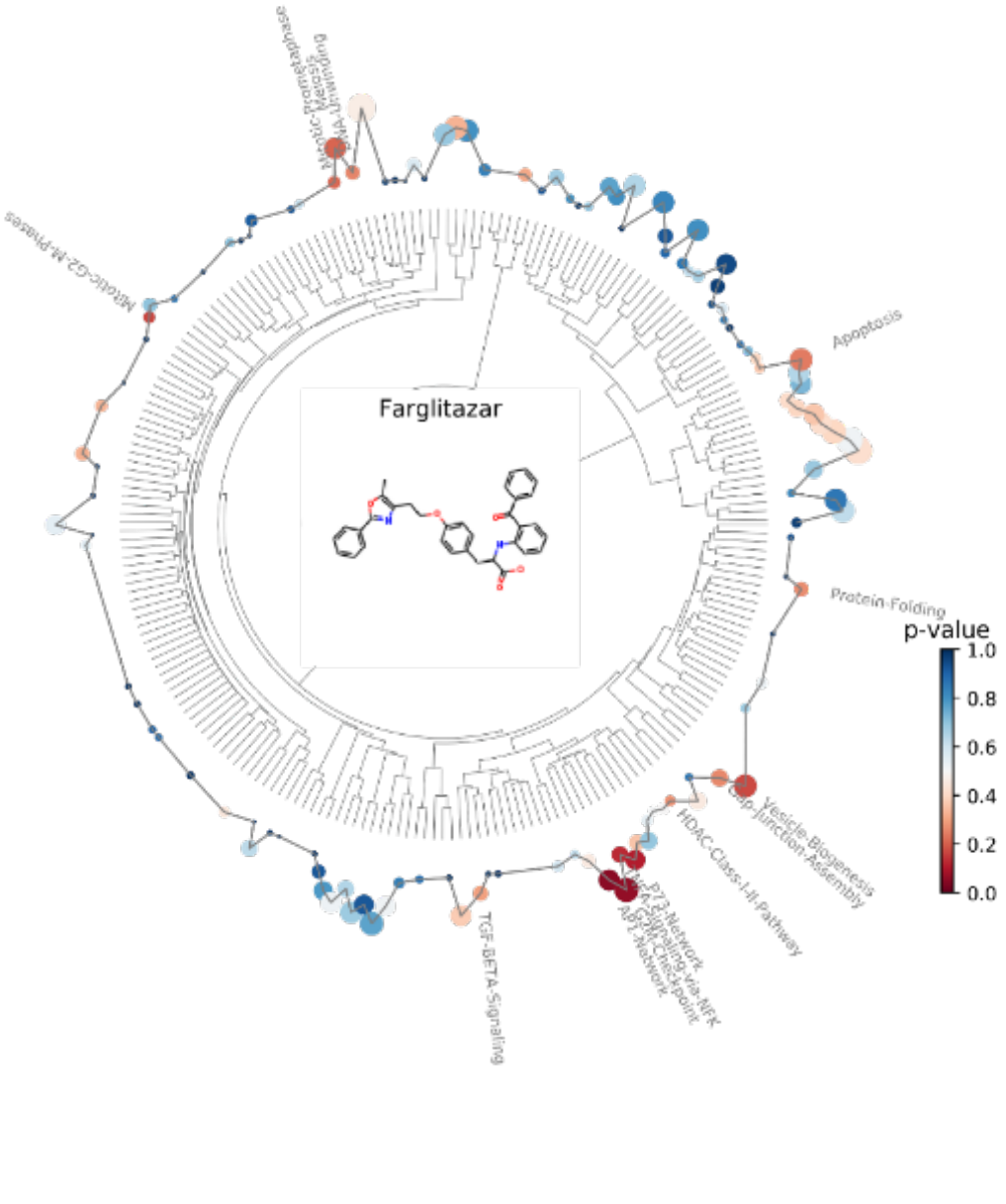


- 224 “Super-Pathways” produced by clustering genes within 1,210 hallmark and canonical pathways from MSigDB V6 (Reactome, KEGG, PID and BioCarta).
- Visualized as radial hierarchical cluster with Super-Pathways as leaves



Super-Pathway Example



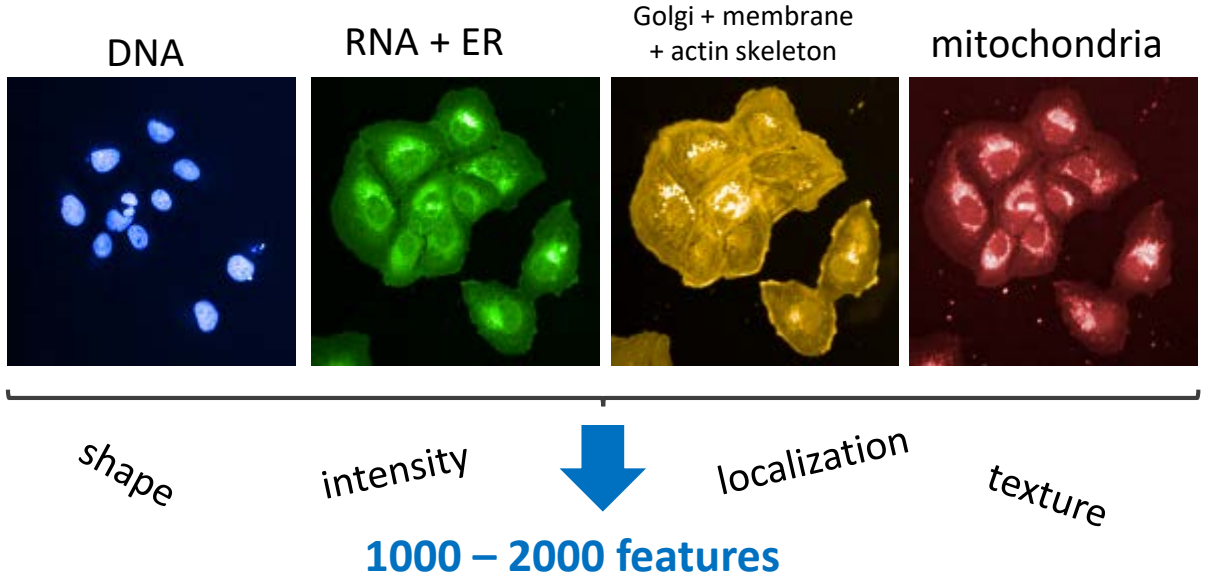


EPA
United States
Environmental Protection
Agency

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

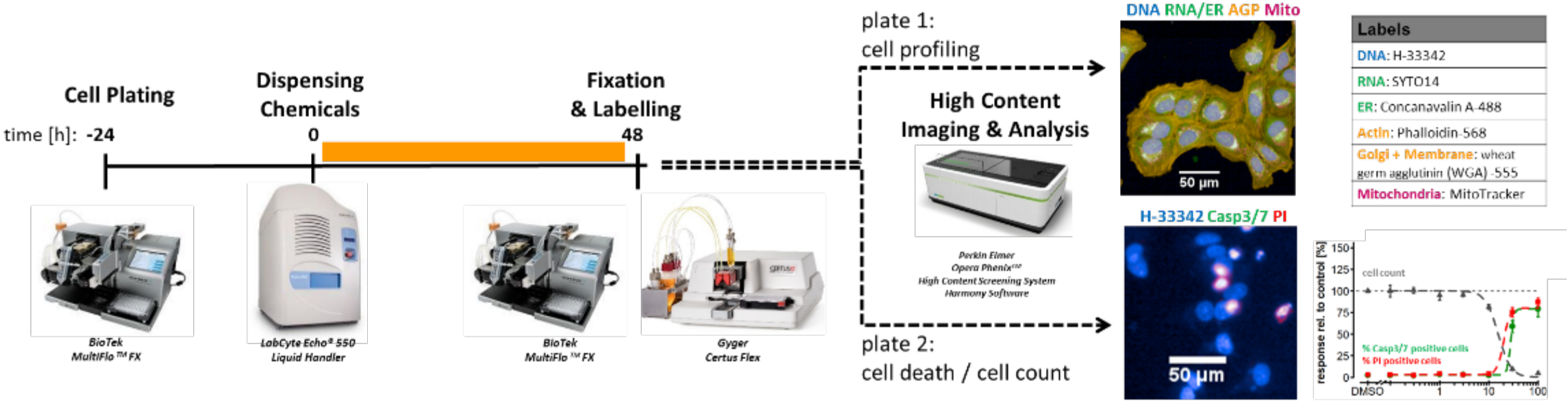
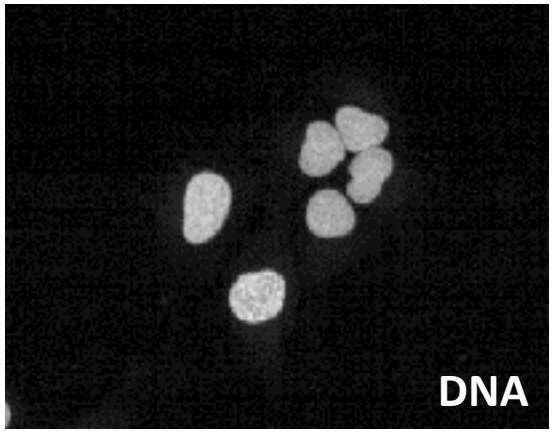
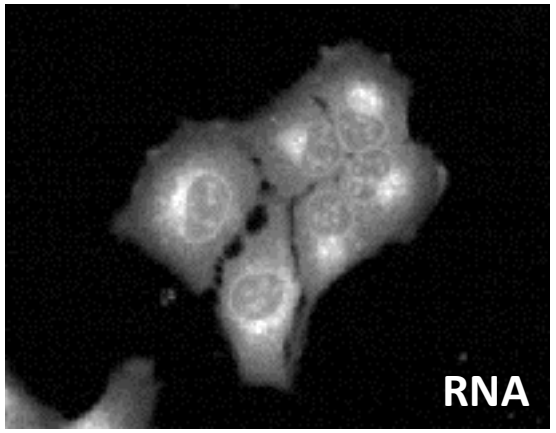


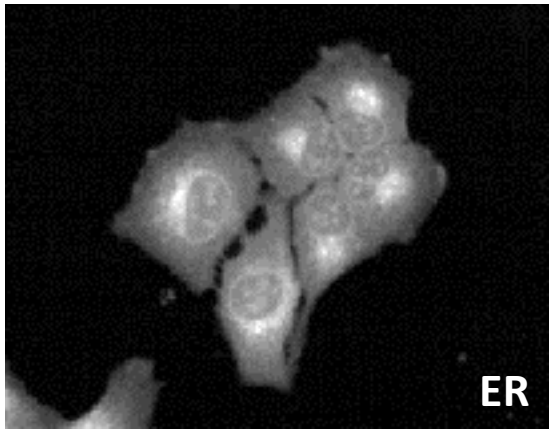
Image Analysis Workflow
Image Acquisition

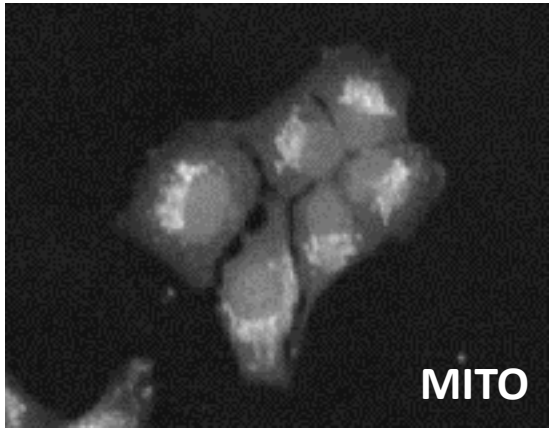
Image Acquisition

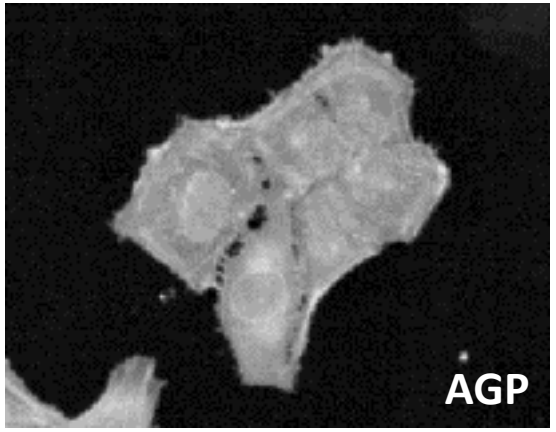
- Perkin Elmer Opera Phenix
- 20x Water Immersion Objective
- Confocal Mode, Single Z
- CellCarrier-384 Ultra Microplates











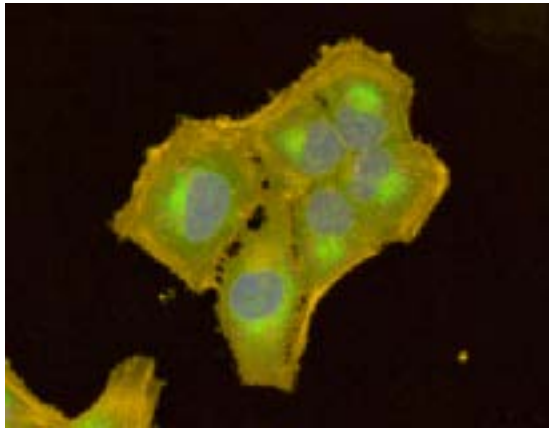


Image Analysis Workflow:
Nucleus and Cell Segmentation

1. find nuclei

2. find cell outline

3. reject border objects

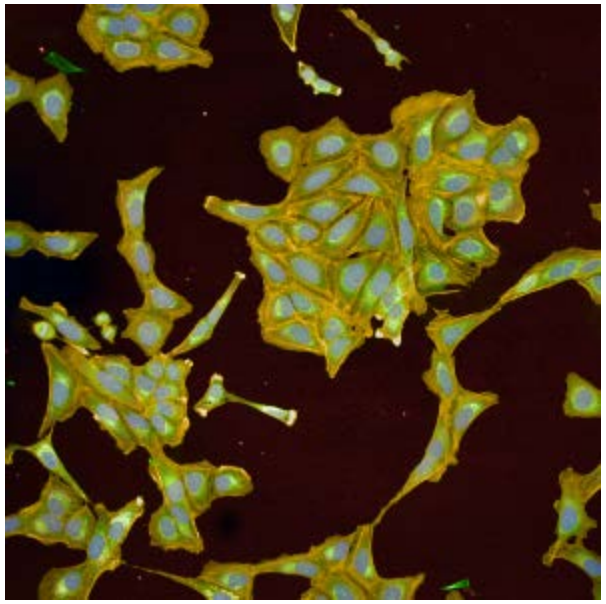
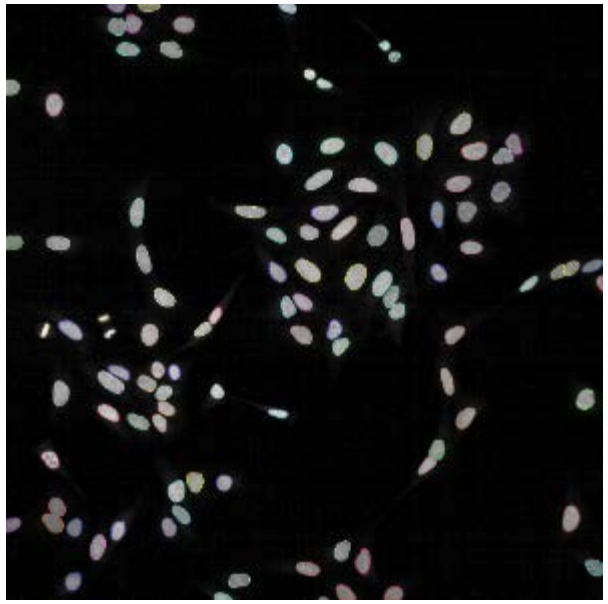
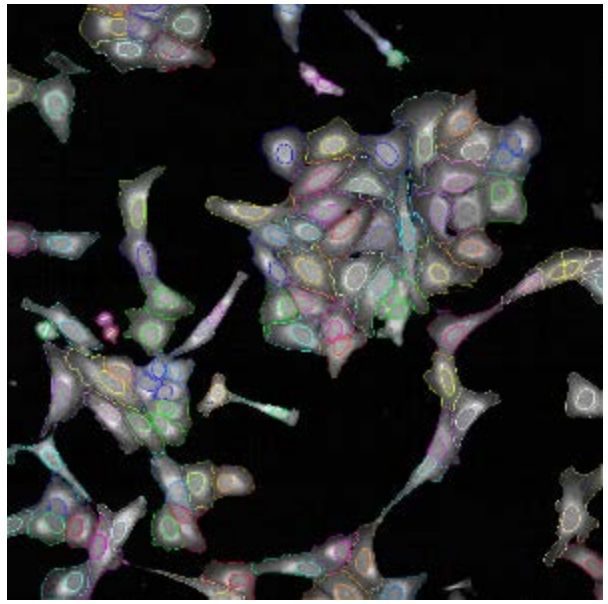
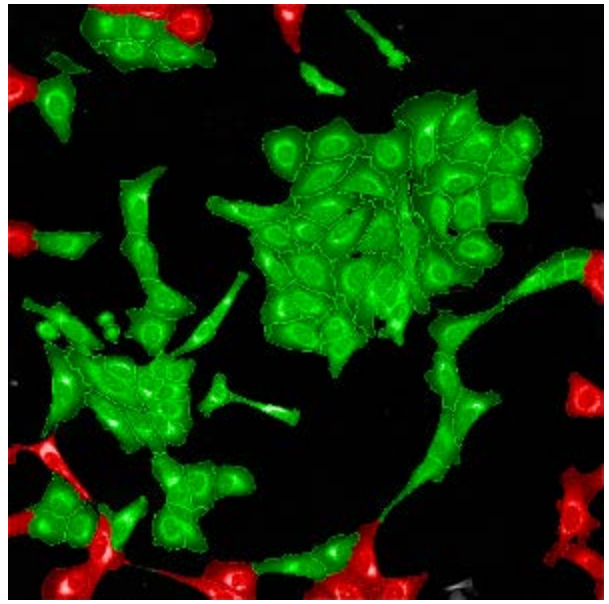
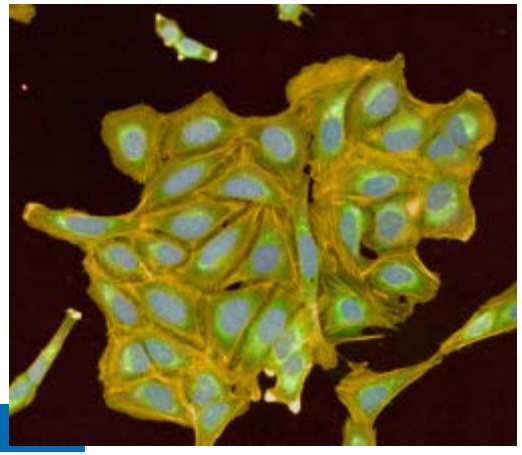
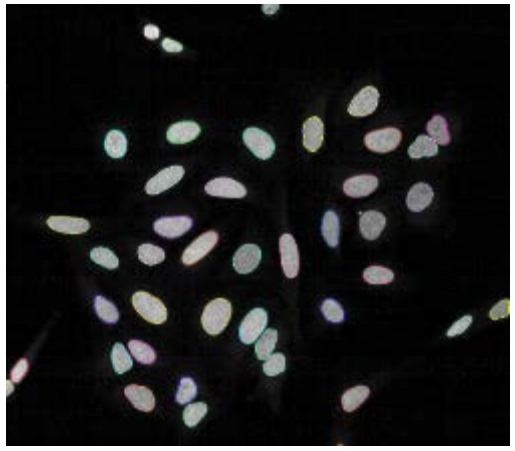
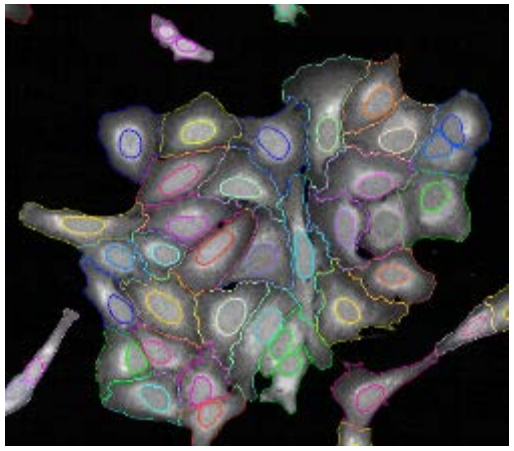
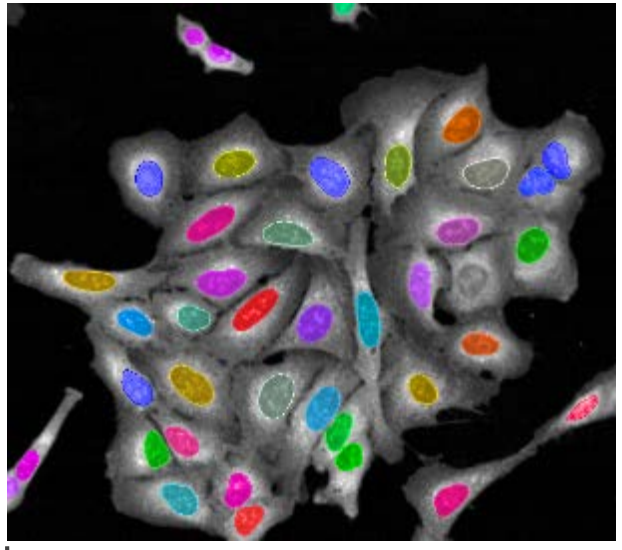








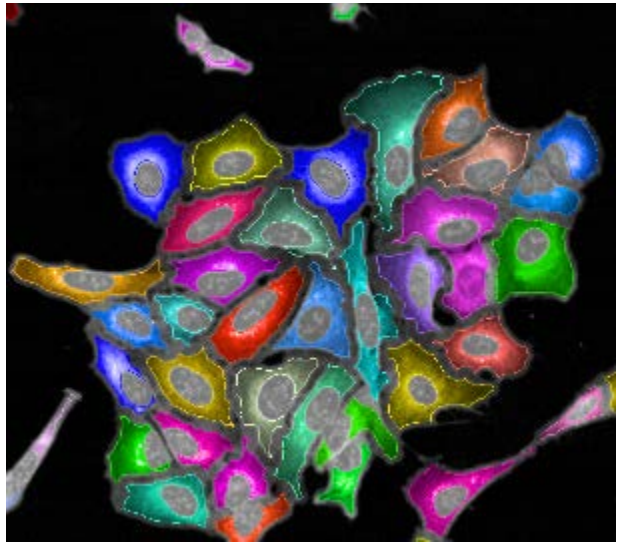
Image Analysis Workflow

Define Cellular Compartments

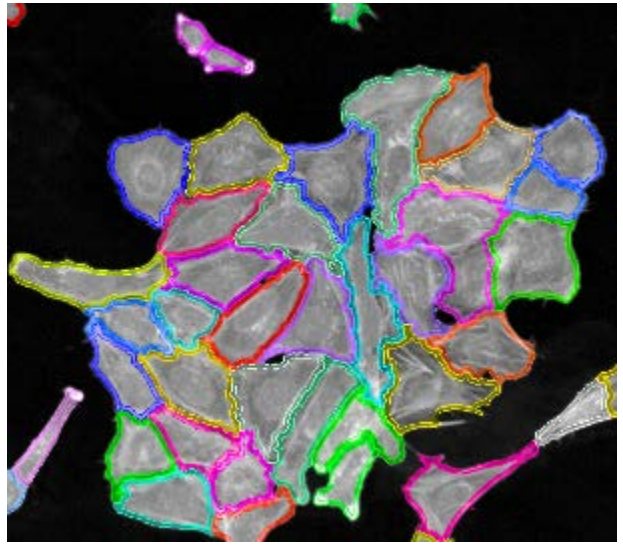
nuclei



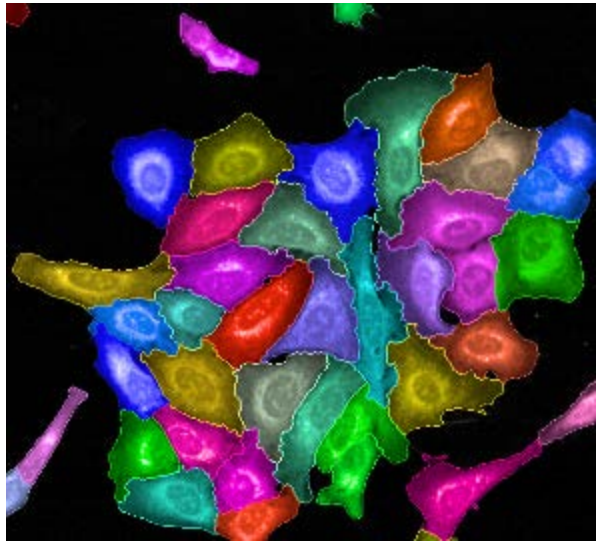
cytoplasm



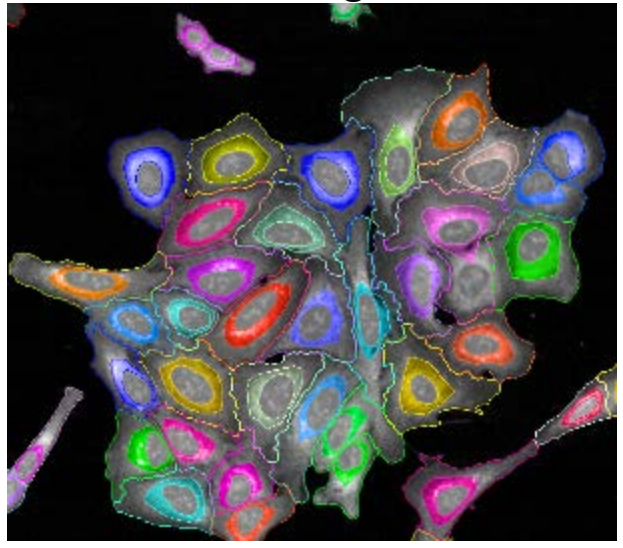
membrane



cell



ring



Assay Outputs & Data Analysis

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

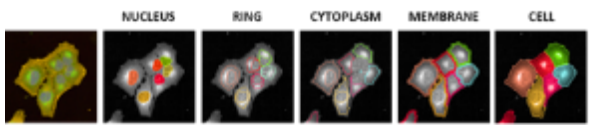
X

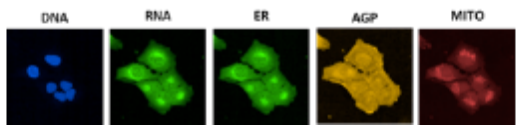
channels (organelles):
- DNA
- RNA
- ER
- AGP (actin skeleton / Golgi/ plasma membrane)
- mitochondria

X

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

= 1700 endpoints



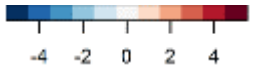


Data Reduction

cell-level data (~500 cells/well)

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

normalized cell-level data



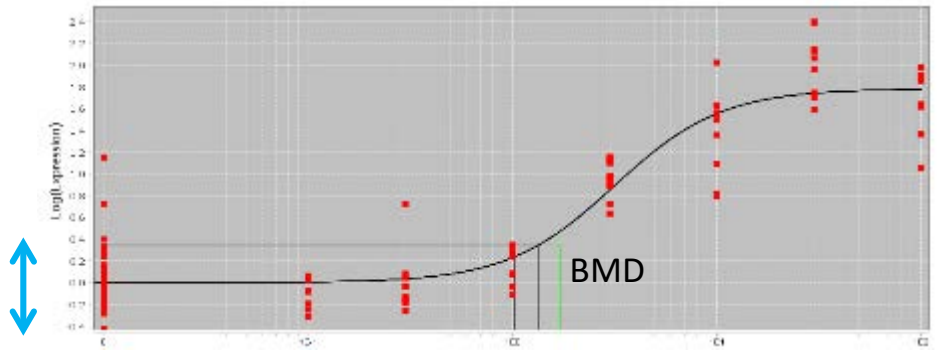
median

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

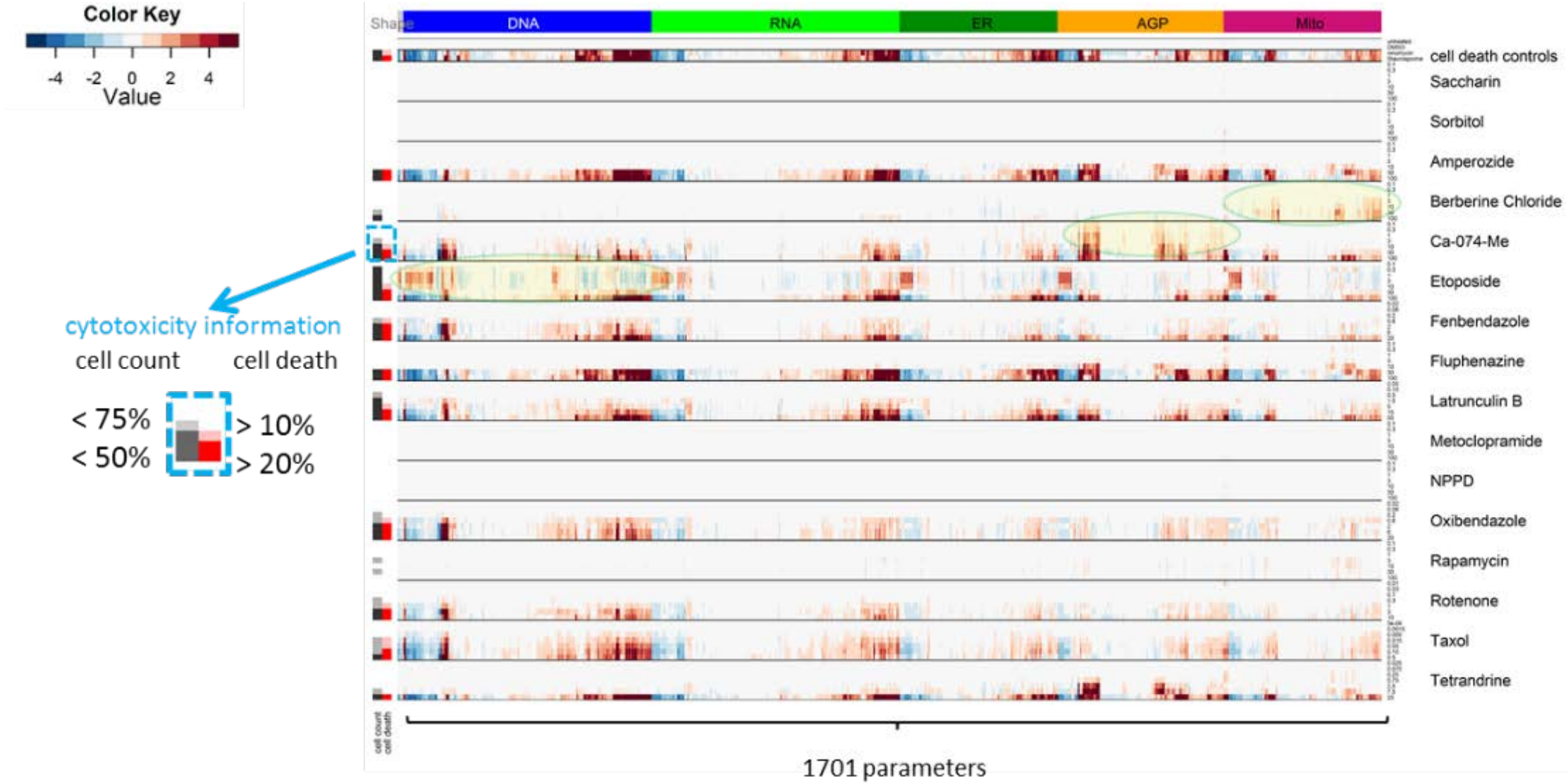


Fig 1.: MAD normalized well-level data of U-2 OS cells were averaged across 3 technical and 3 biological replicates. Endpoints are ordered according to the corresponding channel/organelle. The color key on the left indicates reductions in cell count and increases in cell death. Treatment with different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

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)

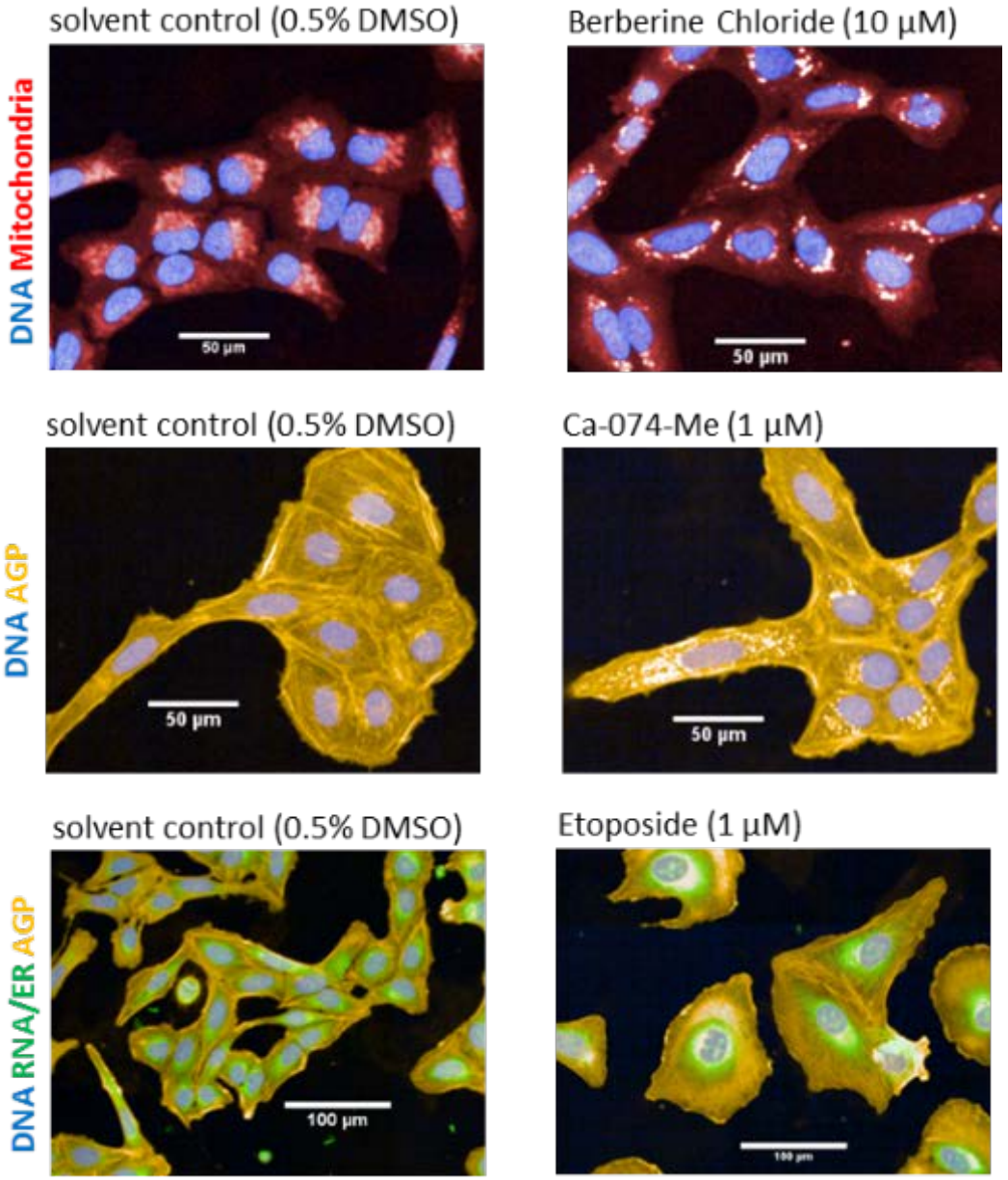
Literature: redistribution of mitochondria

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

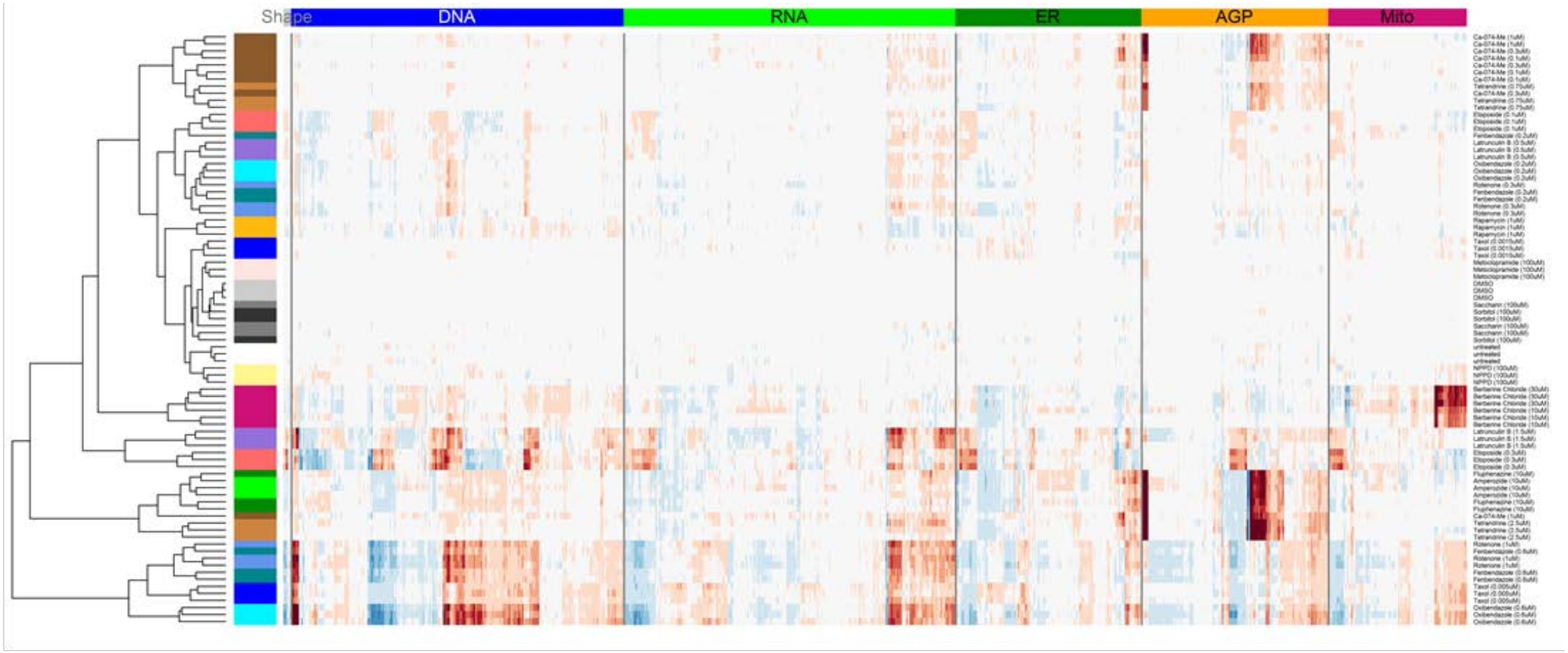
Literature: bright, abundant Golgi stain

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



Reproducibility Across Biological Replicates



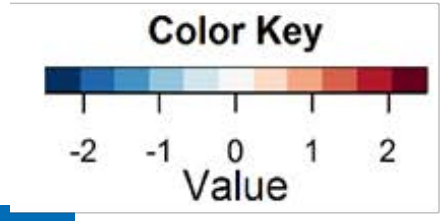


Fig 2.: MAD normalized well-level data of U-2 OS cells were averaged across 3 technical replicates. Each row represents a biological replicate of selected conditions. Conditions were filtered for effects on cell morphology in the absence of pronounced cytotoxicity. Endpoints are ordered according to the corresponding channel/organelle. Only robust endpoints are shown (e.g. their standard deviation in all DMSO control wells was < 0.25). This was the case for 1527/1700 parameters. Biological replicates have similar profiles and cluster together.

Chemical Profiles Across Two Cell Types

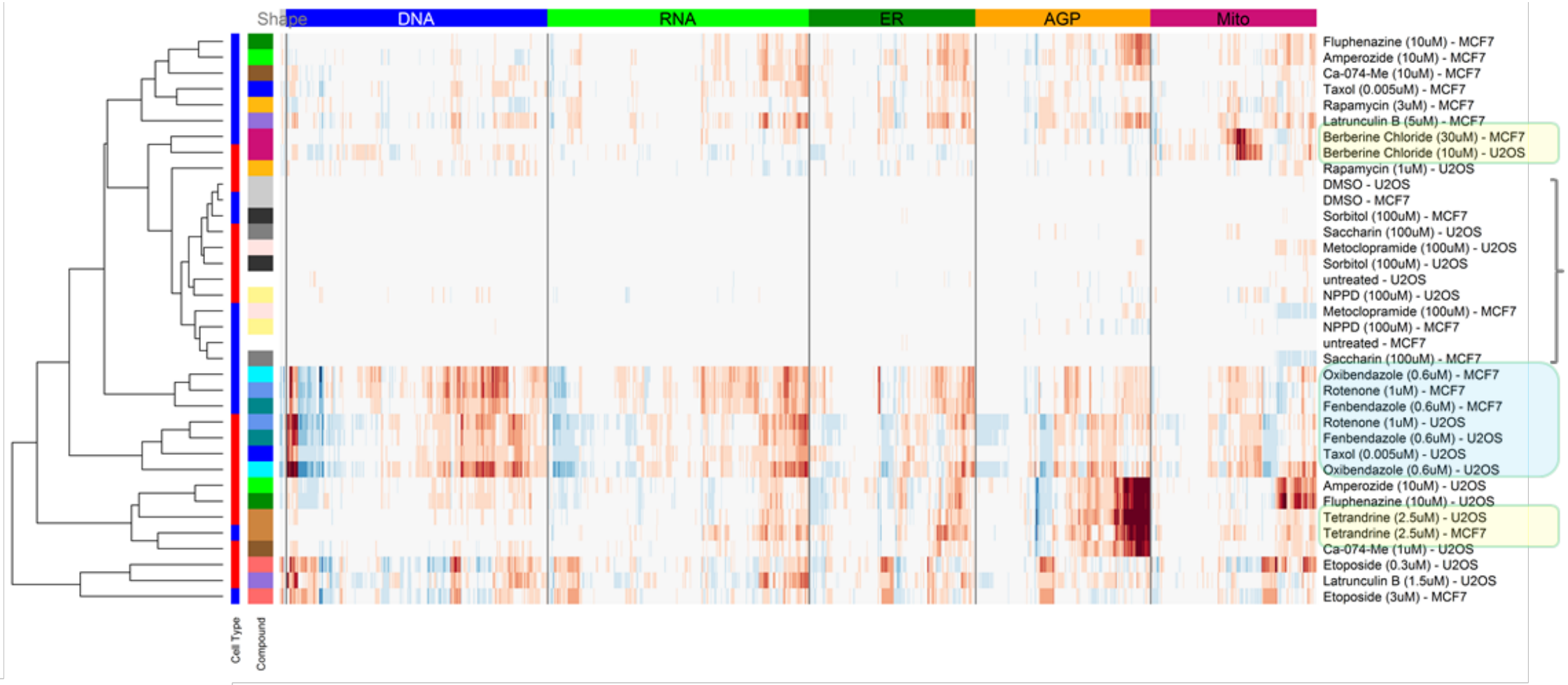


Fig 3.: MAD normalized well-level data of U-2 OS or MCF7 cells were averaged across 3 biological replicates. Each row represents a biological replicate of selected conditions. Conditions were filtered for effects on cell morphology in the absence of pronounced cytotoxicity. Endpoints are ordered according to the corresponding channel/organelle. Treatments with strong effects cluster together across cell types. Profiles of related chemicals are more similar within cell types than across cell types.

Determination of In Vitro Points of Departure (POD)

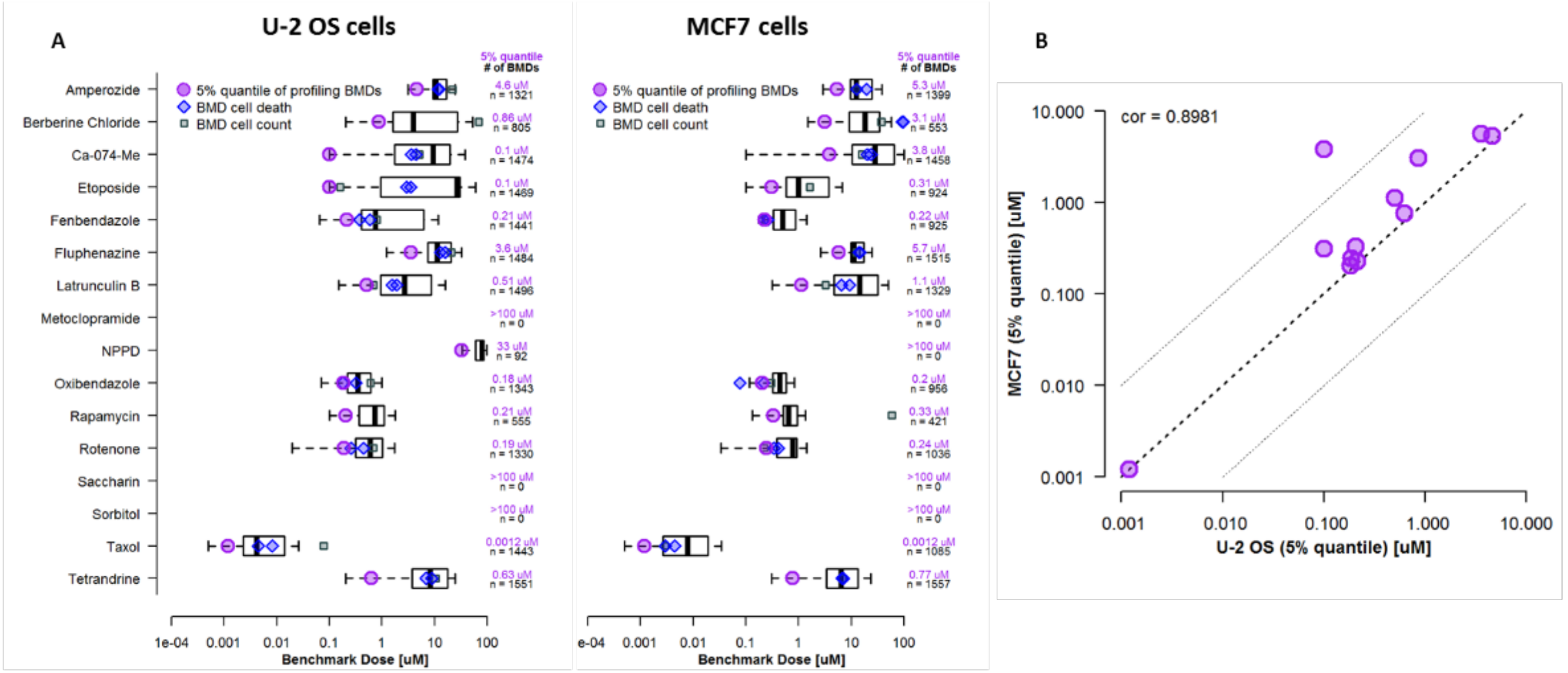
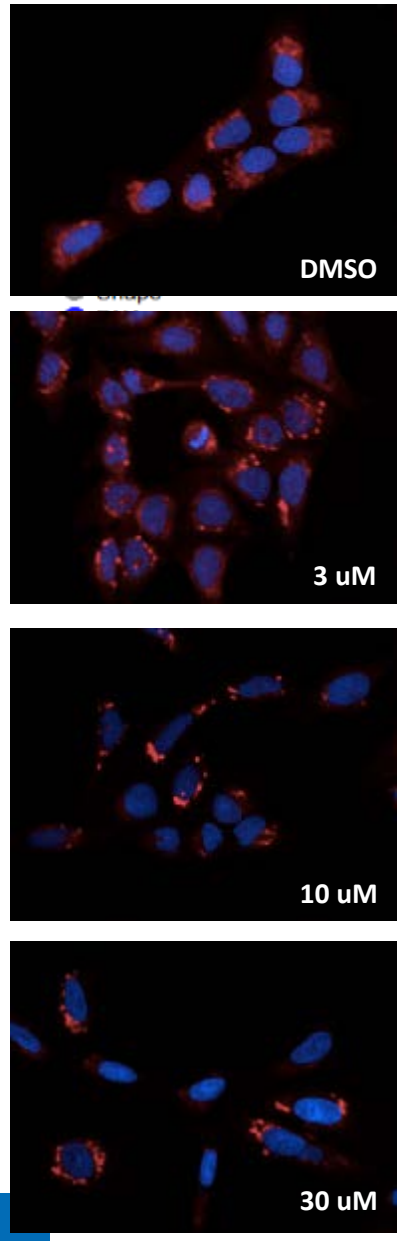
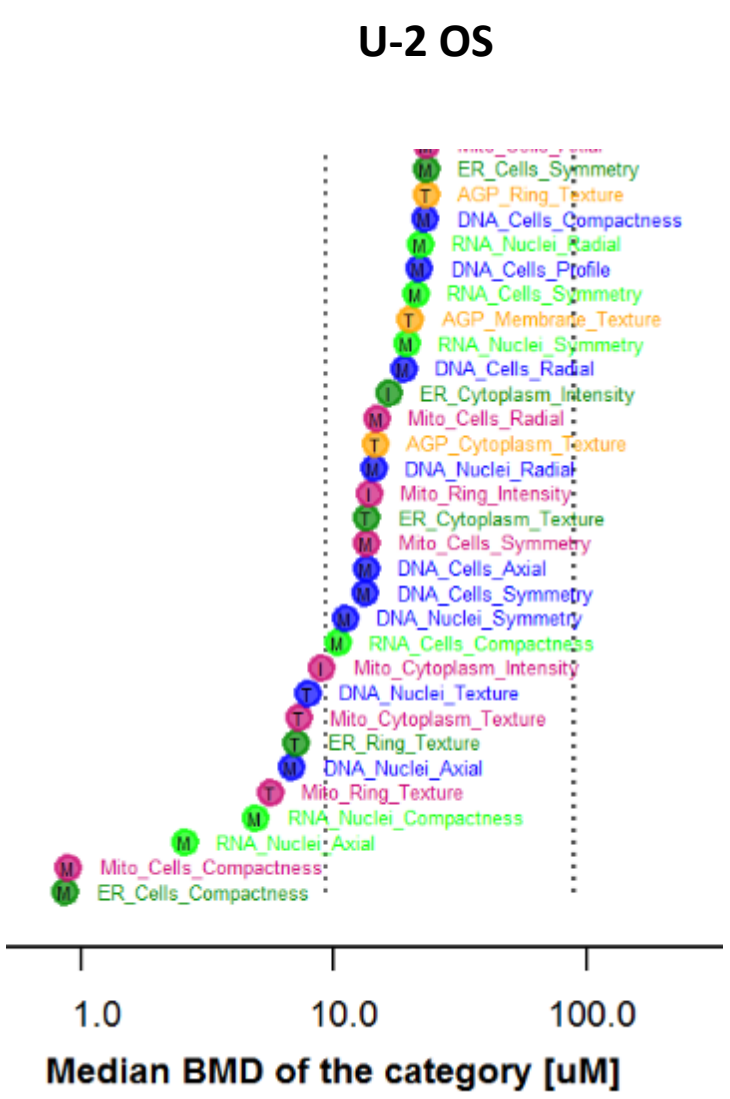
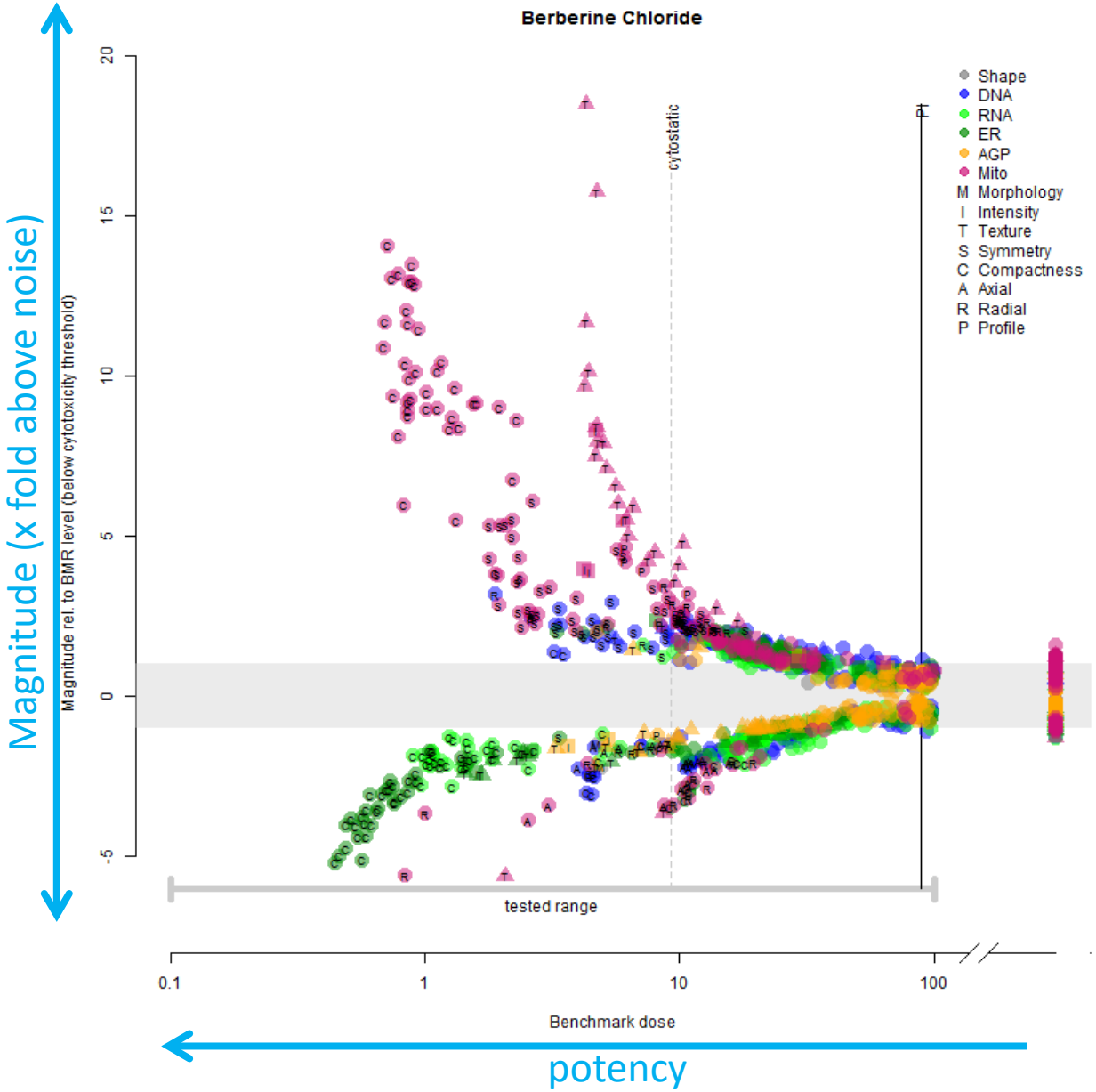


Fig 4.: MAD normalized well-level data were pooled from 3 independent experiments (9 values) to model BMDs. **(A)** The boxplot displays the range of the estimated BMDs from all parameters that were changed. The black line indicates the median: whiskers are at an interquartile range of 1. The 5% quartile of this distribution is considered the point-of-departure (POD) and is indicated in violet. BMDs derived from cytotoxicity and cell count measurement are indicated in blue and green for comparison. **(B)** Comparison of the PODs of the 12 active chemicals across both tested cell types.

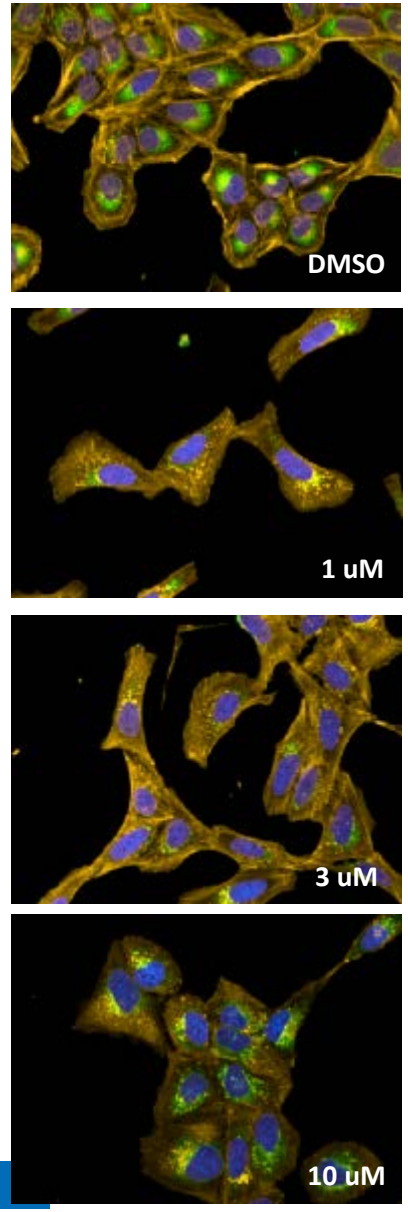
Phenotype Ontogeny Analysis

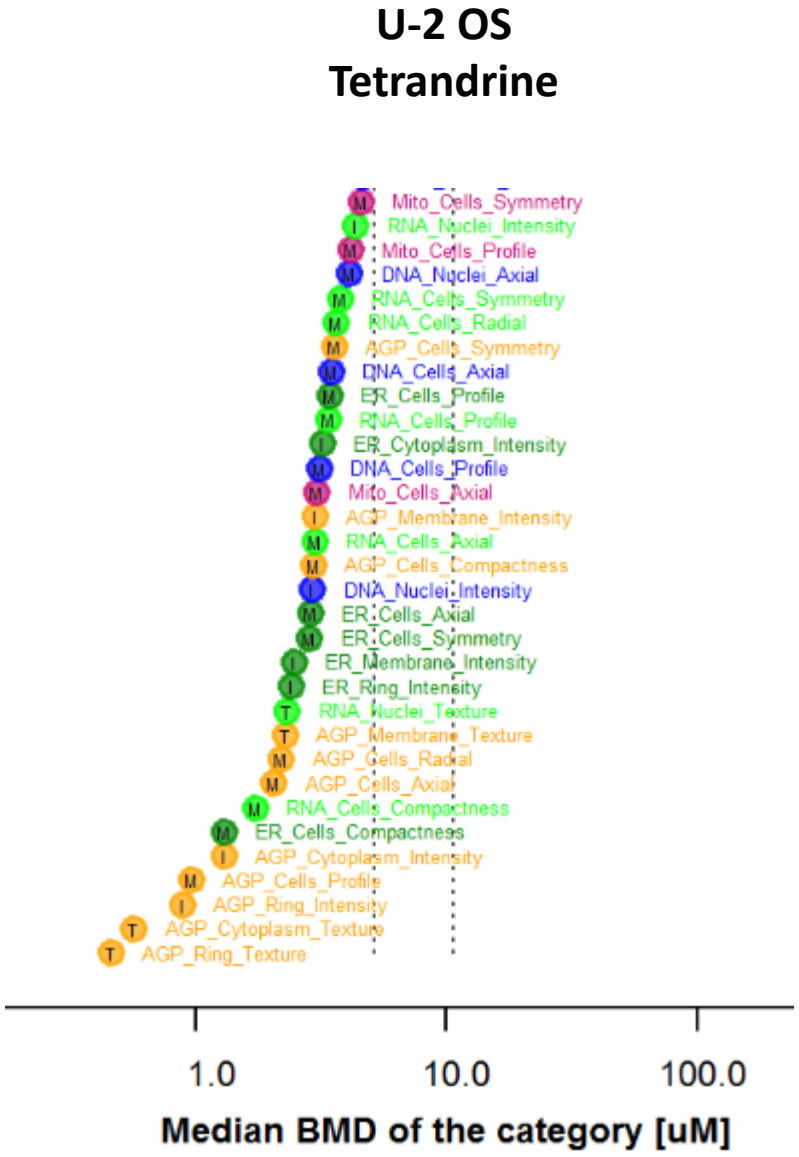


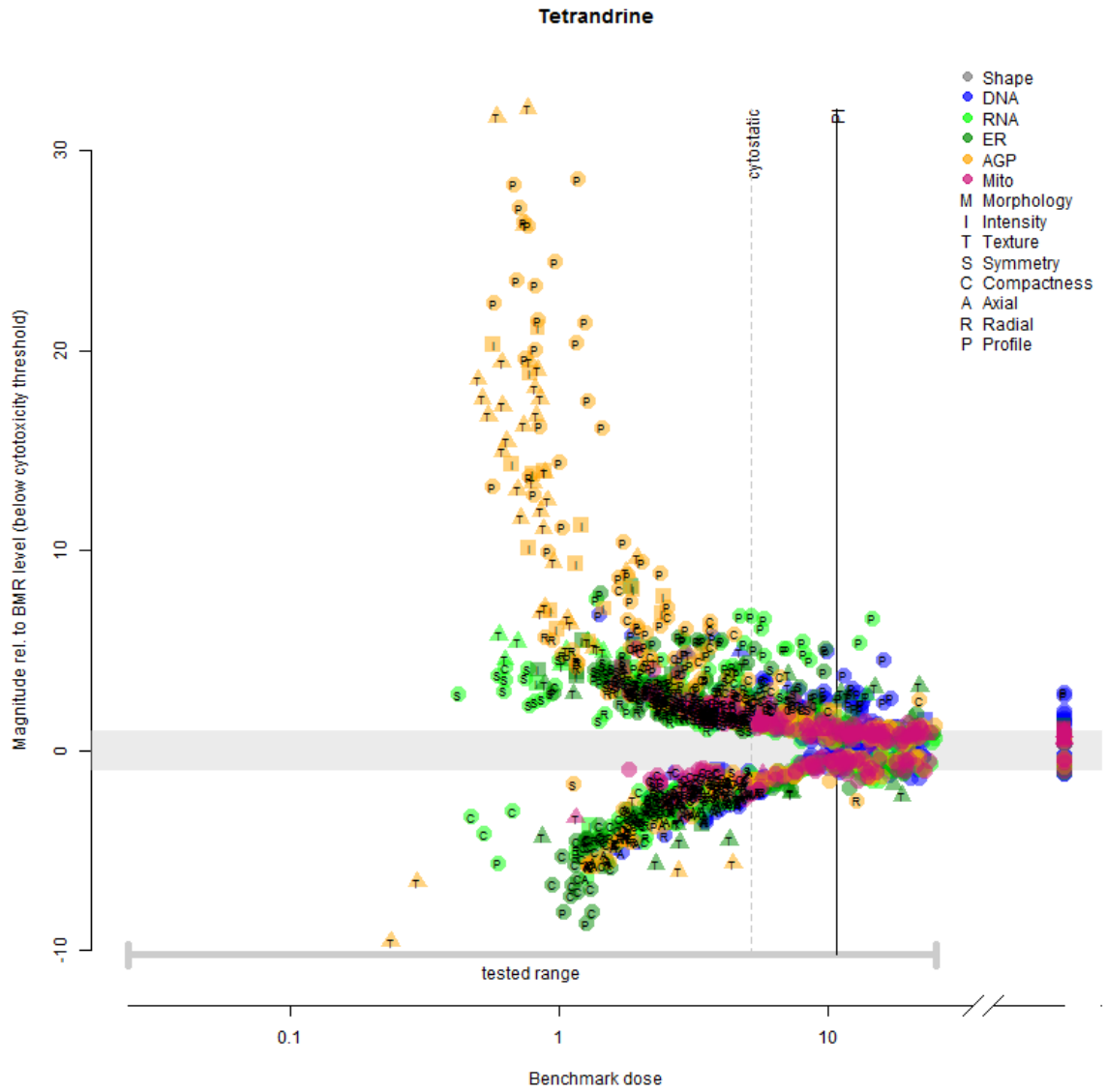




Phenotype Ontogeny Analysis







United States Environmental Protection Agency

Cell Line Selection
Computational Modeling of Biological Diversity

Modeling of Biological Space with Affy Data

- **Concept:** Model biological diversity in an Affymetrix gene expression database (CCLE, NCI-60) to identify candidate cell lines with diversity in basal gene expression levels, and thus representative of different regions of biological space.

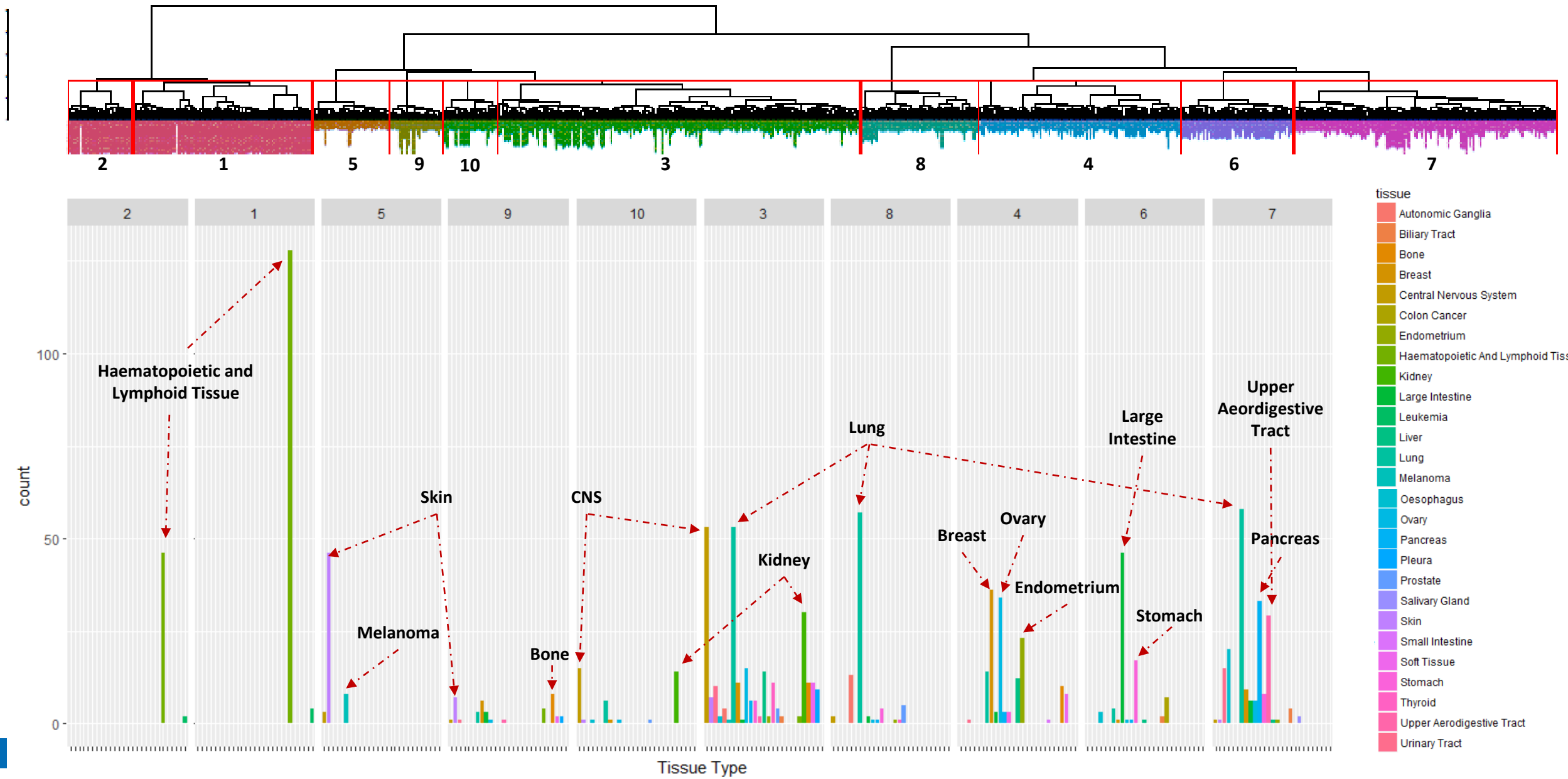
- **Considerations:**
  - Commercial availability,
  - Scalability,
  - Suitability for 384-well microplate culture format,
  - Authentication,
  - Adherent cells preferred, but not required.

- **Two Approaches:**

  - Cluster Analysis

  - “Volume” Maximization

Biological Diversity Approach: Cancer Cell Tissue Origins



Biological Diversity Approach: Cancer Cell Line Candidates

Cluster	Cell Line	Distance from Centroid [a]	Tissue Origin	Histology	Media	Notes	Sourcing
Adherent Cells							
3	HOP-62	11.81	Lung	Adenocarcinoma	RPMI 1640 + 5% FBS + 2 mM L-glutamine	2nd closest to centroid; *(Priority to domestic source)*	NCI
4	MCF7	16.7	Breast	Adenocarcinoma	DMEM + 10% FBS	12th from Centriod; *(Priority for inclusion in test panel)*	NCCT
5	M14	10.3	Skin [c]	Melanoma	RPMI 1640 + 5% FBS + 2 mM L-glutamine	Closest to centroid	NCI
6	HCC2998	11.5	Large Intestine [d]	Adenocarcinoma	RPMI 1640 + 5% FBS + 2 mM L-glutamine	Closest to centroid	NCI
7	Panc 03.27	13.5	Pancreas	Adenocarcinoma	RPMI 1640 + 15% FBS + 10 U/mL insulin	2nd closest to centroid; *(Priority to domestic source)*	ATCC
9	Hs 600.T	8.95	Skin	Melanoma	DMEM + 10% FBS	Closest to centroid	ATCC
10	U-343MG	15.6	Central Nervous System	Glioblastoma	MEM + 0.1 mM NEAA + 2 mM L-glutamine + 10% FBS	20th closest to centroid; *(Priority to availability and authentication)*	CLS
Non-Adherent Cells							
1	MOLT-4	12.79	Hematopoietic & Lymphoid [b]	Acute lymphoblastic leukemia	RPMI 1640 + 5% FBS + 2 mM L-glutamine	Closest to centroid	NCI
2	HL-60	12.54	Hematopoietic & Lymphoid [b]	Acute myeloid leukemia	RPMI 1640 + 5% FBS + 2 mM L-glutamine	Closest to centroid	NCI
8	COR-L51	15.8	Lung	Carcinoma	RPMI 1640 + 10% FBS + 2 mM L-glutamine	2nd closest to centroid; (*Priority to simplified media*)	ECACC

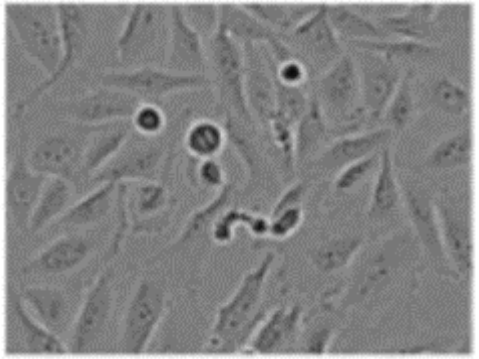
- Broad coverage of tissue origins and tumor types.
- Serum-supplemented “common” basal medias in most cases.
- Transcriptome Coverage:
  - 87% of genes are expressed in the upper 20%-ile of their expression range in at least one line
  - 83% are expressed in the lower 20th %-ile of their expression range in at least one line.

[a] Euclidean Distance for Mean Centered Data
[b] Tissue of origin listed as “leukemia” in database
[c] Tissue of origin listed as “melanoma” in database.
[d] Tissue of origin listed as “colon cancer” in database.

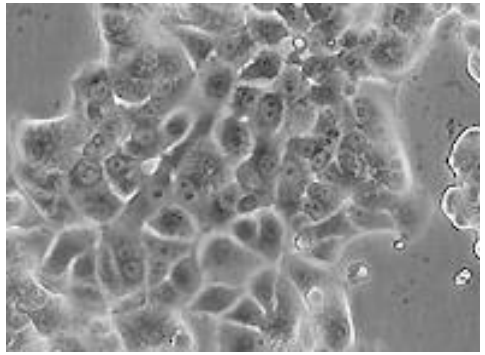
EPA United States Environmental Protection Agency

Morphological Heterogeneity of Cancer Cell Line Candidates

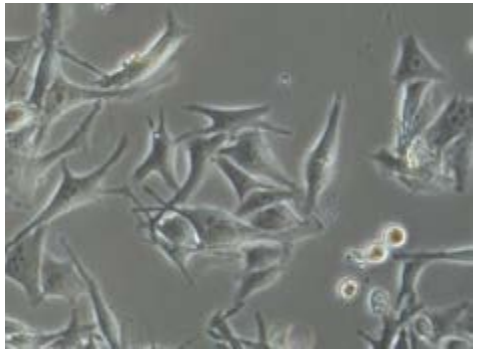
HOP-62



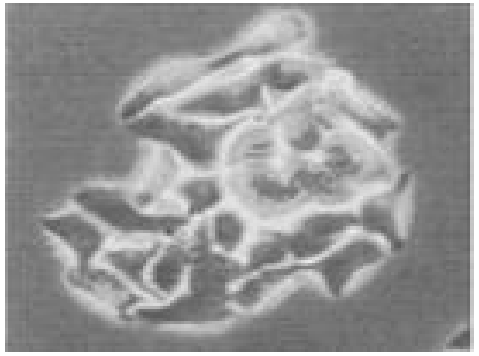
MCF7



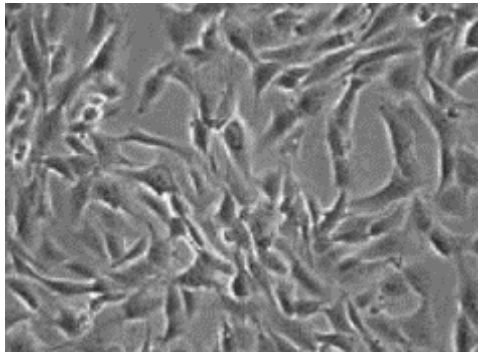
M14



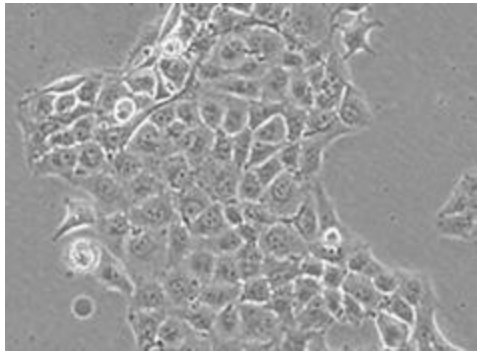
HCC2998



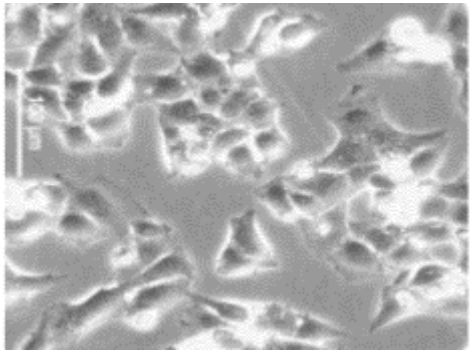
Hs 600.T *



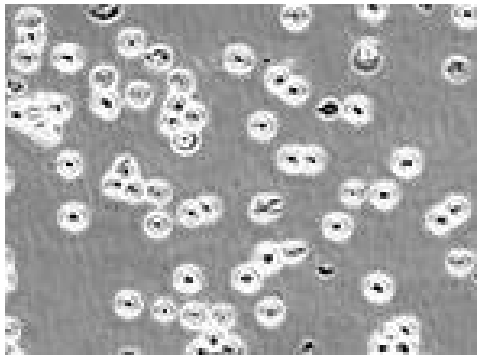
Panc 03.27 *



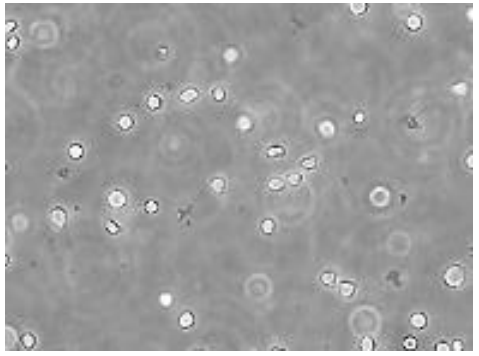
U-343MG



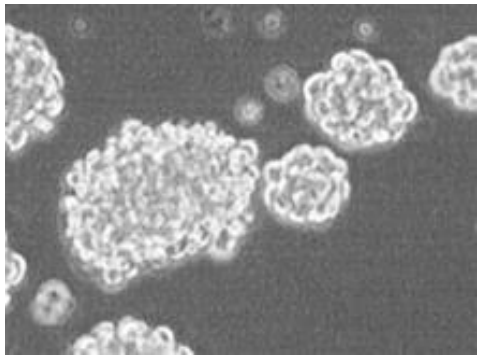
MOLT-4



HL-60



COR-L51 *



Option 1A

United States Environmental Protection Agency

Biological Diversity Approach: "Batting Order"

Prioritization Scheme:

1. Use MCF7 as the starting point.
2. For each new line, pick the remaining line that maximizes the sum of the distances from the cell lines that have already been run.

Order of Screening	Cell Line	Tissue Origin	Cancer Type	Cluster ID
1	MCF7	Breast	Adenocarcinoma	4
2	Hs 600.T	Skin	Melanoma	9
3	COR-L51	Lung	Small Cell Lung Carcinoma	8
4	HL-60	Hematopoietic & Lymphoid	Acute Myeloid Leukemia	2
5	U-343MG	Central Nervous System	Glioblastoma	10
6	Panc 03.27	Pancreas	Adenocarcinoma	7
7	MOLT-4	Hematopoietic & Lymphoid	Acute Lymphoblastic Leukemia	1
8	HCC2998	Large Intestine	Adenocarcinoma	6
9	M14	Skin	Melanoma	5
10	HOP-62	Lung	Adenocarcinoma	3

United States Environmental Protection Agency

Biological Diversity Approach: Volume Maximization

- Use computational methods to select a set of 5 cells lines which maximize biological diversity in basal gene expression within the database.

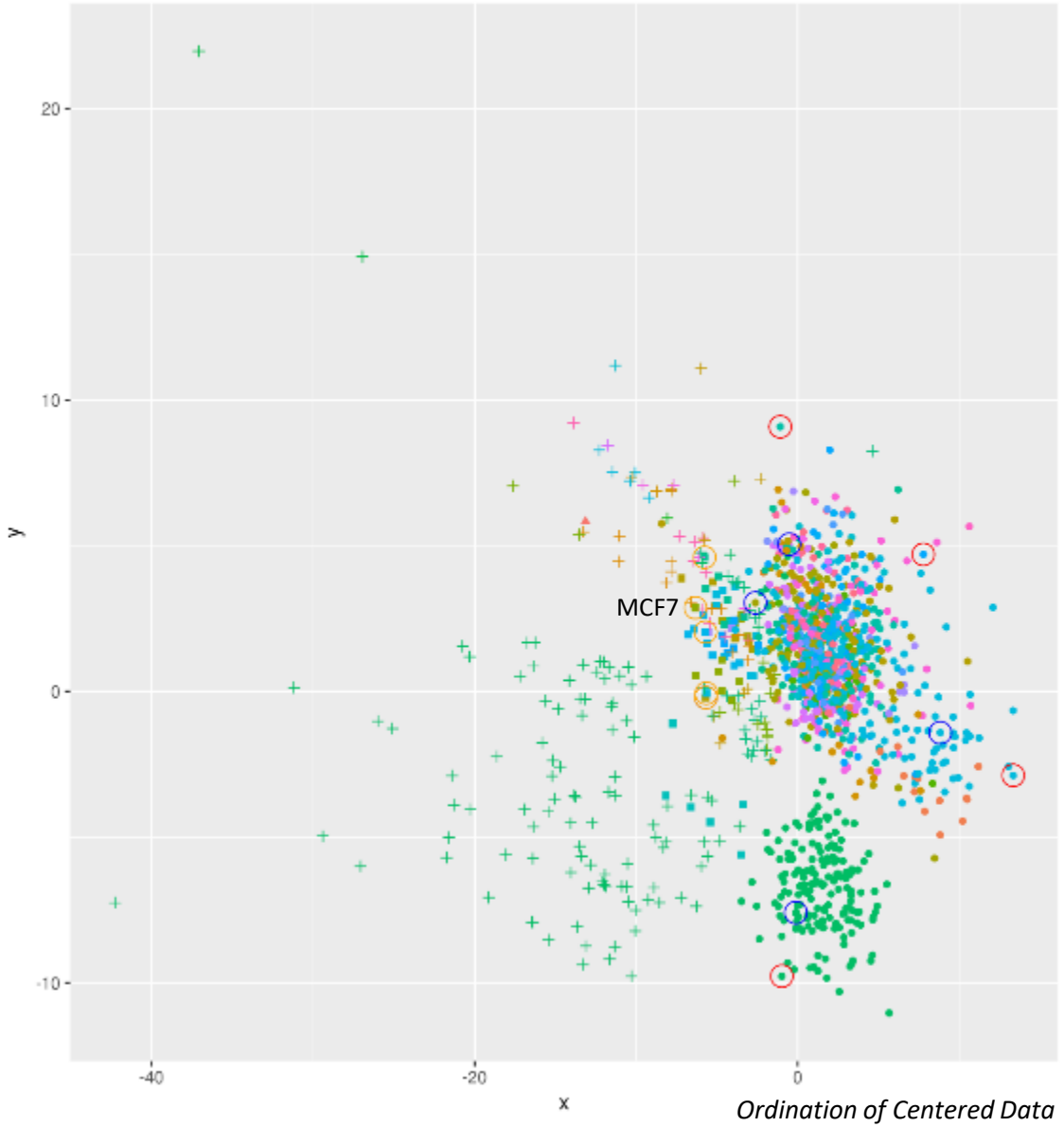
    1. Select a “seed” cell line → MCF7 cells.
    2. Find the cell line “furthest” from the seed.
    3. Find a third cell line that maximizes the area of a triangle between the three points.
    4. Find a forth cell line that maximize the volume of a pyramid between the four points.
    5. Continue selecting cell lines up to *n* points.
    6. Define the “volume” of the *n* dimensional polyhedron.
    7. Identify the list of *n* lines that meet study criteria & have the largest “volume” → covers the most biological space.

- This approach was performed with the CCLE + NCI-60 database and NCI-60 “adherent only” database.

Option 1B

United States Environmental Protection Agency

Biological Diversity Approach: Volume Maximization



Ordination of Centered Data

source
- ccle
- EPA
- nci60
- primary

tissue
- Adrenal
- Autonomic Ganglia
- Biliary Tract
- Blood Vessel
- Bone
- Brain
- Breast
- Central Nervous System
- Colon Cancer
- Embryonic stem cells
- Endometrium
- Fat
- Gametocyte
- Haematopoietic And Lymphoid Tissue
- iPS cells
- Kidney
- Large Intestine
- Leukemia
- Liver
- Lung
- Melanoma
- Oesophagus
- Ovary
- Pancreas
- Pleura
- Prostate
- Salivary Gland
- Skin
- Small Intestine
- Soft Tissue
- Stomach
- Thyroid
- Tissue stem cells
- Upper Aerodigestive Tract
- Urinary Tract

Red [CCLE + NCI-60]
MCF7 (Breast)
NCI-H1092 (Lung)
MOTN-1 (Haem. & Lymphoid)
Hs 255.T (Large Intestine)
JHOM-2B (Ovary)
NCI-60 Adherent Only
MCF7 (Breast)
SK-MEL-28 (Skin)
A-498 (Kidney)
NCI-H522 (Lung)
HT-29 (Colon)
Cluster Centroids
MCF7 (Breast)
Hs 600.T (Skin)
COR-L51 (Lung)
HL-60 (Lymphoid)
U-343MG (Central Nervous System)

Modeling of Biological Space with Affy Data

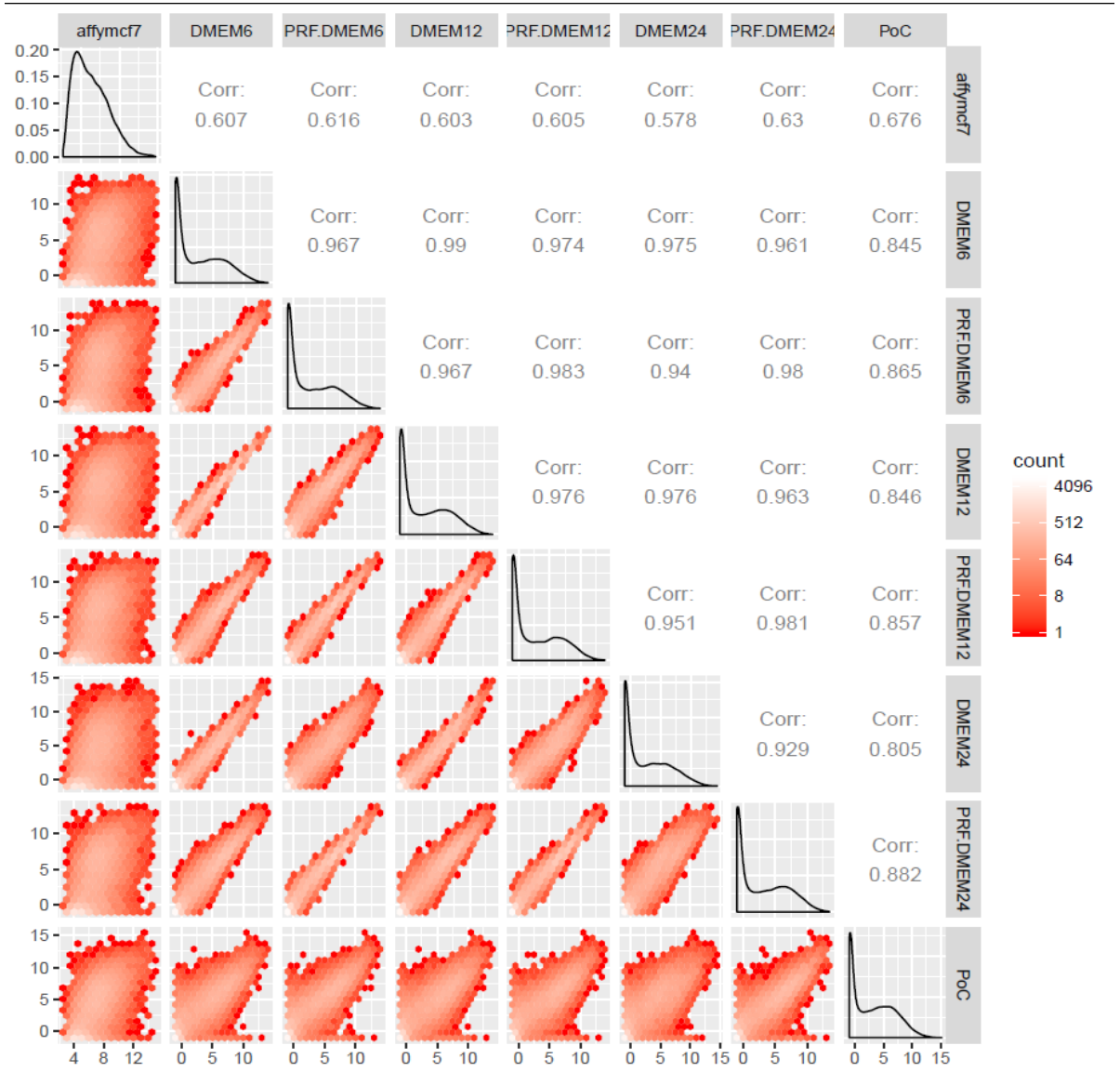
Problem (?):

We observed low correlation in the level of basal gene expression (not FC) when comparing TempO-Seq MCF7 data to Affymetrix data.

Evaluation of biological diversity using “baseline” Affymetrix data may not necessarily be predictive of results using TempO-Seq.

If TempO-Seq is the platform of choice, then…
to guide selection of a cell line panel with maximal biological diversity, TempO-Seq data is required.

There is no existing database of basal gene expression profiles across diverse cells using the TempO-Seq platform.



CRADA Addendum

Objectives

- Characterize basal gene expression in a set of cell culture models using TempO-Seq whole transcriptome profiling.
- Model biological diversity within this cell line panel using computational tools.
- Select a cell line which maximizes the difference in biological space as compared to MCF7 cells
- Screen a set of reference chemicals with well characterized molecular / pathway targets
- Use the combined set of reference profiles to identify potential molecular/pathway targets for the consensus chemicals.

CRADA Addendum

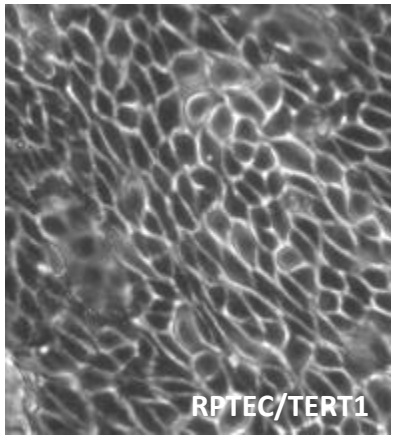
Phase 1

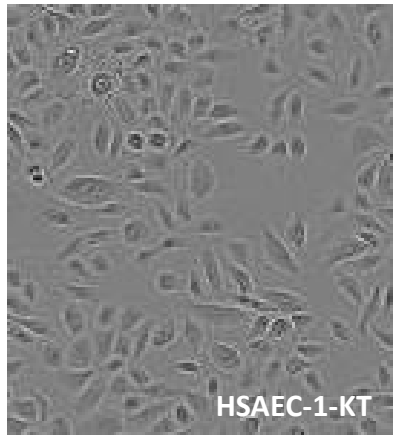
- NCCT will procure a set of human cancer cell lines (n ~ 10) with diverse basal gene expression profiles (as determined from an Affymetrix database, the CCLE). Also include phenotypic profiling cell lines (n ~7).
- NCCT will also procure a set of hTERT-immortalized human primary cells (n ~14) of various tissue origins.
- Cells will be expanded a limited number of passages according to the supplier's recommended protocol.
- Cell lysates will be created from the cells grown in microtiter plate format and whole transcriptome profiles will be generated using the TempO-Seq human whole transcriptome assay.
- The resultant baseline gene expression data will be used to map biological diversity in comparison with the reference cell line (i.e. MCF7 cells).
- The cell line with the maximum distance in biological space as compared to MCF7 cells will be selected for reference chemical screening in Phase 2.
- **Anticipated timeline:** 6-9 months from receipt of material.

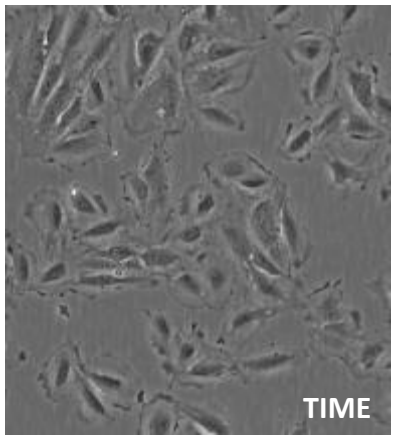
Immortalized Primary Cell Panel

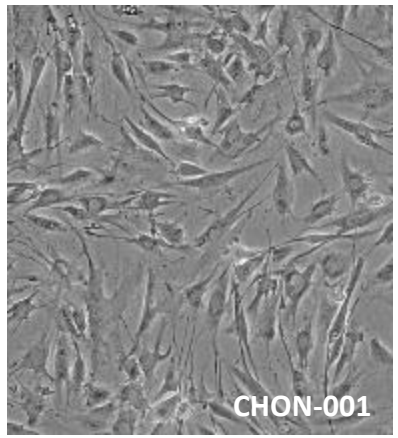
- ATCC offers a variety of cell lines immortalized using human telomerase transcriptase (hTERT).
- Derived for normal (i.e. non-cancerous) tissue
- Cells maintain phenotypic characteristics of source tissue *in vitro*.
- Karyotypes are more “normal” compared to analogous cancer cell lines.
- Most require cocktails of media supplements to support growth *in vitro.*

Cell Type	Tissue	Name	ATCC Number	Disease	Morphology
Endothelial	Aorta	TeloHAEC	CRL-4052	Normal	endothelial
Endothelial	Foreskin; Dermal Microvascular Endothelium	TIME	CRL-4025	Normal	endothelial-like
Endothelial	Foreskin	NFkB-TIME	CRL-4049	Normal	endothelial
Epithelial	Lung, Small Airway	HSAEC-1-KT	CRL-4050	Normal	epithelial, packed cuboidal
Epithelial	Lung, Bronchial	HBEC3-KT	CRL-4051	Normal	epithelial, packed cuboidal
Epithelial	Breast; Mammary Gland	hTERT-HME1	CRL-4010	Normal	epithelial-like
Epithelial	Renal cortex, proximal tubules	RPTEC/TERT1	CRL-4031	Normal	epithelial-like
Epithelial	Retina, Eye	hTERT RPE-1	CRL-4000	Normal	epithelial-like
Epithelial	Esophagus Epithelium	CP-A (KR-42421)	CRL-4027	Non-dysplastic metaplasia	epithelial-like
Epithelial	Lung, Bronchial	NuLi-1	CRL-4011	Normal	epithelial-like
Fibroblast	Long Bone, Cartilage	CHON-001	CRL-2846	Normal	fibroblast-like
Fibroblast	Uterus; Endometrium	T HESCs	CRL-4003	Non-malignant myomas	fibroblast-like
Intermediary	Pancreas, Duct	hTERT-HPNE	CRL-4023	Normal	epithelial-like
Keratinocyte	Foreskin	Ker-CT	CRL-4048	Normal	epithelial
Mesenchymal	Adipose-derived mesenchymal stem cell	ASC52telo	SCRC-4000	Normal (?)	fibroblast-like
Fetal	Long Bone, Cartilage	CHON-002	CRL-2847	Normal	fibroblast-like









Images and information adapted from: www.atcc.org

CRADA Addendum

Phase 2

- NCCT will expand and create a large cryostock of the cell line selected from Phase 1.
- The cell line will then be screened with a set of approximately 350 reference chemicals which were previously tested in the MCF7 cell model.
- The chemicals will be screened in 8-point concentration-response in triplicate cultures.
- Cell lysates from this screening campaign will be analyzed using the BioSpyder human whole transcriptome assay.
- The resultant data will be used to compare transcriptional responses across the two cell lines in terms of the identity and point-of-departure of gene and pathway level effects.
- **Anticipated timeline:** 6-9 months from selection of cell line.

Acknowledgments

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

NCCT Labs:
Clinton Willis
Johanna Nyffeler

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