

# EVALUATION OF THE TOXCAST SUITE OF CELLULAR AND MOLECULAR ASSAYS FOR PREDICTION OF IN VIVO TOXICITY

*Keith Houck, Ph.D.*

*SOT 2009  
Baltimore, MD*

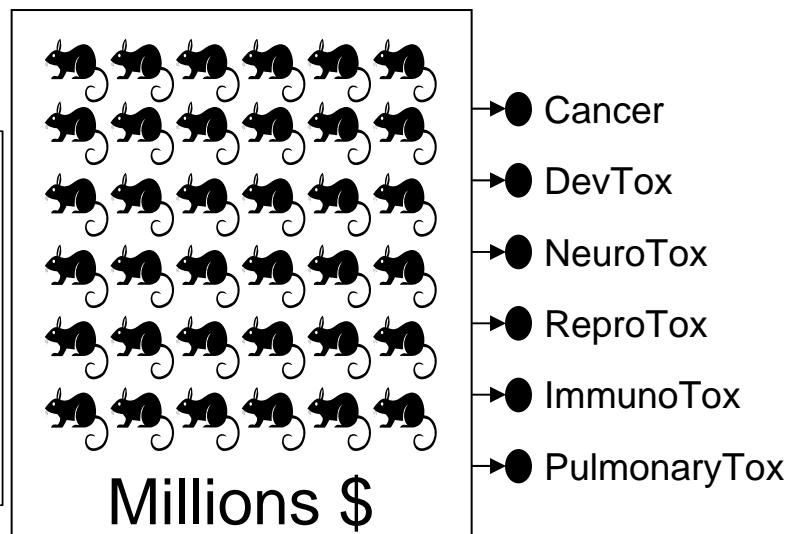
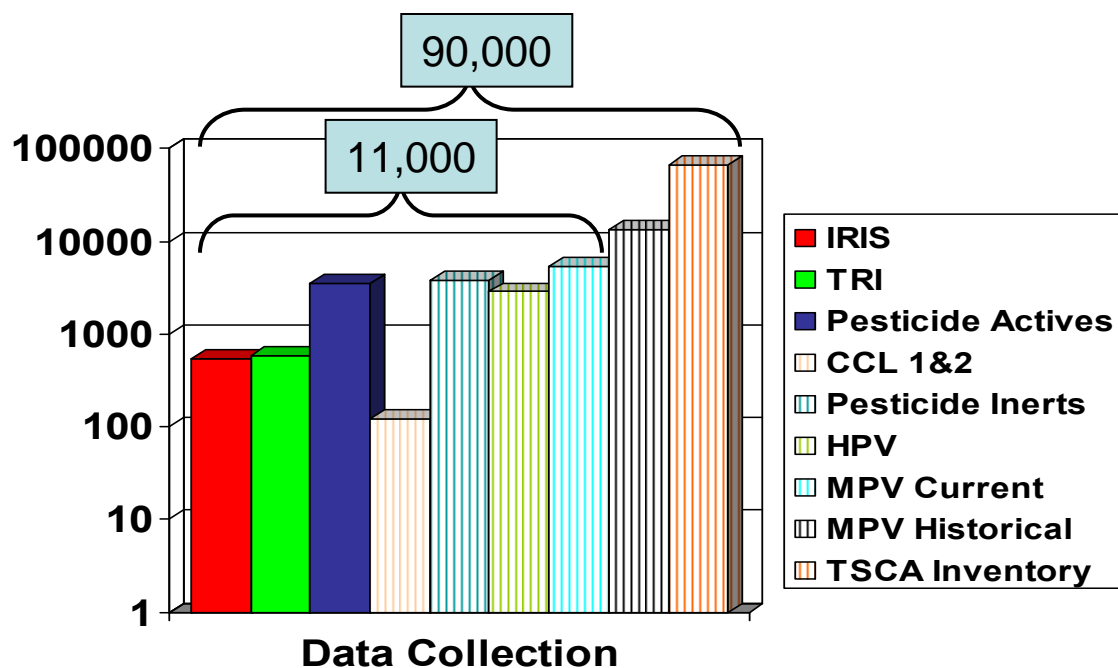
UNITED STATES ENVIRONMENTAL PROTECTION AGENCY



# EPA's Need for Chemical Prioritization

## Too Many Chemicals

## Too High a Cost



...and not enough data.

## How Can We Prioritize?

- Animal studies
  - cost, time, ethical considerations
- Exposure
  - lacks hazard information
- QSAR
  - domain of applicability issues
  - lack of availability of sufficient models
- Bioactivity Profiling
  - biologically relevant chemical characterization
  - high-throughput capacity
  - needs development and validation

 **ToxCast**

# Biological spectra analysis: Linking biological activity profiles to molecular structure

Anton F. Fliri\*, William T. Loging, Peter F. Thadeio, and Robert A. Volkman<sup>\*†</sup> PNAS January 11, 2005 vol. 102 no. 2 261–266

Pfizer Global Research and Development, Groton, CT 06340

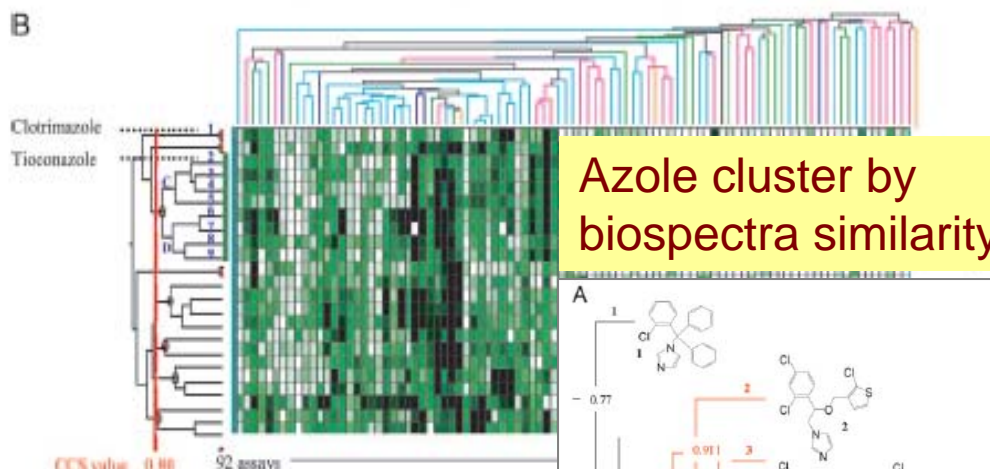
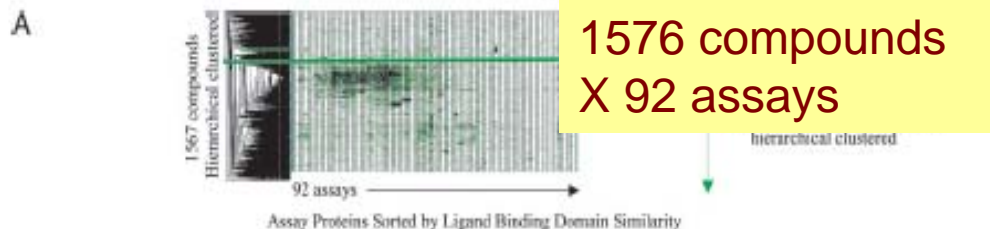
Communicated by Larry E. Overman, University of California, Irvine, CA, October 25, 2004 (received for review September 4, 2004)

Establishing quantitative relationships between molecular structure and broad biological effects has been a longstanding challenge in science. Currently, no method exists for forecasting broad biological activity profiles of medicinal agents even within narrow boundaries of structurally similar molecules. Starting from the premise that biological activity results from the capacity of small organic molecules to modulate the activity of the proteome, we set out to investigate whether descriptor sets could be developed for measuring and quantifying this molecular property. Using a 1,567-compound database, we show that percent inhibition values, determined at single high drug concentration in a battery of *in vitro* assays representing a cross section of the proteome, provide precise molecular property descriptors that identify the structure of molecules. When broad biological activity of molecules is represented in spectra form, organic molecules can be sorted by quantifying differences between biological spectra. Unlike traditional structure–activity relationship methods, sorting of molecules by using biospectra comparisons does not require knowledge of a molecule’s putative drug targets. To illustrate this finding, we selected as starting point the biological activity spectra of clotrimazole and tioconazole because their putative target, lanosterol demethylase (CYP51), was not included in the bioassay array. Spectra similarity obtained through profile similarity measurements and hierarchical clustering provided an unbiased means for establishing quantitative relationships between chemical structures and biological activity spectra. This methodology, which we have termed biological spectra analysis, provides the capability not only of sorting molecules on the basis of biospectra similarity but also of predicting simultaneous interactions of new molecules with multiple proteins.

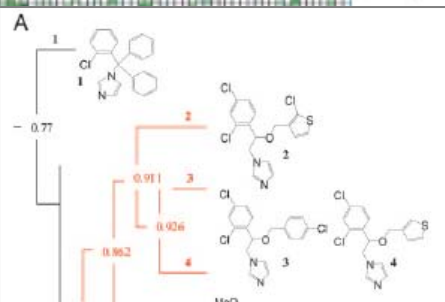
biospectra | proteome | structure–function relationships

differences in biological environments (8). Considering the complexity of this requirement, computational solutions that precisely link molecular structure to broad biological response are currently not possible (9, 10). We report here an approach to structure–function studies that is based on measurements of the capacity of molecules to interact with the proteome (11).

## Translation of Chemical Property Information into Biological Activity Spectra



Azole cluster by biospectra similarity

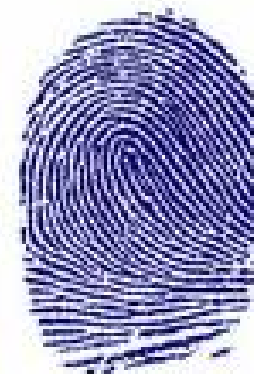


Approximate “proteomic” signature



# ToxCast™ : a computational toxicology approach based on high-throughput bioactivity profiling

- Research program of EPA's National Center for Computational Toxicology
- Addresses chemical screening and prioritization needs for pesticidal inerts, anti-microbials, CCLs, HPVs and MPVs
- Comprehensive use of HTS technologies
- Coordinated with NIH: NTP and NHGRI/NCGC via Tox21
- Committed to stakeholder involvement and public release of data
  - Chemical Prioritization Community of Practice
  - NCCT website- <http://www.epa.gov/ncct/toxcast>
  - ACToR- Aggregated Computational Toxicology Resource  
<http://actor.epa.gov/actor/>



# Prioritization Product Timeline

Phase	Number of Chemicals	Chemical Criteria	Purpose	Number of Assays	Cost per Chemical	Target Date
I	320	Data Rich (pesticides)	Signature Development	552	\$20k	FY08
Ib	15	Nanomaterials	Pilot	166	\$10K	FY09
Ila	>300	Data Rich Chemicals	Validation	>400	~\$20-25k	FY09
Ilb	>100	Known Human Toxicants	Extrapolation	>400	~\$20-25k	FY09
Ilc	>300	Expanded Structure and Use Diversity	Extension	>400	~\$20-25k	FY10
Ild	>12	Nanomaterials	PMN	>200	~\$15-20K	FY09-10
III	Thousands	Data poor	Prediction and Prioritization	>300	~\$15-20k	FY11-12

FY07

FY08

FY09

FY10

FY11

FY12

Proof of Concept: ToxCast

Verification/Extension

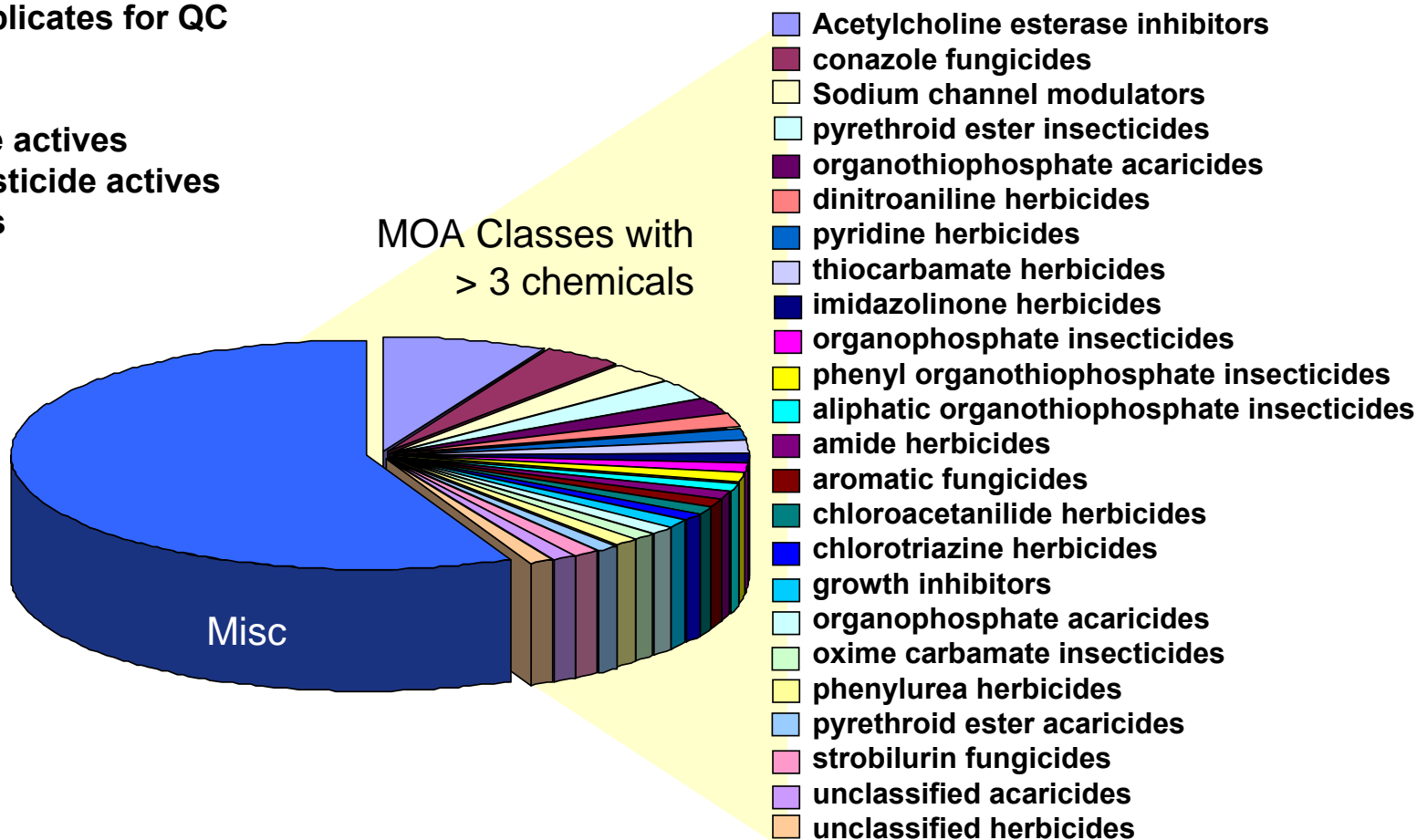
Reduce to Practice

Tox21

# ToxCast\_320 Phase I Chemicals

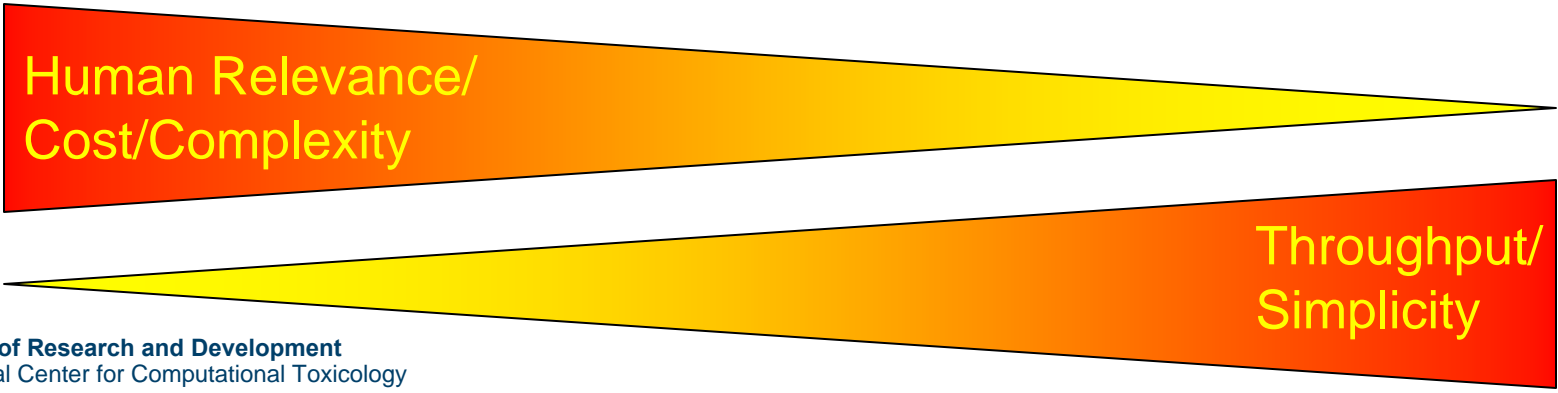
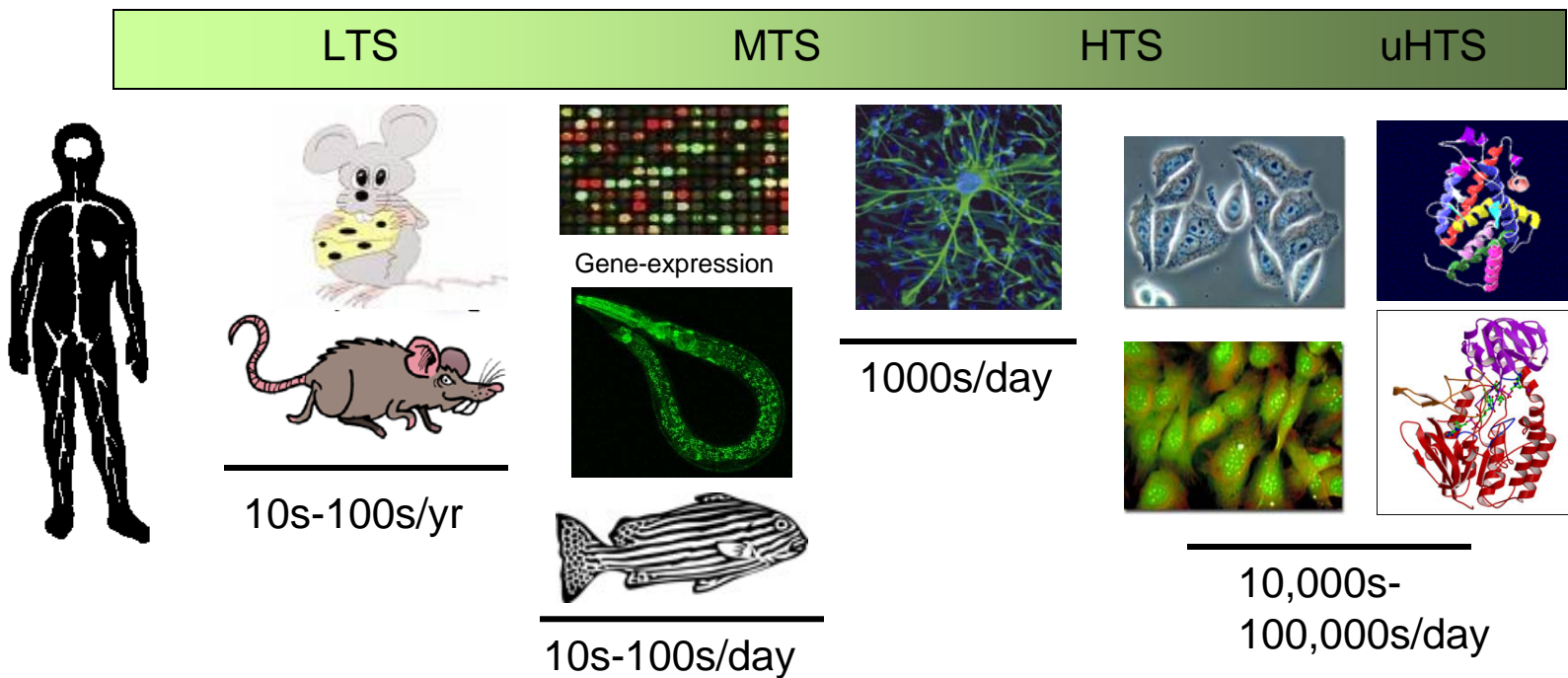
**309 unique structures**  
**3 triplicates, 5 duplicates for QC**  
**8 metabolites**

**291 total pesticide actives**  
**273 registered pesticide actives**  
**22 pesticide inerts**  
**33 antimicrobials**



# High-Throughput Screening Assays

*batch testing of chemicals for pharmacological/toxicological endpoints  
using automated liquid handling, detectors, and data acquisition*





# ToxCast Phase I Datasets

- Released to Data Analysis Partners:
  - ACEA - Real-time Cell Electronic Sensing (7 assays)
  - Attagene - Transcription factor assays (81 assays)
  - BioSeek - Cell-based protein level assays (87 assays)
  - Cellumen - Cell imaging assays (11 assays)
  - CellzDirect – NR target-gene expression assays (16 assays)
  - Gentrionix - GreenScreen GeneTox assay (1 assay)
  - NCGC - nuclear receptor assays (22 assays)
  - Novascreen / Caliper - receptor binding and enzyme inhibition assays (239 assays)
  - Solidus - P450 vs. cytotoxicity assays (4 assays)
- Upcoming Dataset Additions:
  - Neurite outgrowth HCS (NHEERL)
  - Cell proliferation (NHEERL)
  - Zebrafish developmental toxicity (NHEERL)
  - Organ toxicity; dosimetry (Hamner Institutes)
  - C. elegans WormTox (NIEHS)
  - Gene markers from microscale cultured hepatocytes (MIT)
  - 3D Cellular Zebrafish vascular/cardiotoxicity (Zygogen)
  - HTS stress response (NHEERL+NCGC)

**468 Endpoints**

# ToxCast Assays

## Biochemical Assays

- Protein families
  - GPCR
  - NR
  - Kinase
  - Phosphatase
  - Protease
  - Other enzyme
  - Ion channel
  - Transporter
- Assay formats
  - Radioligand binding
  - Enzyme activity
  - Co-activator recruitment

## Model Organism Assays

- Zebrafish (development)
- *C. elegans* (development)

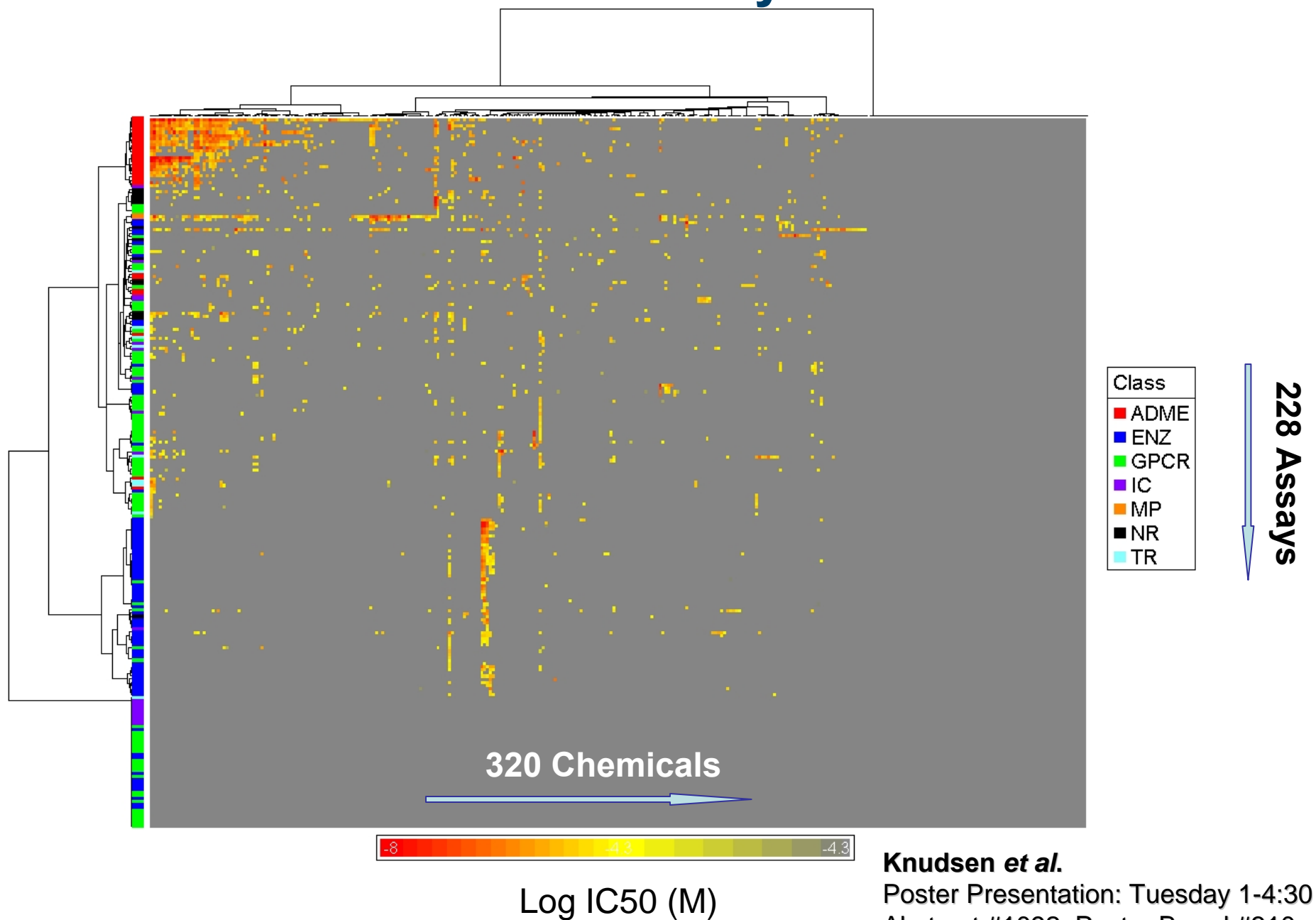
## Cellular Assays

- Cell lines
  - HepG2 human hepatoblastoma
  - A549 human lung carcinoma
  - HEK 293 human embryonic kidney
- Primary cells
  - Human endothelial cells
  - Human monocytes
  - Human keratinocytes
  - Human fibroblasts
  - Human proximal tubule kidney cells
  - Human small airway epithelial cells
- Biotransformation competent cells
  - Primary rat hepatocytes
  - Primary human hepatocytes
- Assay formats
  - Cytotoxicity
  - Reporter gene
  - Gene expression
  - Biomarker production
  - High-content imaging for cellular phenotype

# Biochemical Assays

- Diverse selection of targets
- Attempted to cover known toxicity targets
- Attempted to broadly profile major protein super-families
  - GPCR
  - Kinase
  - Phosphatase
  - Nuclear receptor
  - Proteases
  - Ion channels
  - others
- Screened at 25  $\mu$ M initially with actives screened in concentration-response format

# Biochemical Assay Results

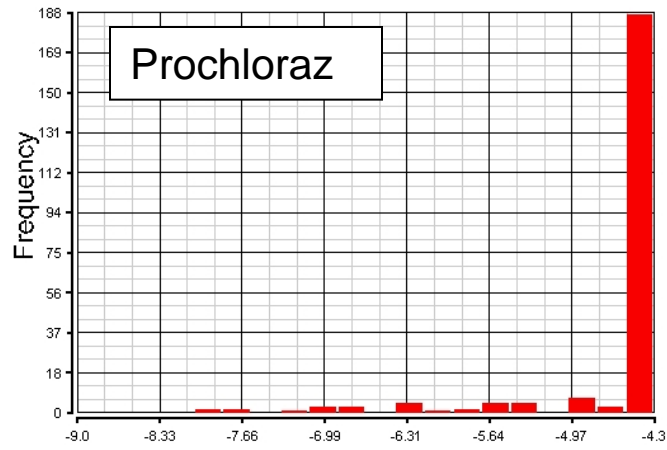


**Knudsen *et al.***

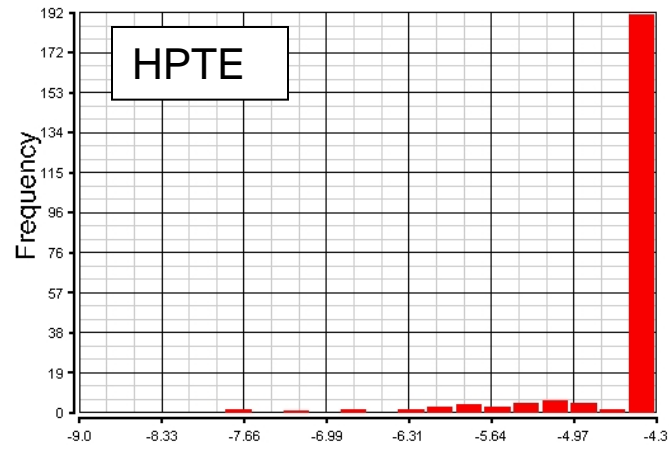
Poster Presentation: Tuesday 1-4:30  
Abstract #1092; Poster Board #310



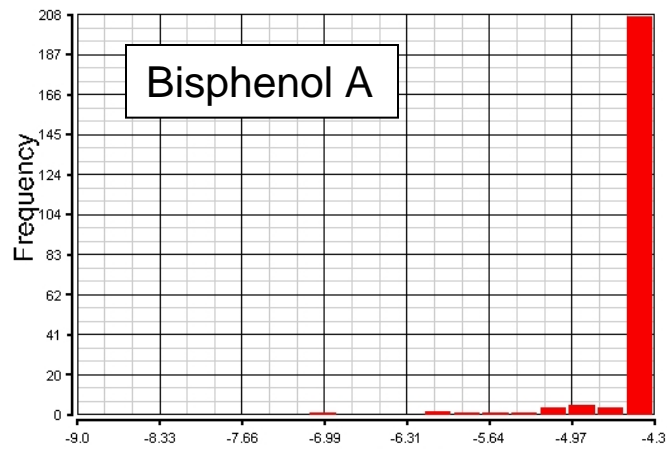
# Examples of Chemical Activity in Biochemical Assays



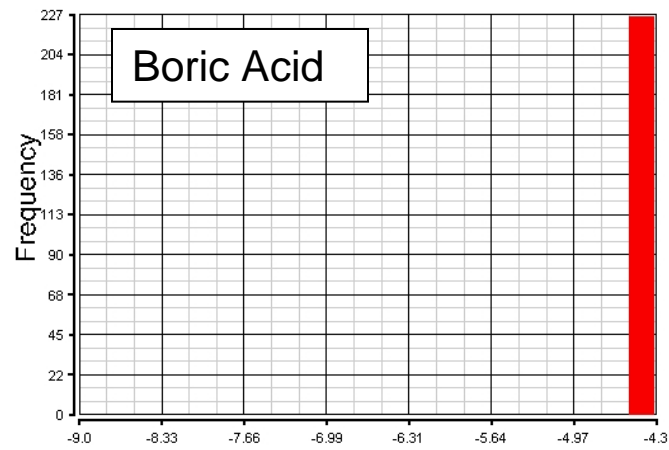
(IC50 log M)



(IC50 log M)

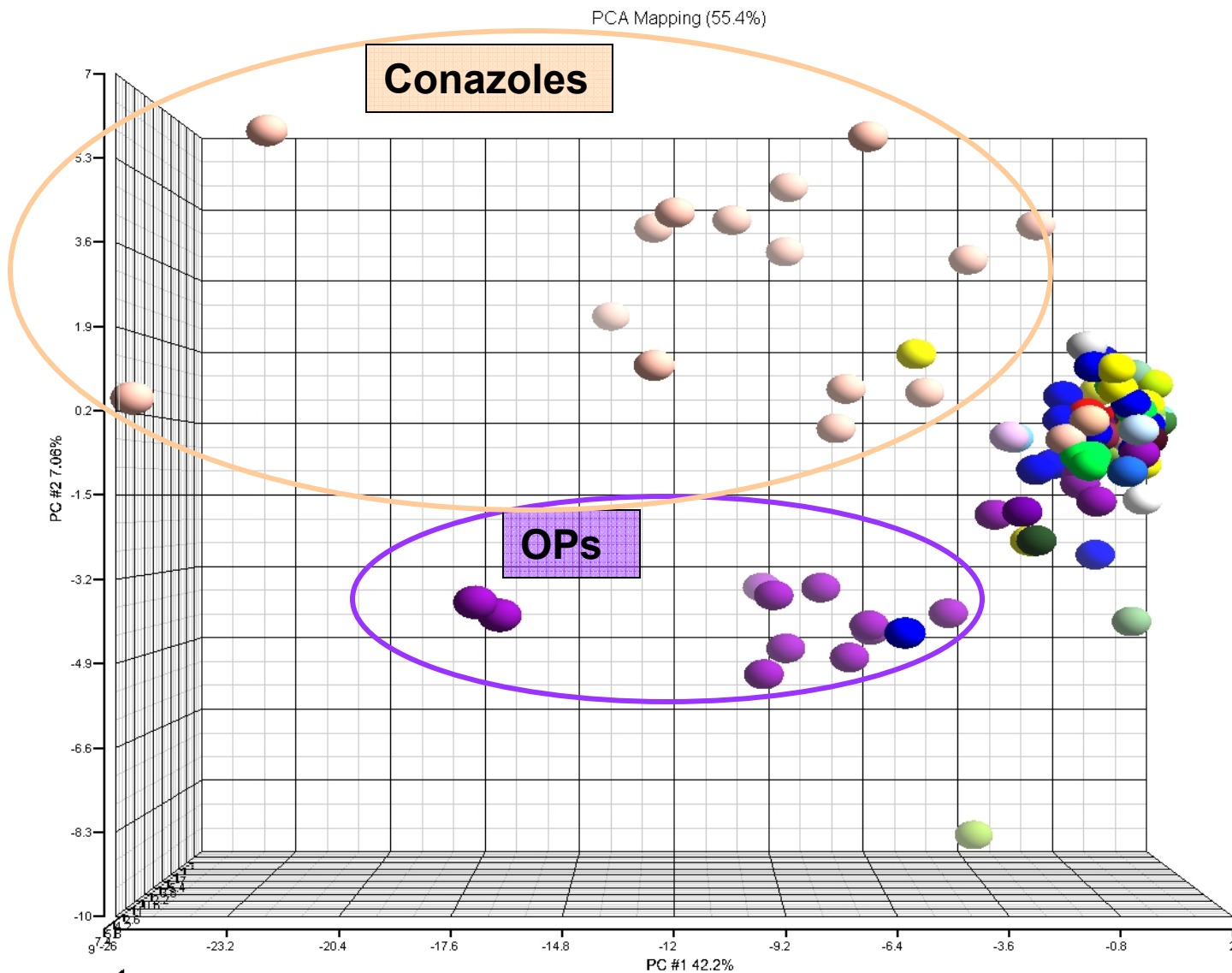


(IC50 log M)



(IC50 log M)









# PCA Mapping of CYP Inhibition



# Cellular Assays

- Types of Assays
  - Known toxicity pathways and targets
    - biomarker measurements
    - reporter gene assays
  - General cytotoxicity
  - Toxicity cellular phenotypes
- Cell lines and primary cells
- Generally screened at up to 100-200  $\mu\text{M}$  or MTC
- Concentration-response format used and  $\text{AC}_{50}$  generated

# Primary Human Cell Systems (BioSeek, Inc.)

System	Cell Types	Environment	Readouts
3C	 Endothelial cells	IL-1 $\beta$ +TNF- $\alpha$ +IFN- $\gamma$	MCP-1, VCAM-1, ICAM-1, Thrombomodulin, Tissue Factor, E-selectin, uPAR, IL-8, MIG, HLA-DR, Prolif., Vis., SRB (13)
4H	 Endothelial cells	IL-4+histamine	VEGFR2, P-selectin, VCAM-1, uPAR, Eotaxin-3, MCP-1, SRB (7)
LPS	 Peripheral Blood Mononuclear Cells + Endothelial cells	TLR4	CD40, VCAM-1, Tissue Factor, MCP-1, E-selectin, IL-1a, IL-8, M-CSF, TNF-a, PGE2, SRB (11)
SAg	 Peripheral Blood Mononuclear Cells + Endothelial cells	TCR	MCP-1, CD38, CD40, CD69, E-selectin, IL-8, MIG, PBMC Cytotox, SRB, Proliferation (10)
BE3C	 Bronchial epithelial cells	IL-1 $\beta$ +TNF- $\alpha$ +IFN- $\gamma$	uPAR, IP-10, MIG, HLA-DR, IL-1a, MMP-1, PAI-1, SRB, TGF-b1, tPA, uPA(11)
HDF3CGF	 Fibroblasts	IL-1 $\beta$ +TNF- $\alpha$ +IFN- $\gamma$ +bFGF+EGF+PDGF-BB	VCAM-1, IP-10, IL-8, MIG, Collagen III, M-CSF, MMP-1, PAI-1, Proliferation, TIMP-1, EGFR, SRB (12)
KF3CT	 Keratinocytes + Fibroblasts	IL-1 $\beta$ +TNF- $\alpha$ +IFN- $\gamma$ +TGF- $\beta$	MCP-1, ICAM-1, IP-10, IL-1a, MMP-9, TGF-b1, TIMP-2, uPA, SRB (9)
SM3C	 Vascular smooth muscle cells	IL-1 $\beta$ +TNF- $\alpha$ +IFN- $\gamma$	MCP-1, VCAM-1, Thrombomodulin, Tissue Factor, IL-6, LDLR, SAA, uPAR, IL-8, MIG, HLA-DR, M-CSF, Prolif., SRB (14)

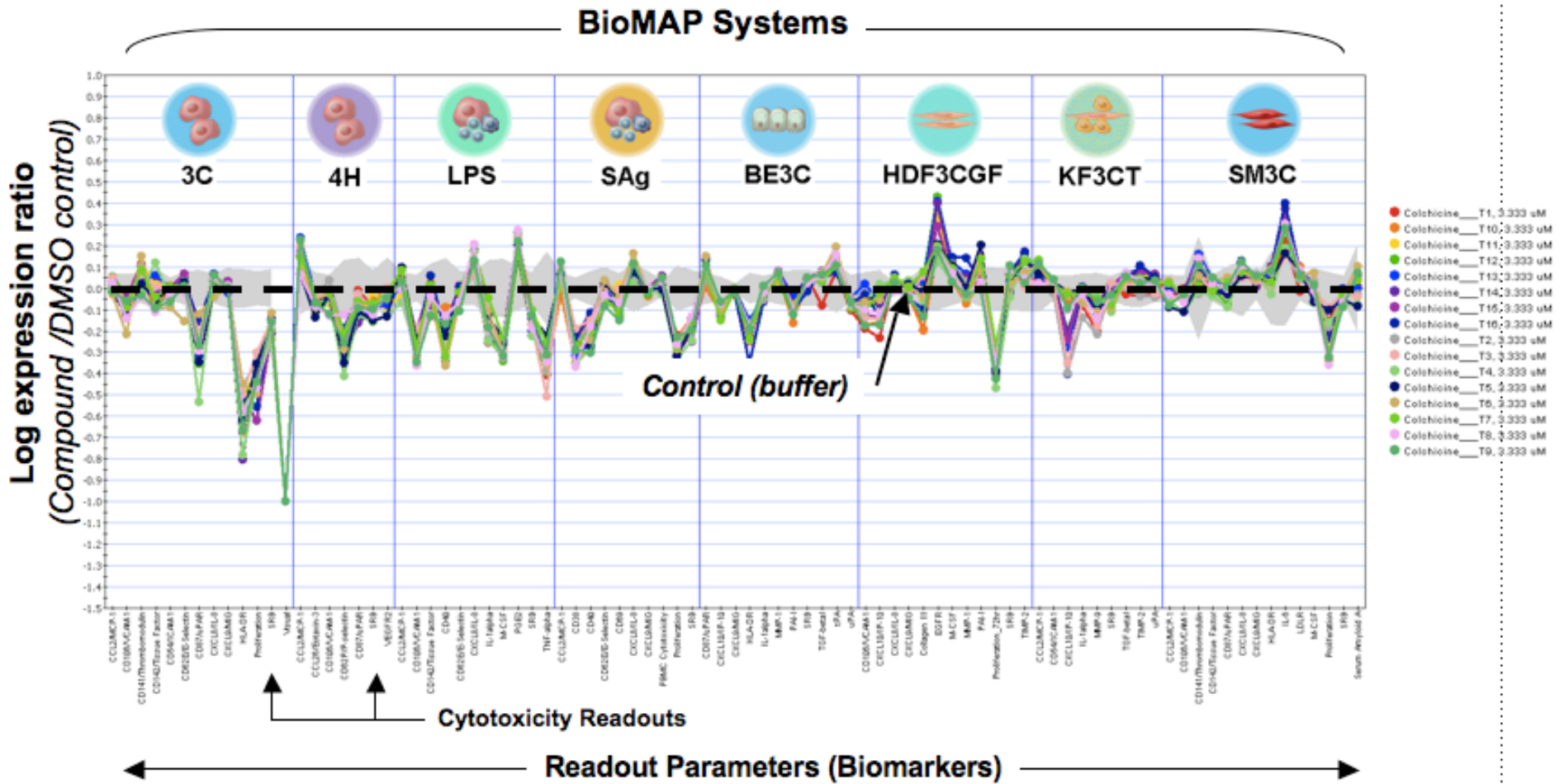
- 8 Assay systems
- 87 endpoints
- 4 concentrations



# Primary Human Cell Systems

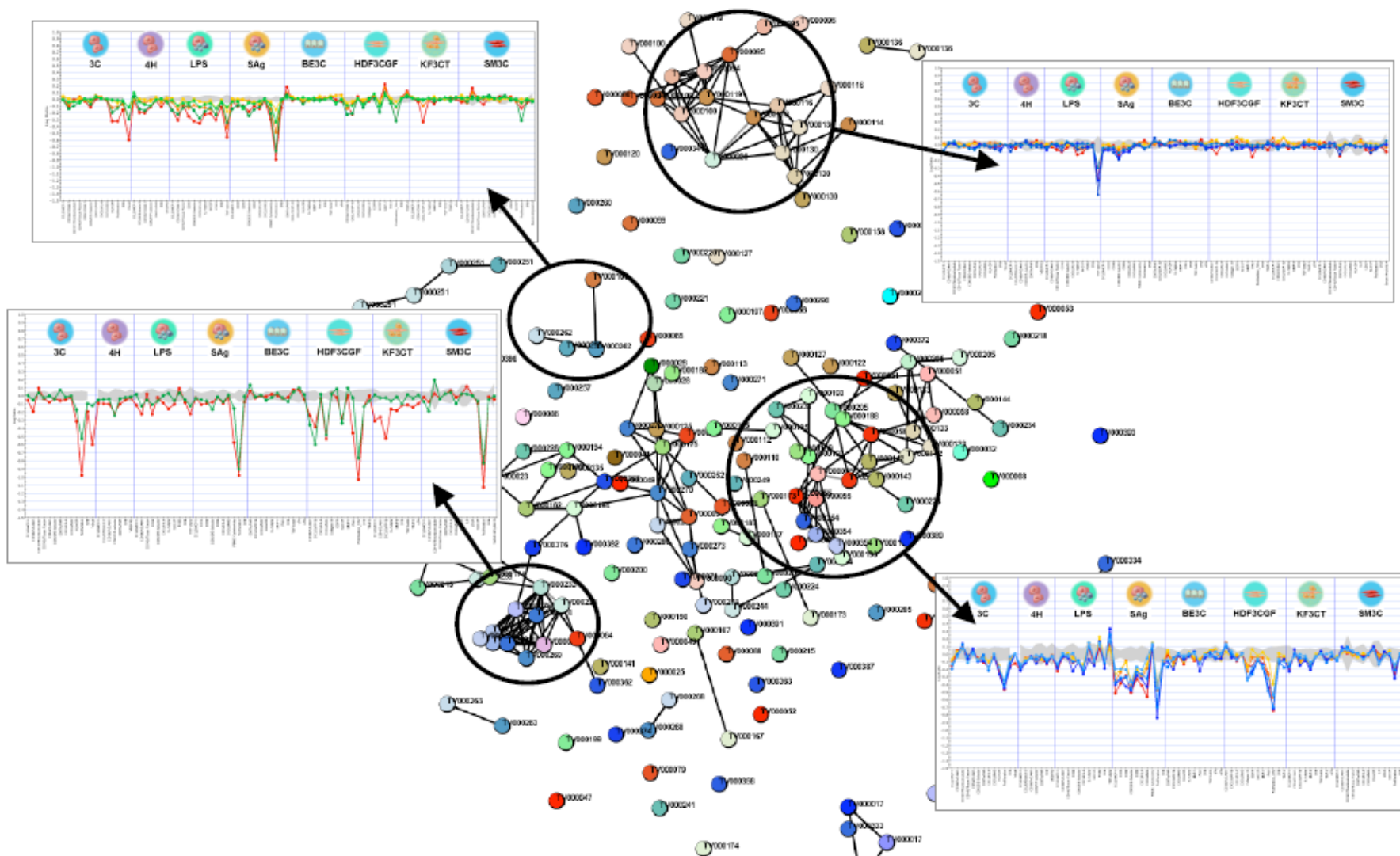
## Reference Compound Variability

### Colchicine

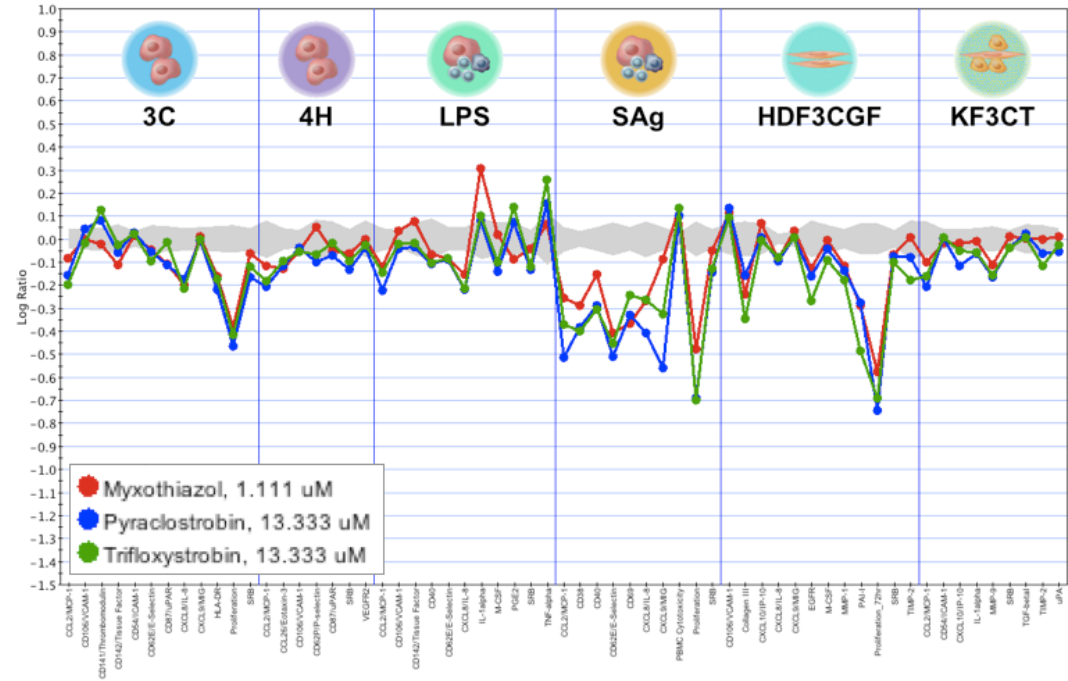
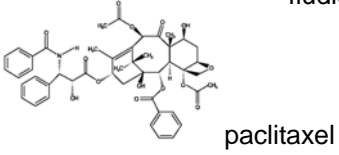
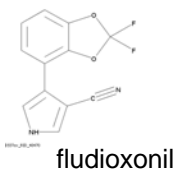
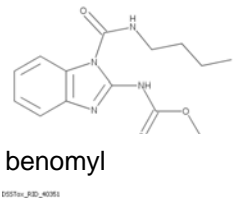
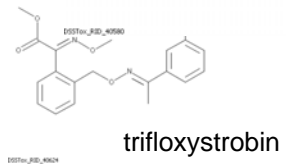
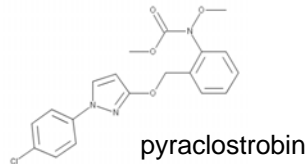
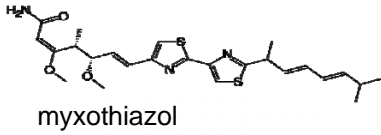


# Functional Similarity Map of ToxCast Library

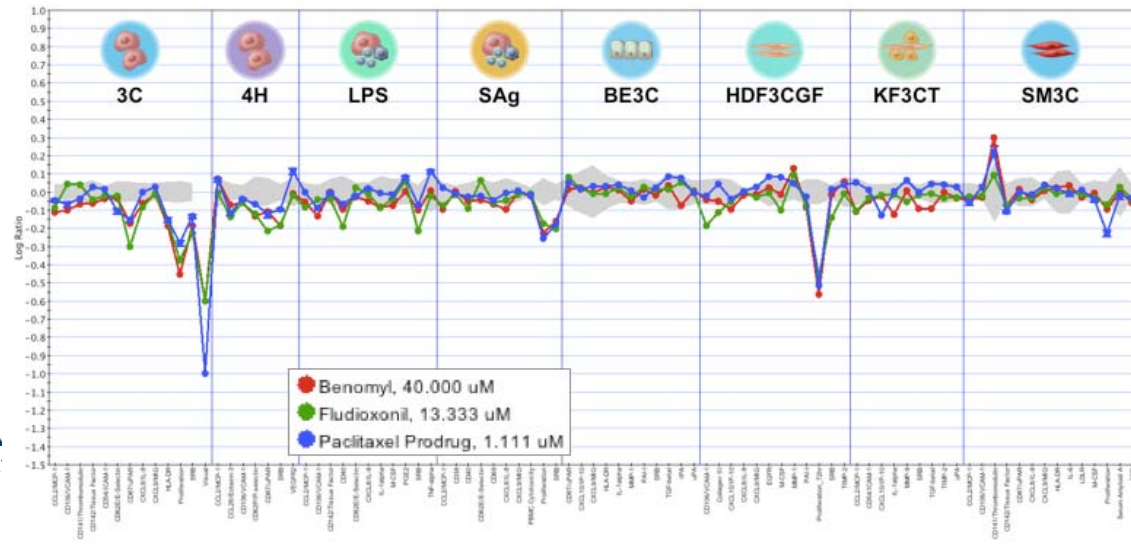
## Diversity of Mechanisms



# Mechanism Class Examples



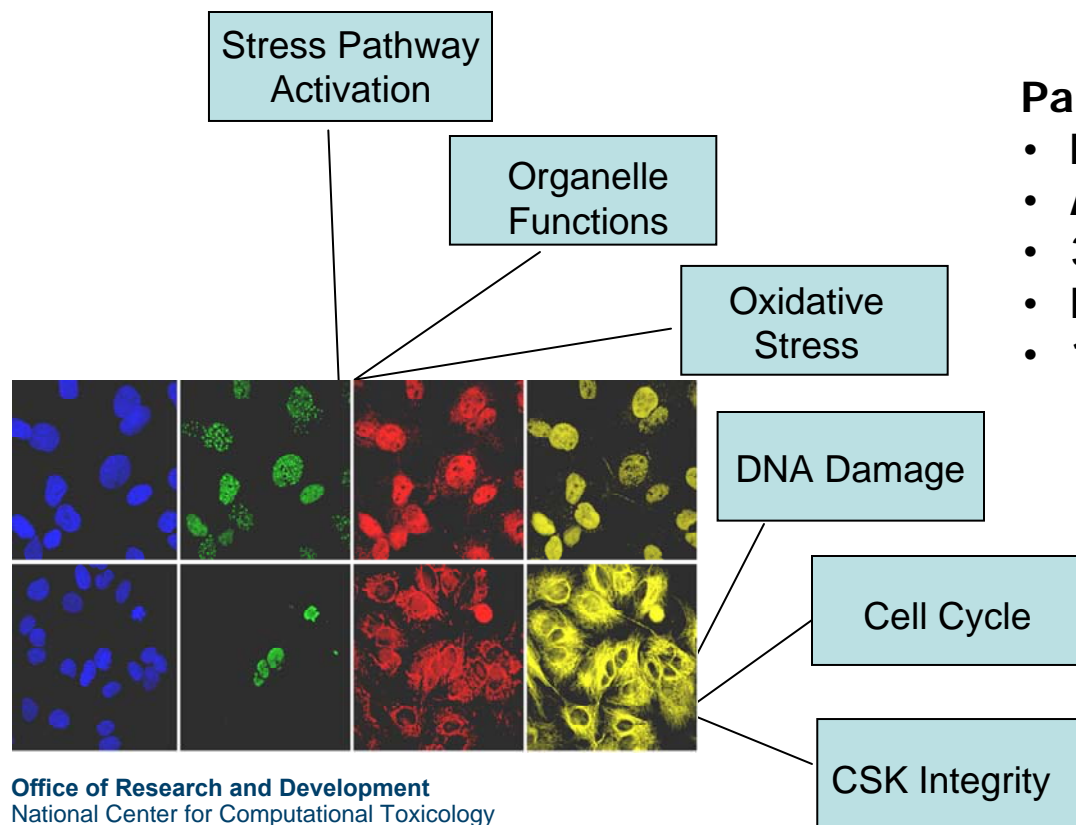
**Mitochondria  
disruptors**



**Microtubule  
disruptors**

# High-Content Screening of Cellular Phenotypic Toxicity Parameters (Cellumen, Inc.)

- Technology: automated fluorescent microscopy
- Objective: Determine effects of chemicals on toxicity biomarkers in a cell culture of HepG2 and primary rat hepatocytes



## Panel 1 design\*:

- Multiple mechanisms of toxicity
- Acute, early & chronic exposure
- 384-well capacity
- HepG2
- 1<sup>o</sup> rat hepatocytes

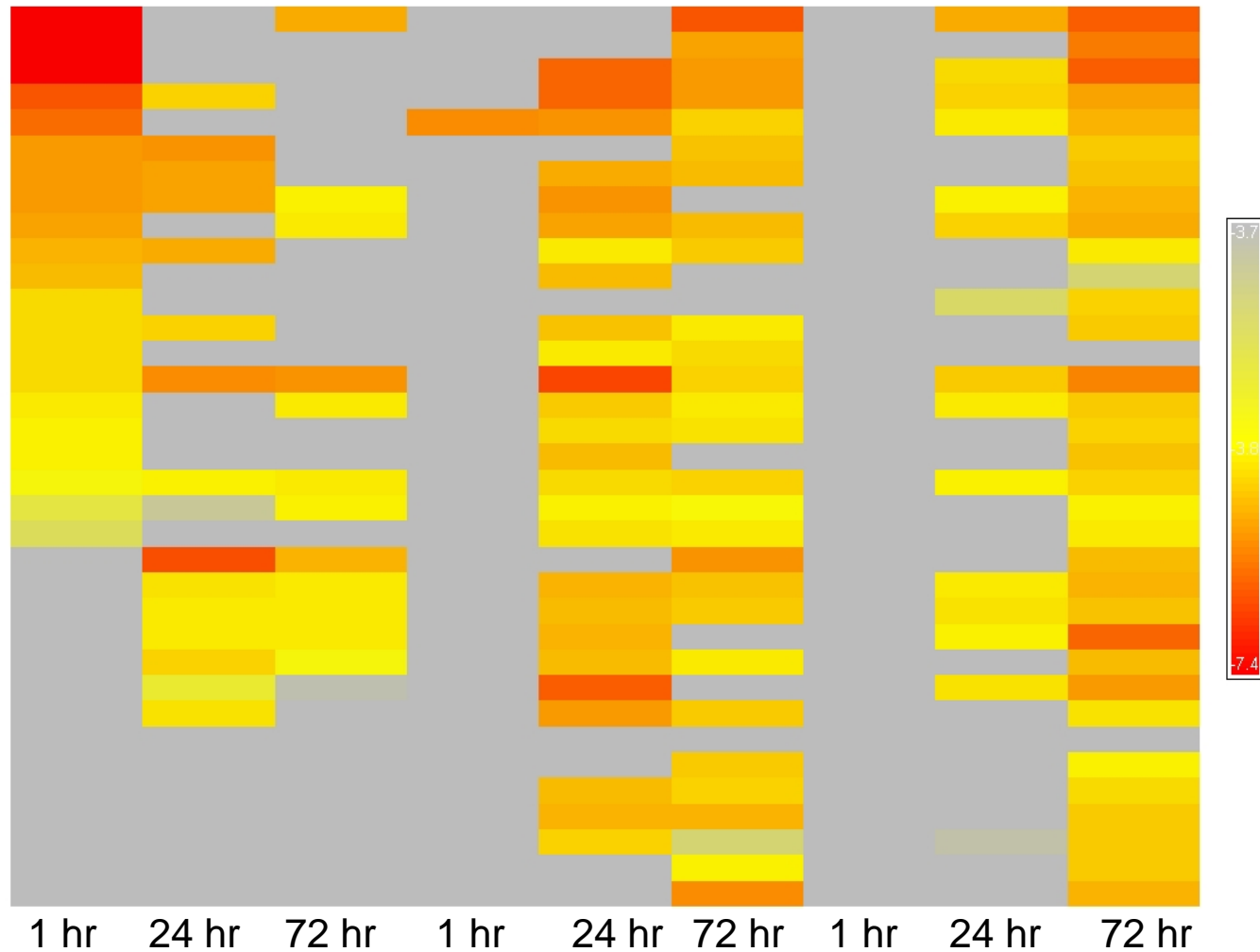



# CellCiphr™ Cytotoxicity Panel


- 10-point conc-response (200  $\mu$ M-39 nM)
- Three time points (1 hr, 24 hr, 72 hr)
- 11 endpoints per assay


Biomarker	Measurement	Positive Control	Z'
Stress Pathway	Phospho-c-jun	Anisomycin	0.63
Oxidative Stress	Phospho-Histone H2A.X	Camptothecin	0.7
Mitochondrial Function	Mitochondrial membrane potential	CCCP	0.55
Mitochondrial Mass	Mitochondrial mass	CCCP	0.35
Cell Loss	Cell number	Camptothecin	0.56
Cell Cycle	DNA content	Paclitaxel	0.54
DNA Degradation	DNA structure	Paclitaxel	0.6
Nuclear Size	Area of nuclear region	Paclitaxel	0.63
DNA Damage	Detection of p53	Camptothecin	0.43
Mitotic Arrest	Phospho-Histone-H3	Paclitaxel	0.63
Cytoskeletal Integrity	Detection of $\alpha$ -tubulin	Paclitaxel	0.3

# Correlation of BioSeek Mitochondrial Dysfunction Class with HCS Mitochondrial Function Endpoints



 Mitochondria Membrane Potential

 Mitochondria Mass

 Cytotoxicity



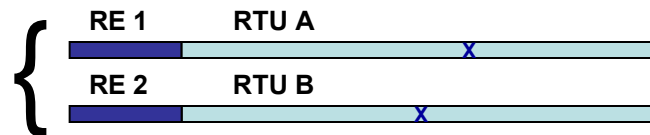
# Multiplexed Reporter Gene Assay (Attagene, Inc.)

- Measures activation/inhibition of transcription factors (TF)
- TF integrate signals arising from changing cellular environments and coordinate cellular response to such change
- Similar to genomics but many fewer TF than genes
- Compounds with similar mechanism of toxicity should bear similar patterns
- Patterns should reflect the changes that precede or accompany the compounds' toxicity
- Use signatures for prediction of toxicological outcomes of compounds

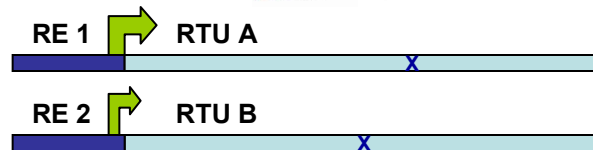
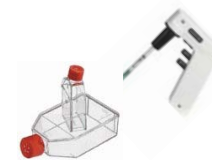


# Multiplexed Reporter Gene Assay

Library of RTUs



Cell Transfection

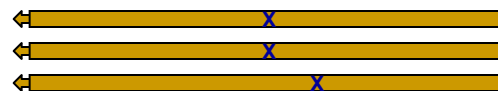


Transcription

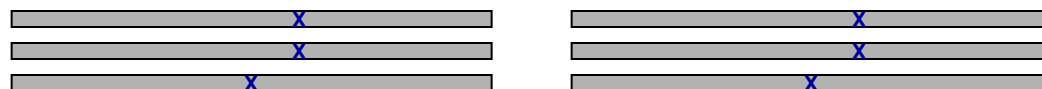
RNA Isolation



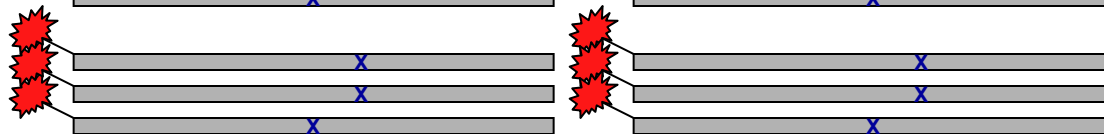
Reverse transcription



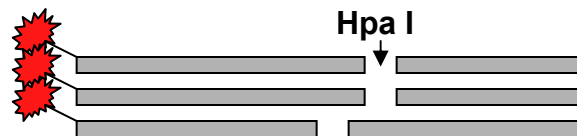
PCR amplification



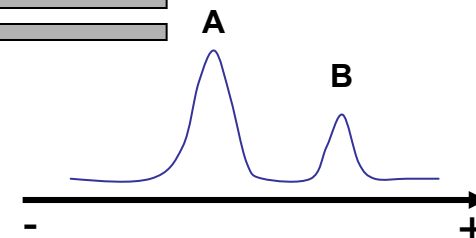
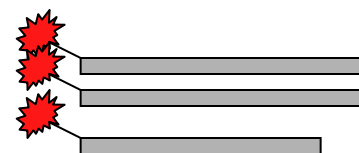
Labeling



Processing (Hpa I)



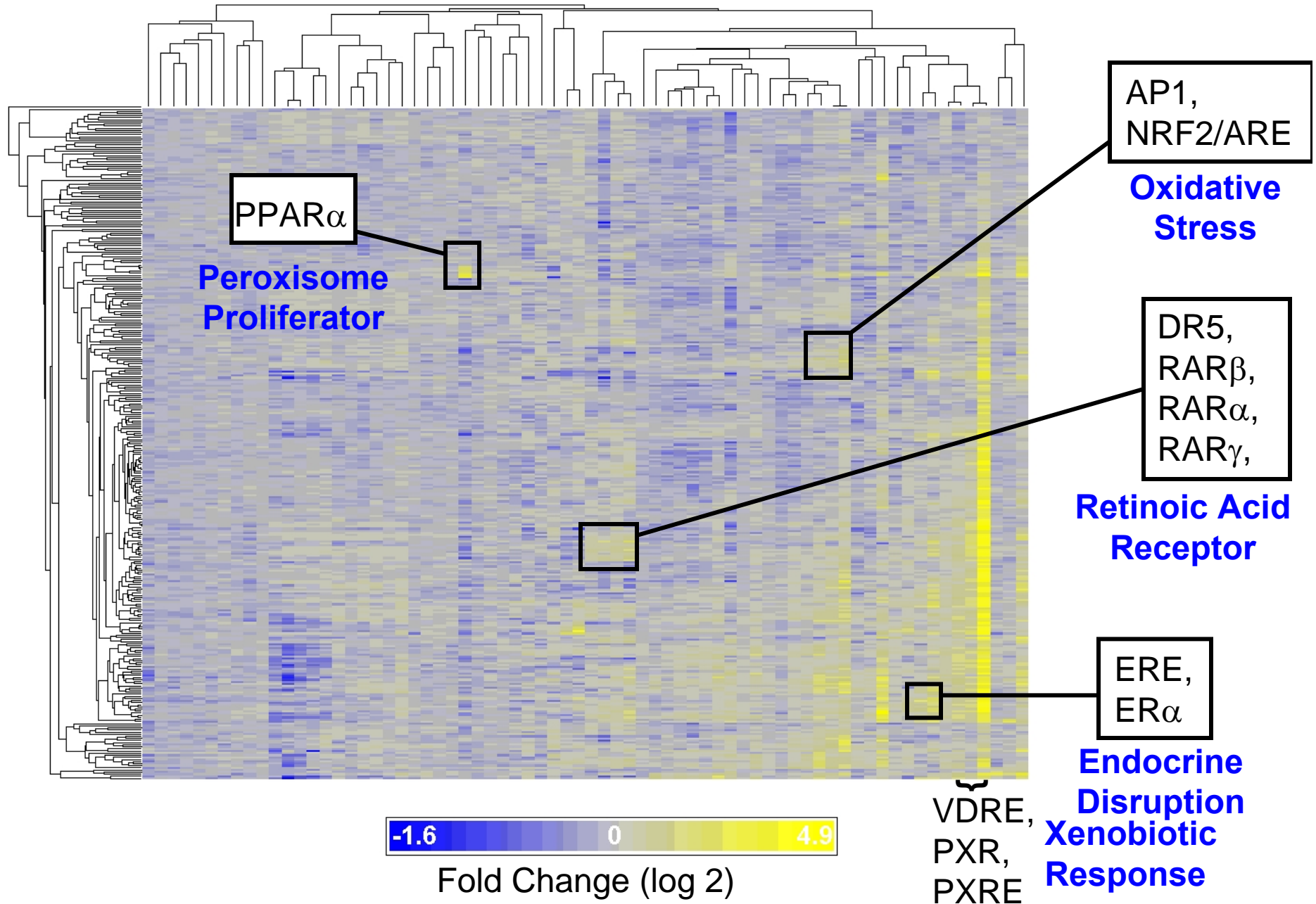
Separation and detection  
(capillary electrophoresis)





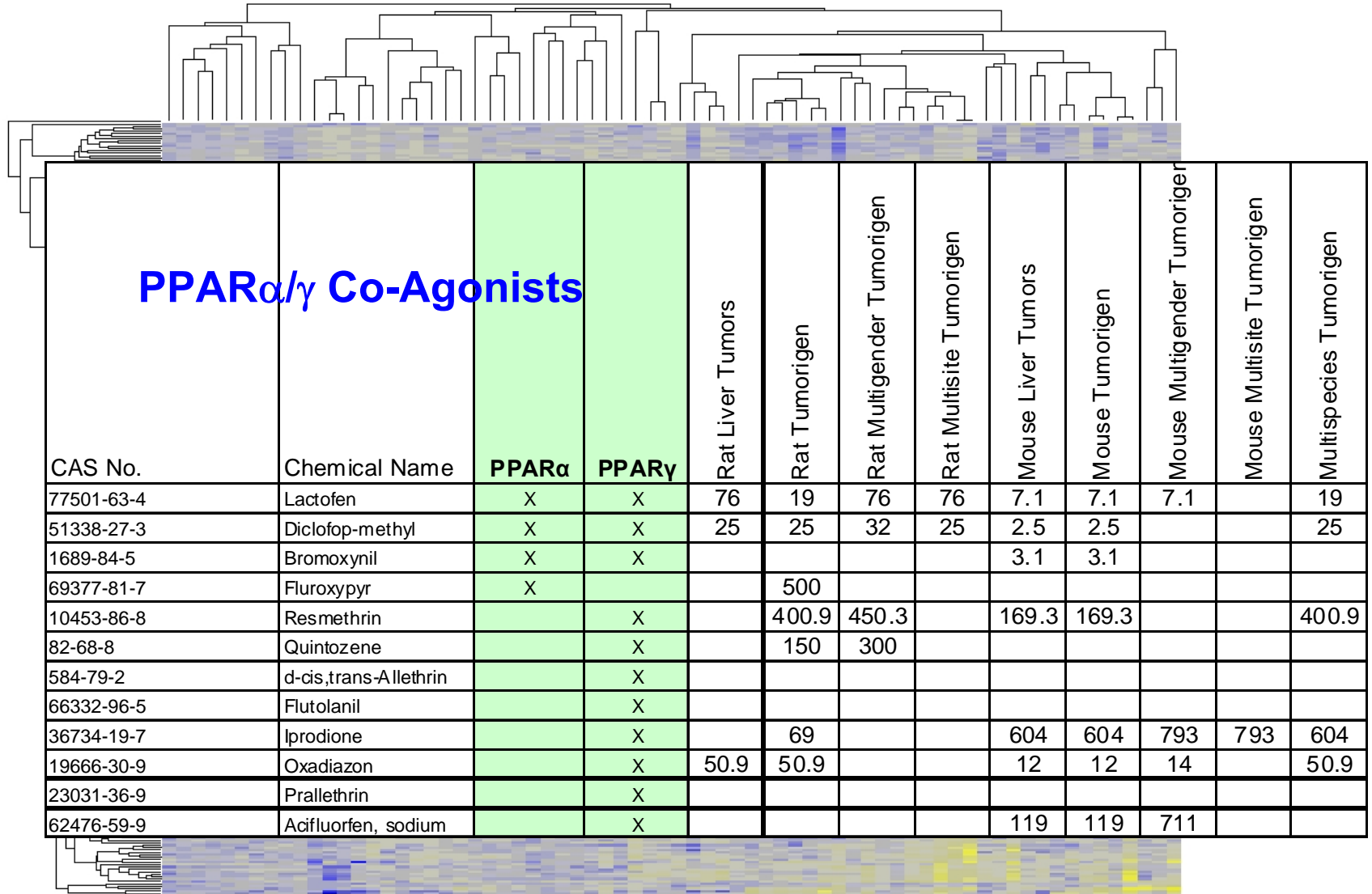


# Hierarchical Cluster Attagene Results





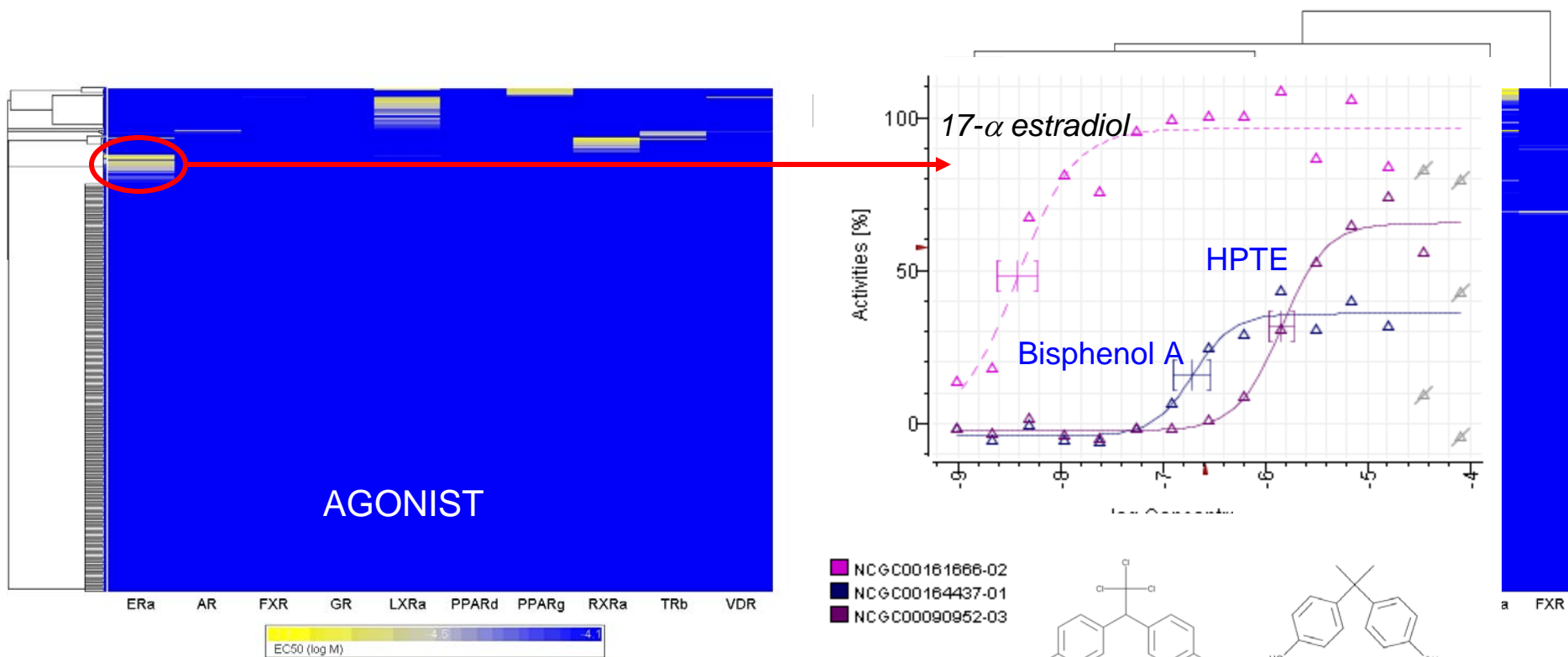
# Hierarchical Cluster Attagene Results

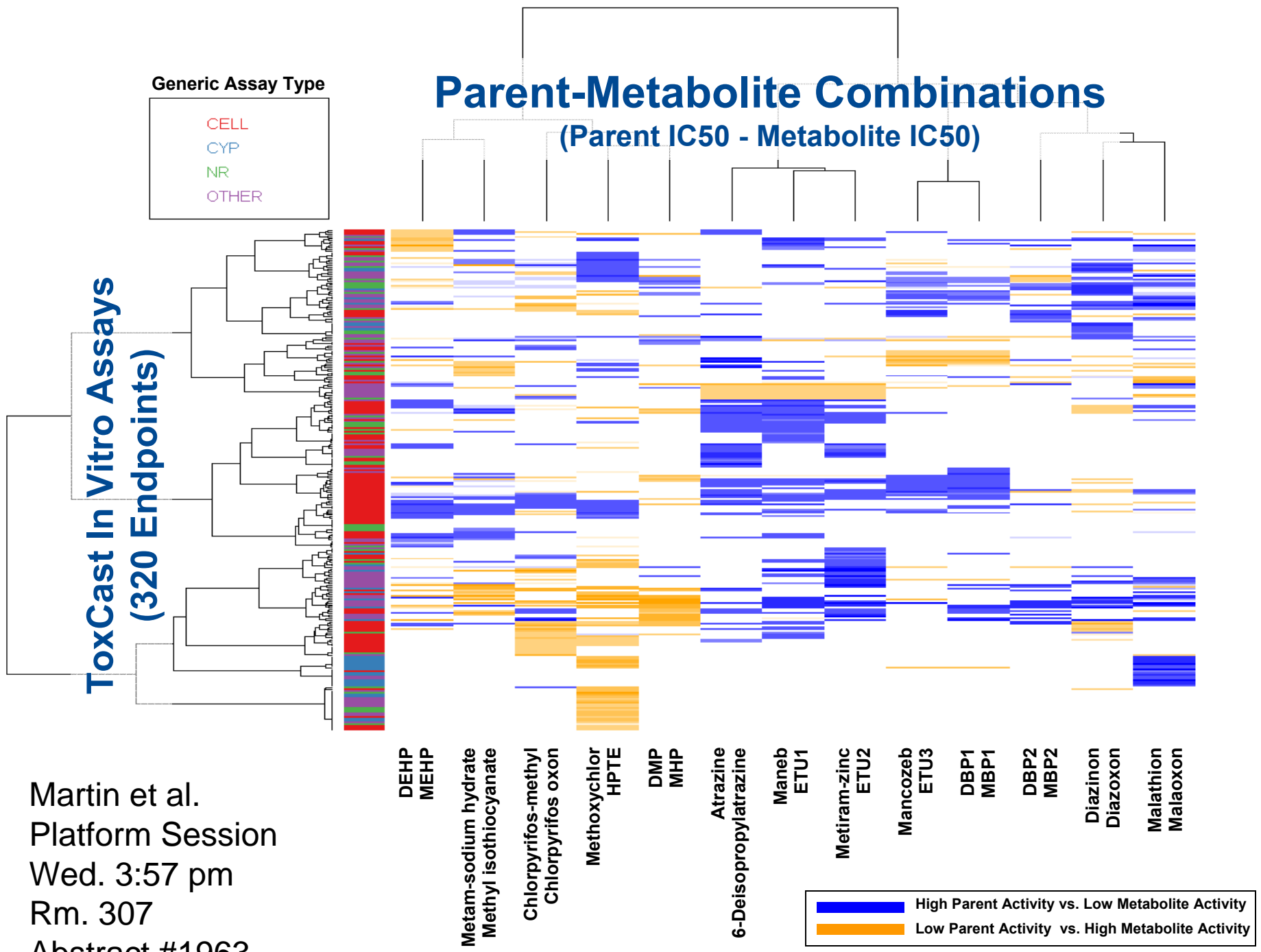


Fold Change (log 2)

# Nuclear Receptor Screening NIH Chemical Genomics Center (NCGC)

- 10 Nuclear Receptors (more in queue)
- Cellular Reporter Assays
- Agonist and Antagonist modes
- Concentration-Response Format (15 conc)
- 1462 Compounds (ToxCast 320 is a subset)



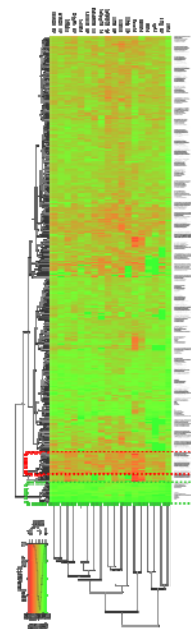
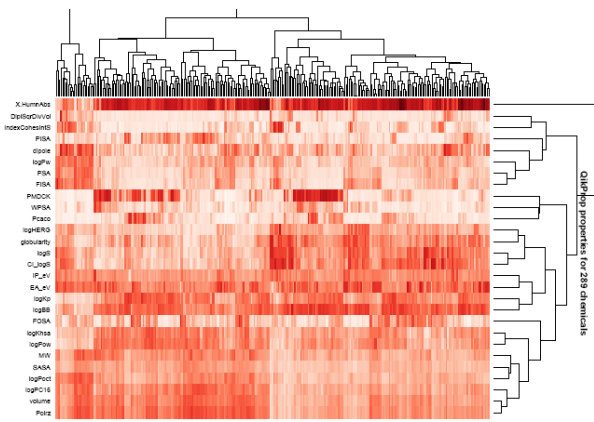
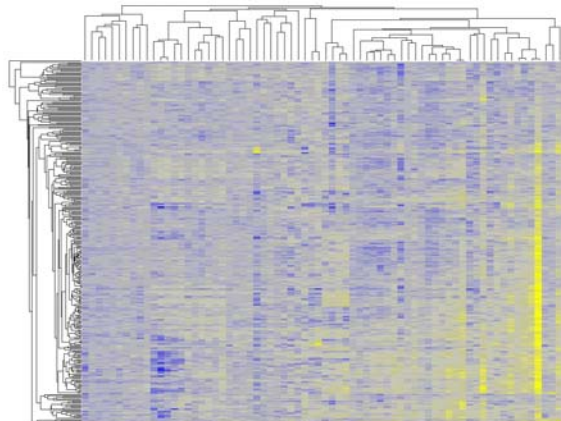


Martin et al.  
 Platform Session  
 Wed. 3:57 pm  
 Rm. 307  
 Abstract #1963

# ToxCast Phase I Data Status

- Data collection complete
- Finishing data prep for analysis (QC, normalization, etc.)
- Data sets to be made available to public through manuscript publications
- Methods for correlative analysis being tested on limited data sets
- Data Summit
  - RTP, NC May 14-15

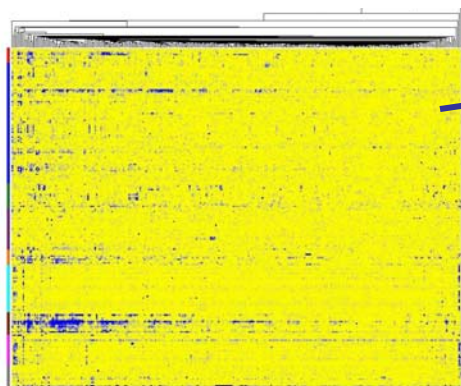
# ToxCast Data Analysis



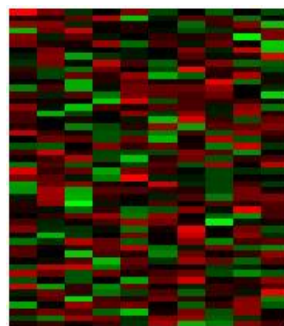
Physical chemical properties

In silico Predictions

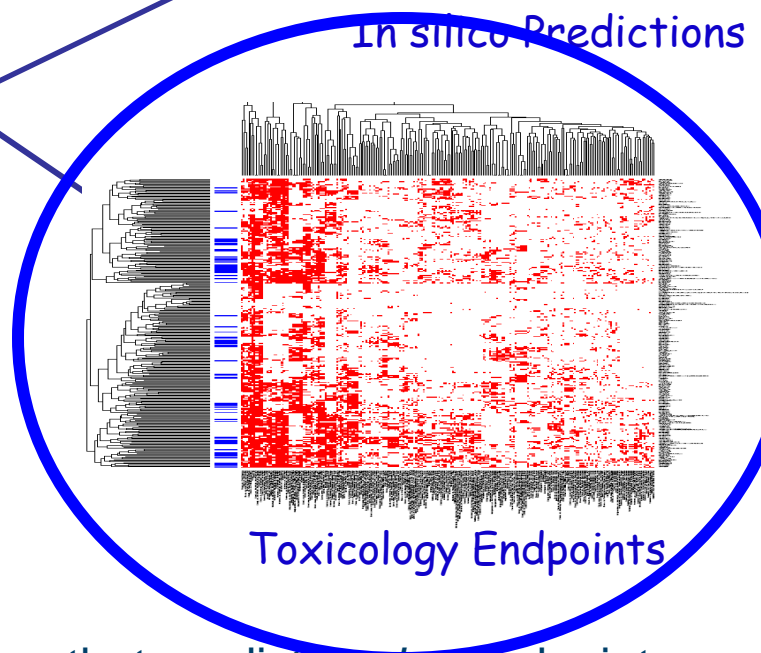
Profile Matching



Biochemical Assays



Genomic Signatures



Toxicology Endpoints

Find "Signatures" from *in vitro* & *in silico* assays that predict *in vivo* endpoints.



## National Center for Computational Toxicology

U.S. ENVIRONMENTAL PROTECTION AGENCY

Bookmark

Contact Us Search:  All EPA  This Area  Go

You are here: [EPA Home](#) » [National Center for Computational Toxicology](#) » [Toxicology Reference Database \(ToxRefDB\)](#)

### ToxRefDB Program Toxicology Reference Database

ToxRefDB was developed by the National Center for Computational Toxicology (NCCT) in partnership with EPA's Office of Pesticide Programs (OPP), to store data from in vivo animal toxicity studies. The initial focus was populating ToxRefDB with pesticide registration toxicity data that has been historically stored as hard-copy and scanned documents by OPP. A significant portion of these data have now been processed into ToxRefDB in a standardized and structured format. ToxRefDB currently includes chronic, cancer, sub-chronic, developmental, and reproductive studies on hundreds of chemicals, many of which are pesticide active ingredients. These data are now accessible and computable within ToxRefDB, and are serving as reference toxicity data for ORD research and OPP retrospective analyses. The primary research application of ToxRefDB is to provide toxicity endpoints for the development of ToxCast™ predictive signatures.

Data Set	Description	Download	Publication
Data Entry Tool & Controlled Vocabulary	The Data Entry Tool provided the user interface for all initial data input into ToxRefDB. The controlled vocabulary standardized the capturing of regulatory animal toxicity studies performed across various study types.	<a href="#">Download</a> (15.5 MB, ZIP)	Martin et al. (2008) " <a href="#">Profiling Chemicals Based on Chronic Toxicity Results from the U.S. EPA ToxRef Database</a> " Environmental Health Perspectives doi:10.1289/ehp.0800074
Chronic & Cancer Endpoints	Based on incidence, severity and potency, 26 primarily tissue-specific pathology endpoints were selected to uniformly classify 310 chemicals included in the manuscript's analysis. The 310 chemicals in this analysis largely overlap with the 320 ToxCast Phase I chemicals.	<a href="#">Download</a> (2.7 MB, XLS)	Martin et al. (2008) " <a href="#">Profiling Chemicals Based on Chronic Toxicity Results from the U.S. EPA ToxRef Database</a> " Environmental Health Perspectives doi:10.1289/ehp.0800074

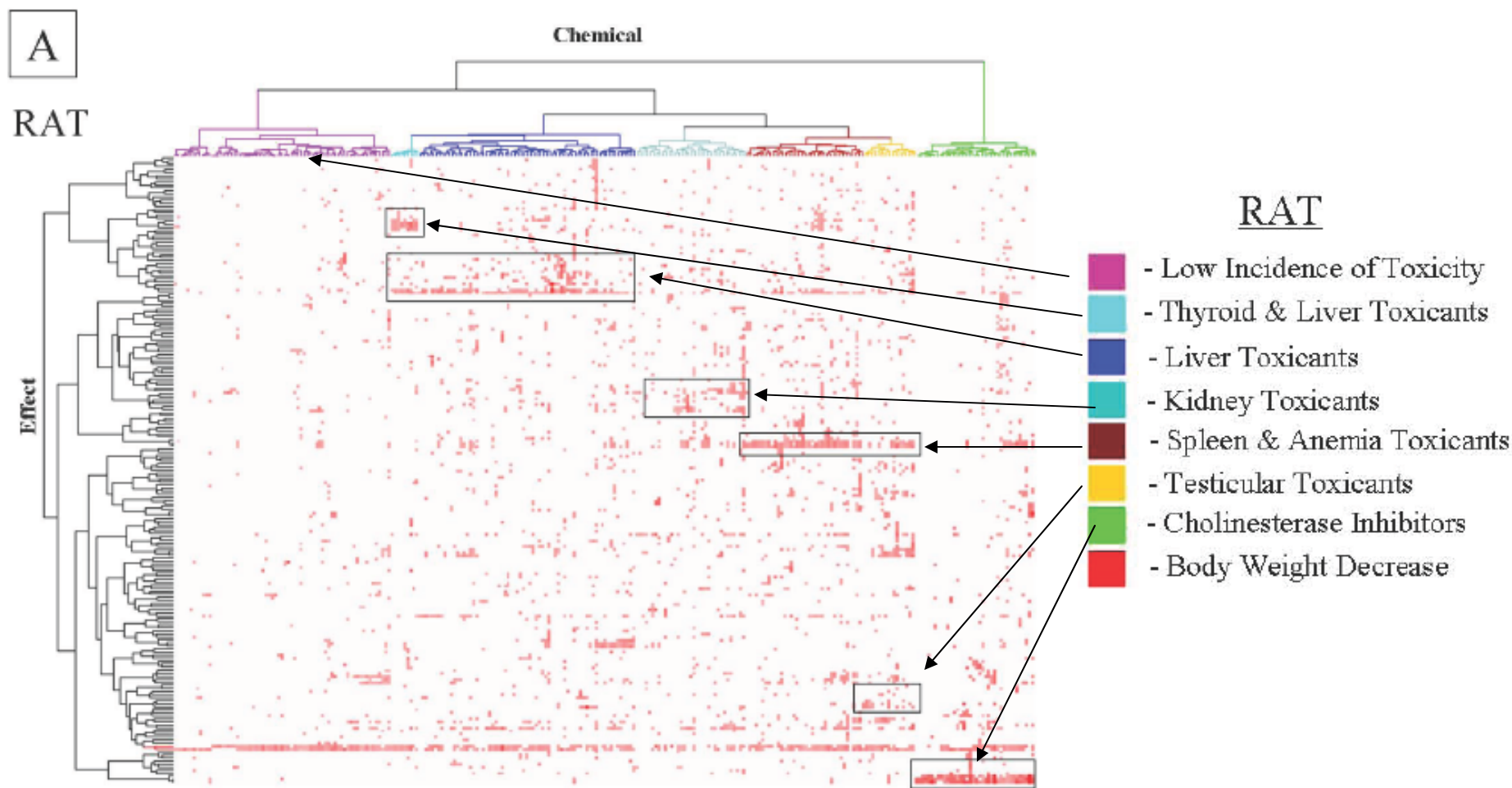
[EPA Home](#) | [Privacy and Security Notice](#) | [Contact Us](#)

Last updated on Tuesday, October 21st, 2008.

[Print As-Is](#)

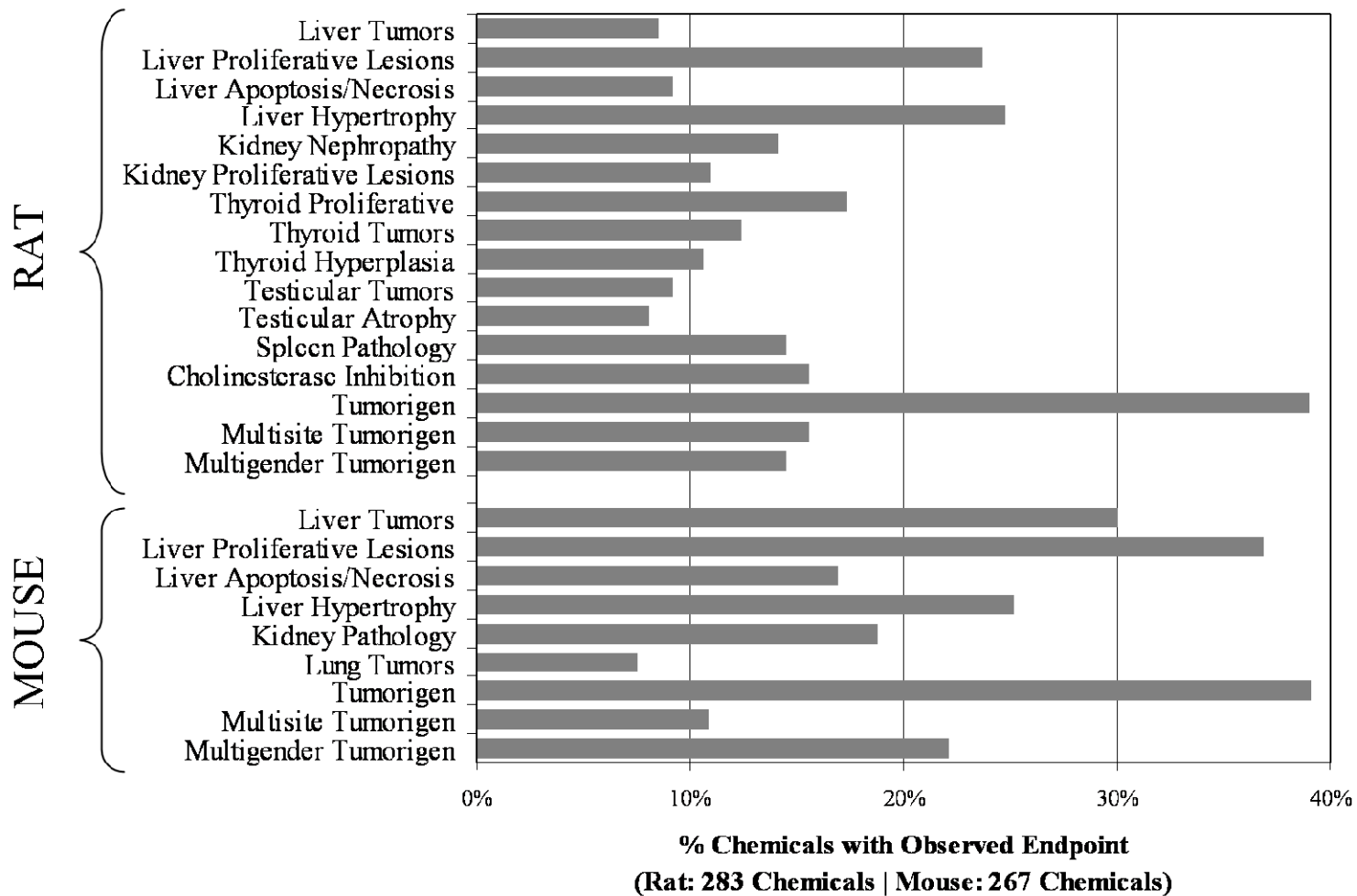
ToxRefDB website: <http://www.epa.gov/ncct/toxrefdb/>

# \$1B in Toxicology Now Stored in ToxRefDB

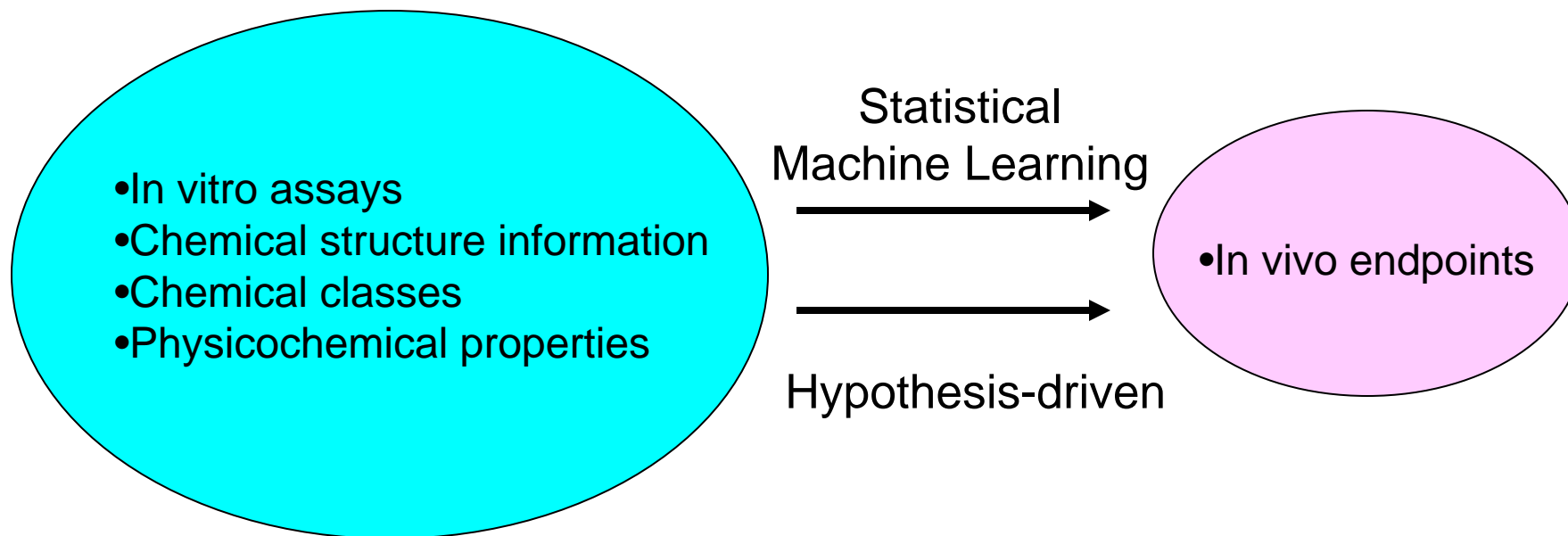




# Selected Chronic Rat & Mouse Endpoints for Predictive Modeling

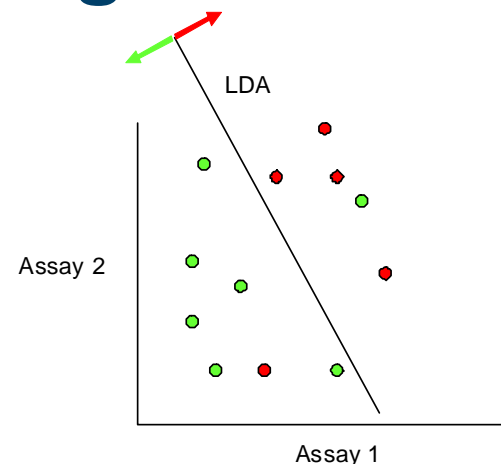


# ToxCast Analysis Approaches



# Association Analysis / Signatures

- Use Machine Learning methods
  - SLR: Stepwise Logistic Regression
  - LDA: Linear Discriminant Analysis
  - SVM: Support Vector Machines
  - Many others
- For each binary endpoint, build models of form
  - $Predictor = F(\text{assay values})$
  - If
    - $Predictor$  for a chemical meets criteria
  - Then
    - Predict endpoint to be positive for the chemical

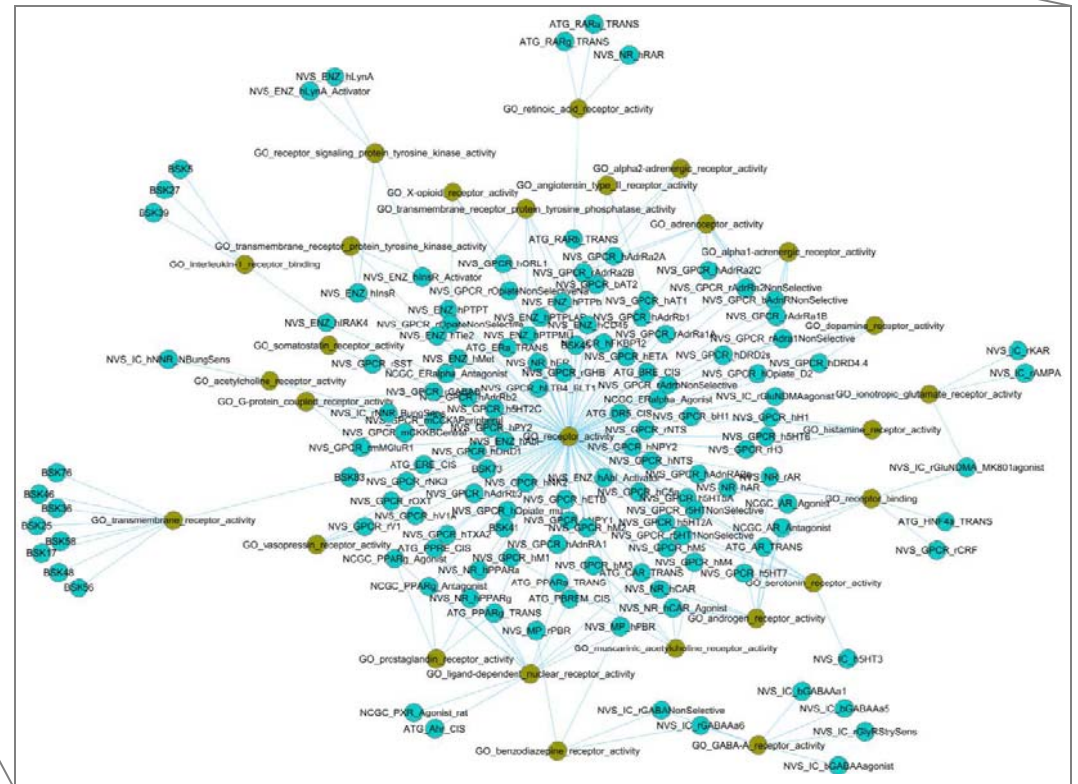


		Truth	
		+	-
Test	+	TP	FP
	-	FN	TN

Assay	Gene	Endpoint	TP	FP	FN	TN	Sens	Spec	BA	Rel Risk	p(dichot)
GENE_Ache_Rattus_norvegicus	Ache	CHR_Rat_CholinesteraseInhibition	14	4	31	208	0.31	0.98	0.65	6	7.00E-09
GENE_HTR2C_Homo_sapiens	HTR2C	CHR_Rat_KidneyProliferativeLesions	6	3	25	223	0.19	0.99	0.59	6.61	1.30E-04
logP_TPSA		CHR_Rat_LiverHypertrophy	57	109	9	82	0.86	0.43	0.65	3.47	1.10E-05
PATH_KEGG_hsa00980_Metabolism_of_xenobiotics_by_cytochrome_P450		CHR_Rat_LiverNecrosis	20	101	3	133	0.87	0.57	0.72	7.49	4.70E-05
PATH_KEGG_hsa00591_Linoleic_acid_metabolism		CHR_Rat_LiverNecrosis	20	102	3	132	0.87	0.56	0.72	7.38	9.50E-05
GENE_AR_Homo_sapiens	AR	CHR_Rat_LiverNecrosis	16	65	7	169	0.7	0.72	0.71	4.97	9.70E-05
ATG_PPARG_TRANS	PPARG	CHR_Rat_LiverNecrosis	9	19	14	215	0.39	0.92	0.66	5.26	1.70E-04
NCGC_ERalpha_Antagonist	ESR1	CHR_Rat_LiverNecrosis	11	32	12	202	0.48	0.86	0.67	4.56	2.50E-04
BSK_SM3C_Thrombomodulin	THBD	CHR_Rat_LiverNecrosis	8	18	15	216	0.35	0.92	0.64	4.74	6.60E-04
GENE_PLAT_Homo_sapiens	PLAT	CHR_Rat_LiverProliferativeLesions	14	8	52	183	0.21	0.96	0.59	2.88	9.30E-05
ATG_PPARG_TRANS	PPARG	CHR_Rat_LiverProliferativeLesions	16	12	50	179	0.24	0.94	0.59	2.62	1.70E-04
OxidativeStress_24hr	H2AFX	CHR_Rat_LiverProliferativeLesions	13	8	53	183	0.2	0.96	0.58	2.76	2.70E-04
GENE_CYP3A4_Homo_sapiens	CYP3A4	CHR_Rat_LiverProliferativeLesions	10	4	56	187	0.15	0.98	0.57	3.1	2.90E-04
ATG_PPARG_TRANS	PPARG	CHR_Rat_LiverTumors	10	18	13	216	0.43	0.92	0.68	6.29	2.10E-05
CellLoss_72hr		CHR_Rat_ThyroidProliferativeLesions	34	109	9	105	0.79	0.49	0.64	3.01	6.80E-04



# ToxCast Assays Associated with Gene Ontology (GO) Receptor Activity Processes



Holly M. Mortensen, *et al.*

Poster Presentation: Tuesday 1-4:30

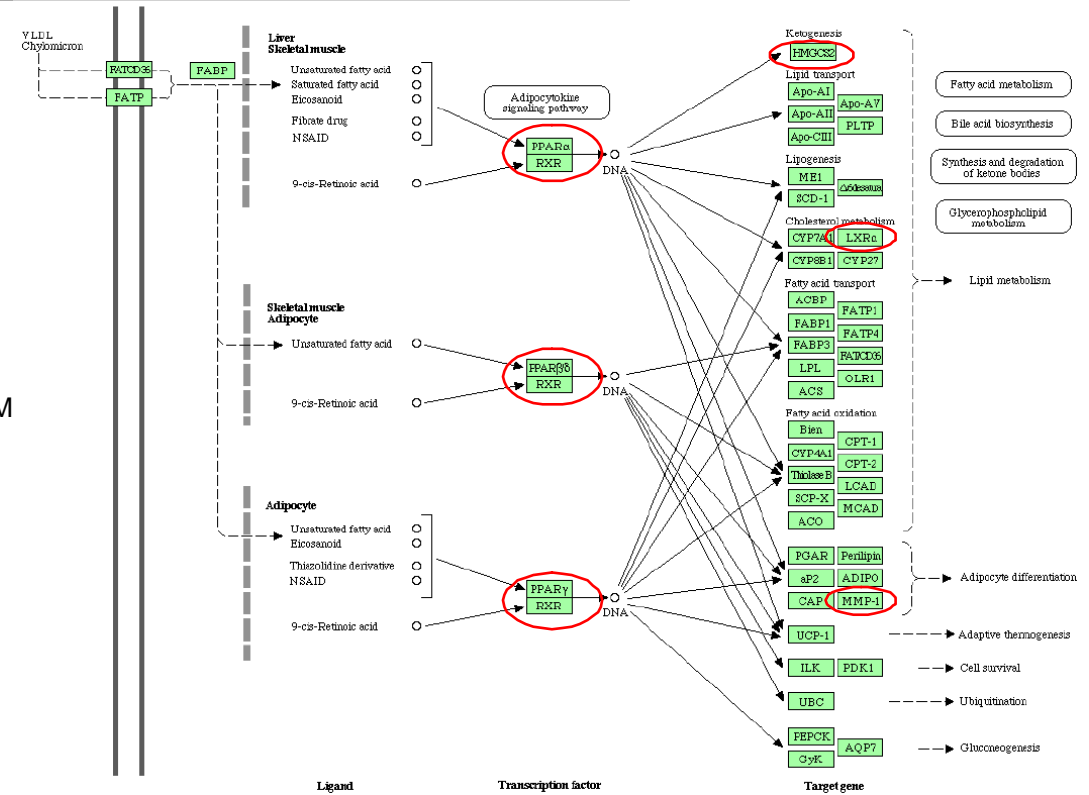
Abstract #1090; Poster Board #308

# Data Analysis: A Pathway-Based Approach

## ToxCast Assays in KEGG PPAR Signaling Pathway

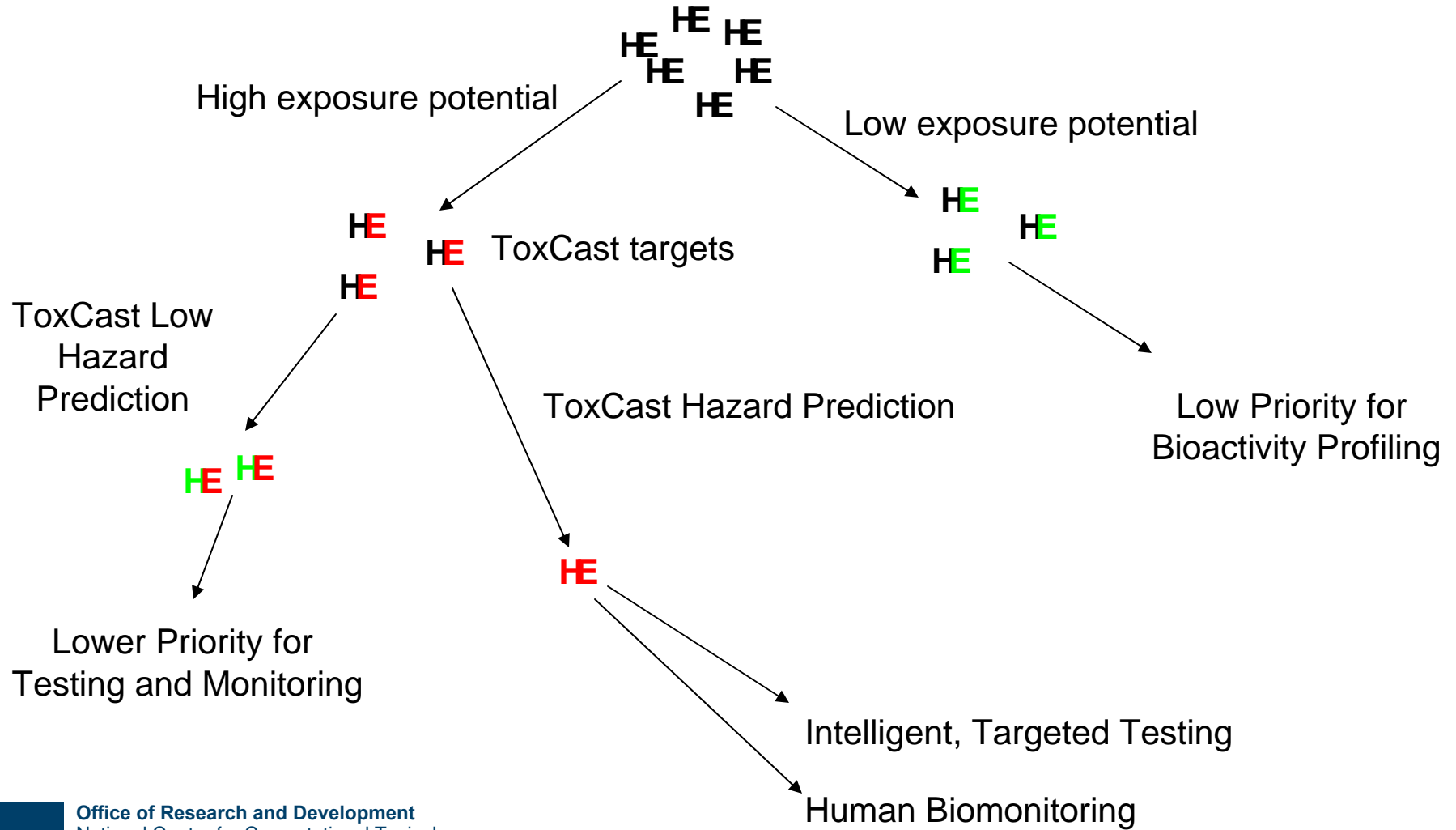
- Attagene Factorial cis DR4/LXRE
- Attagene Factorial cis PPRE
- Attagene Factorial cis PPRE
- Attagene Factorial cis PPRE
- Attagene Factorial trans LXRA
- Attagene Factorial trans PPARa
- Attagene Factorial trans PPARd
- Attagene Factorial trans PPARg
- Attagene Factorial trans RXRa
- Attagene Factorial trans RXRb
- BioSeek BrEP IL1b / TNFa / IFNg / MMP1
- BioSeek HDFn / IL1b / TNFa / IFNg / EGF / FGF / PDGFbb / M
- CellzDirect HMGCS2
- NCGC Reporter Gene Assay PPARd Agonist
- NCGC Reporter Gene Assay PPARd Antagonist
- NCGC Reporter Gene Assay PPARg Agonist
- NCGC Reporter Gene Assay PPARg Antagonist
- NCGC Reporter Gene Assay RXRa Agonist
- NCGC Reporter Gene Assay RXRa Antagonist
- Novascreen Human MMP1
- Novascreen Human PPARa
- Novascreen Human PPARg

### KEGG PPAR Signaling Pathway



Endpoint	TP	FP	FN	TN	Sens	Spe c	Relati ve Risk	p-value	Rule 1	Rule 2	Rule 3
Rat Thyroid Proliferative Lesions	6	24	1	226	0.86	0.90	45.4	1.0E-05	SM3C Proliferation AND KEGG(Focal adhesion) AND KEGG(Thyroid_cancer)	ACEA(Cell growth) AND KEGG(Complement and coagulation cascades)	
Rat Liver Proliferative Lesions	1	74	0	173	1	0.7		9.5E-06	NRF2 AND GABPA AND Bioavailability	PPARG AND KEGG(Limonene and pinene degradation) AND KEGG(Chronic myeloid leukemia)	MCSE AND KEGG(PPAR signaling) AND KEGG(Cell adhesion molecules)

# The Future State: Using Hazard and Exposure Information for Prioritizing Testing and Monitoring



# The ToxCast Team

