Finding patterns in a complex in vitro data set: insights from the ToxCast Project

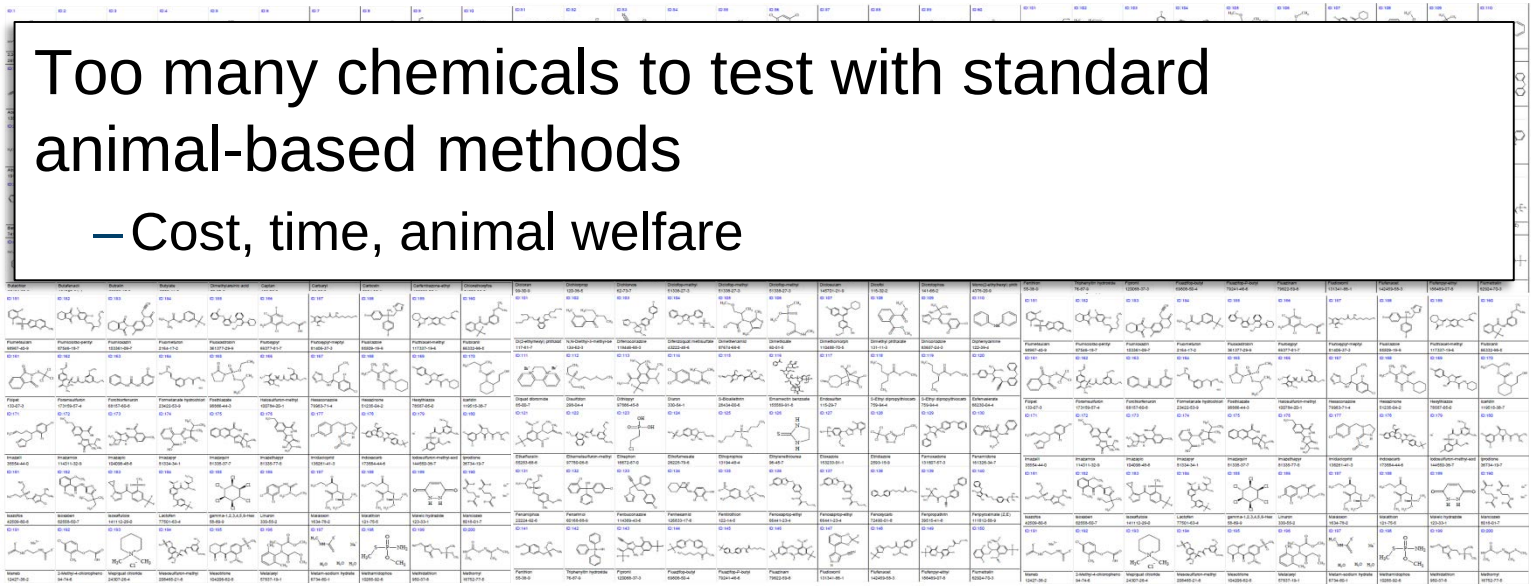
Richard Judson
U.S. EPA, National Center for Computational Toxicology
Office of Research and Development

BASF, RTP February 2014

Problem Statement

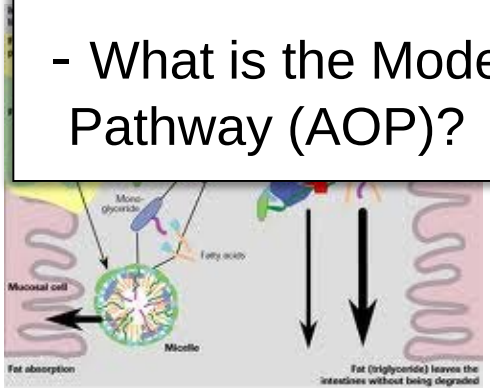
Too many chemicals to test with standard animal-based methods

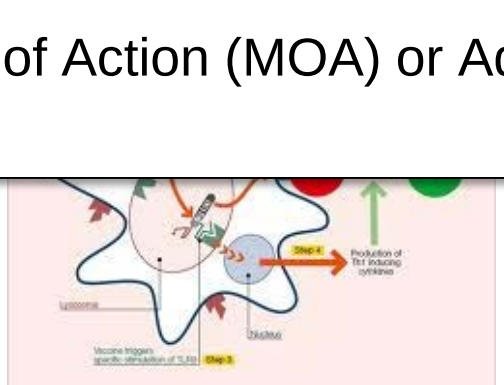
- Cost, time, animal welfare

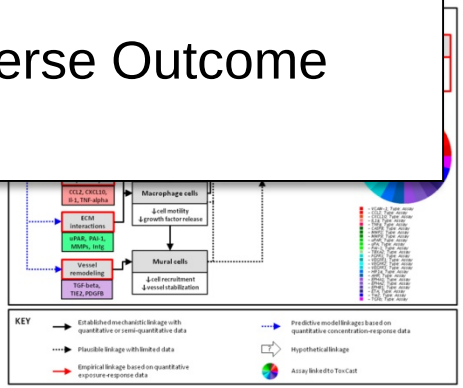


Need for better mechanistic data

- Determine human relevance
- What is the Mode of Action (MOA) or Adverse Outcome Pathway (AOP)?







ToxCast Overall Strategy

- Develop high-throughput, *in vitro* assays for “pathways” linked to toxicity
- Develop predictive hazard models
  - *in vitro, in silico → in vivo*
- Develop high-throughput exposure predictions
- Use models:
  - Prioritize chemicals for targeted testing
  - Distinguish possible AOP for chemicals (human relevance)
  - High Throughput Risk Assessments (semi-quantitative)

High-Throughput Screening Assays

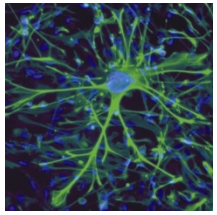
batch testing of chemicals for pharmacological/toxicological endpoints using automated liquid handling, detectors, and data acquisition

LTS MTS HTS uHTS

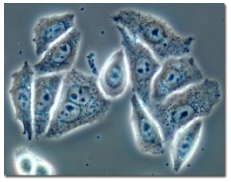
10s-100s/yr

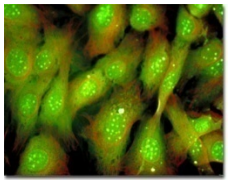
Gene-expression

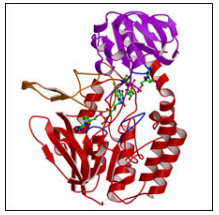
10s-100s/day



1000s/day







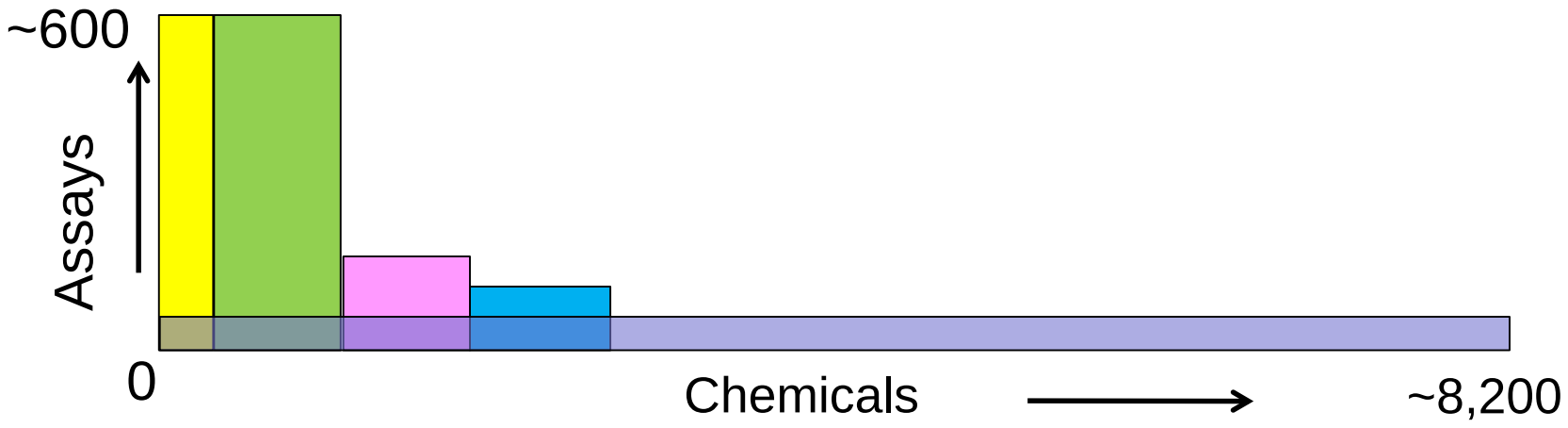
10,000s-
100,000s/day

Human Relevance/
Cost/Complexity

Throughput/
Simplicity

Testing under ToxCast and Tox21 Chemicals, Data and Release Timelines

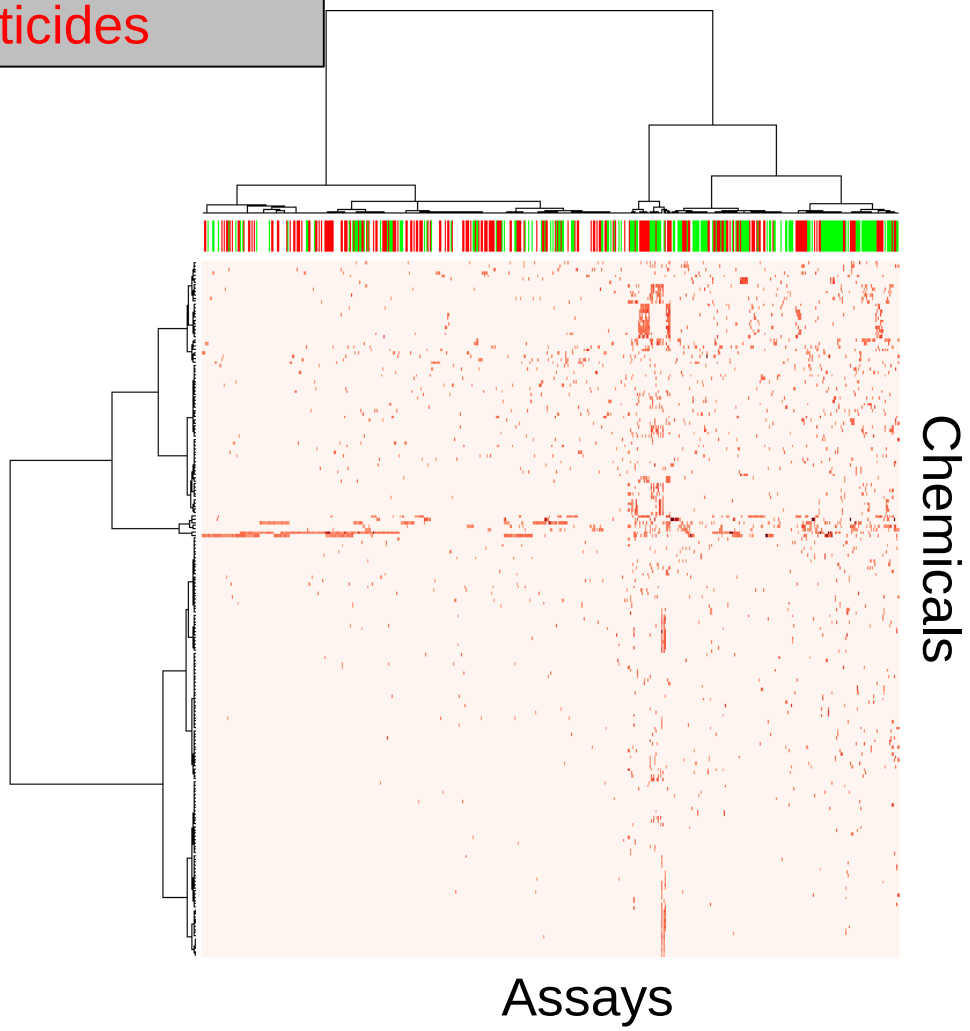
Set	Chemicals	Assays	Endpoints	Completion	Available
ToxCast Phase I	293	~600	~700	2011	Now
ToxCast Phase II	767	~600	~700	03/2013	10/2013
ToxCast Phase IIIa	1001	~100	~100	Just starting	2014
E1K (endocrine)	880	~50	~120	03/2013	10/2013
Tox21	8,193	~25	~50	Ongoing	Ongoing

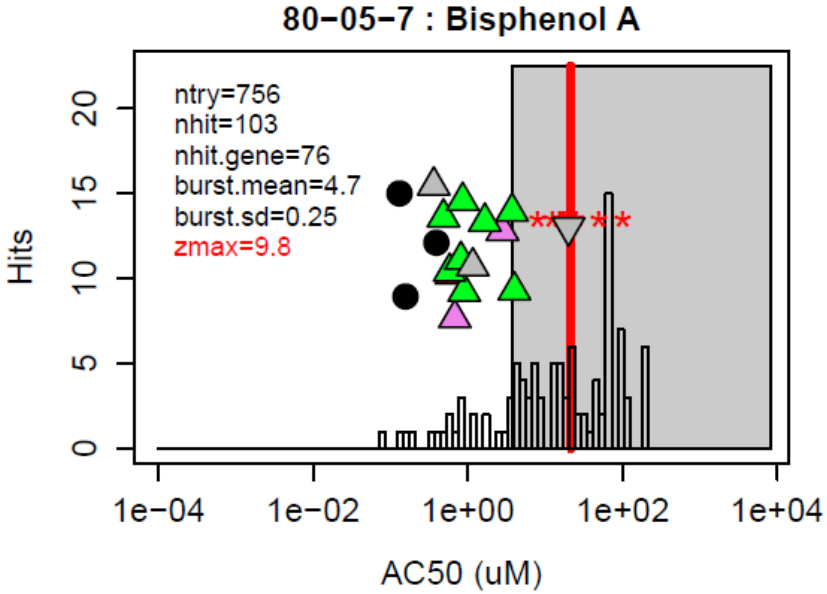


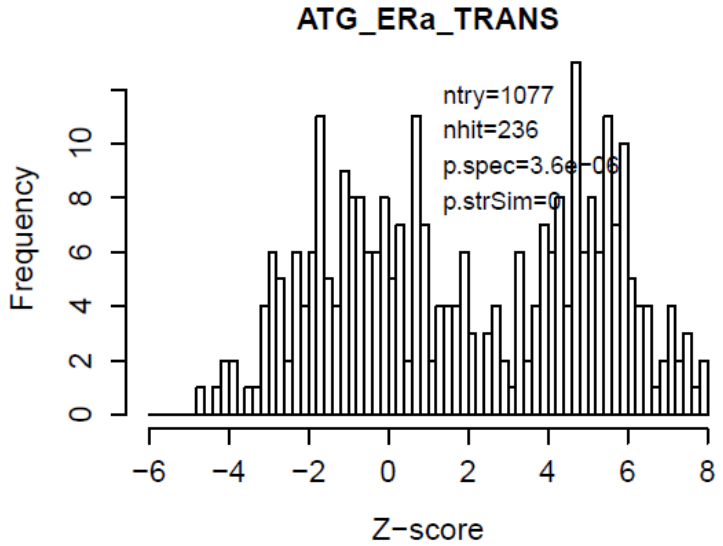
ToxCast Phase II:
1051 Chemicals x 790 Assay Readouts

United States Environmental Protection Agency

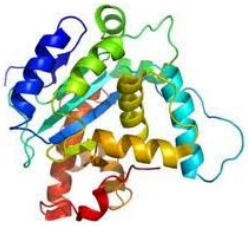
Pharmaceuticals
Pesticides

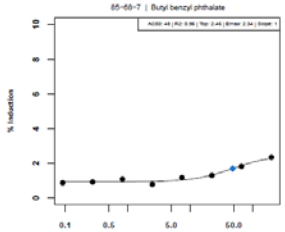


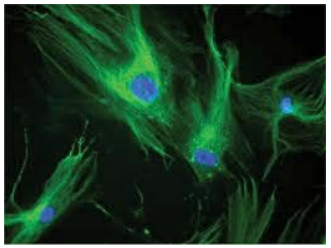


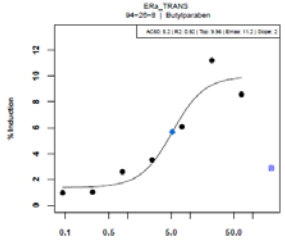


Model or Signature Generation



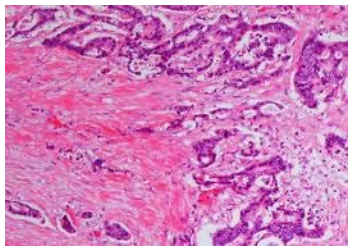


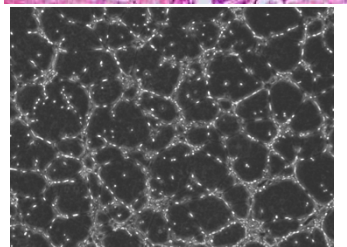


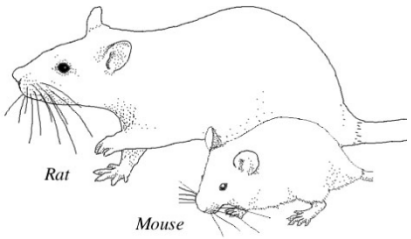


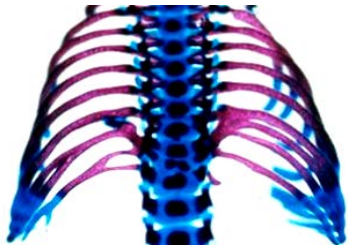
In Vitro Data
ToxCastDB

Predictive Models – “Signatures”







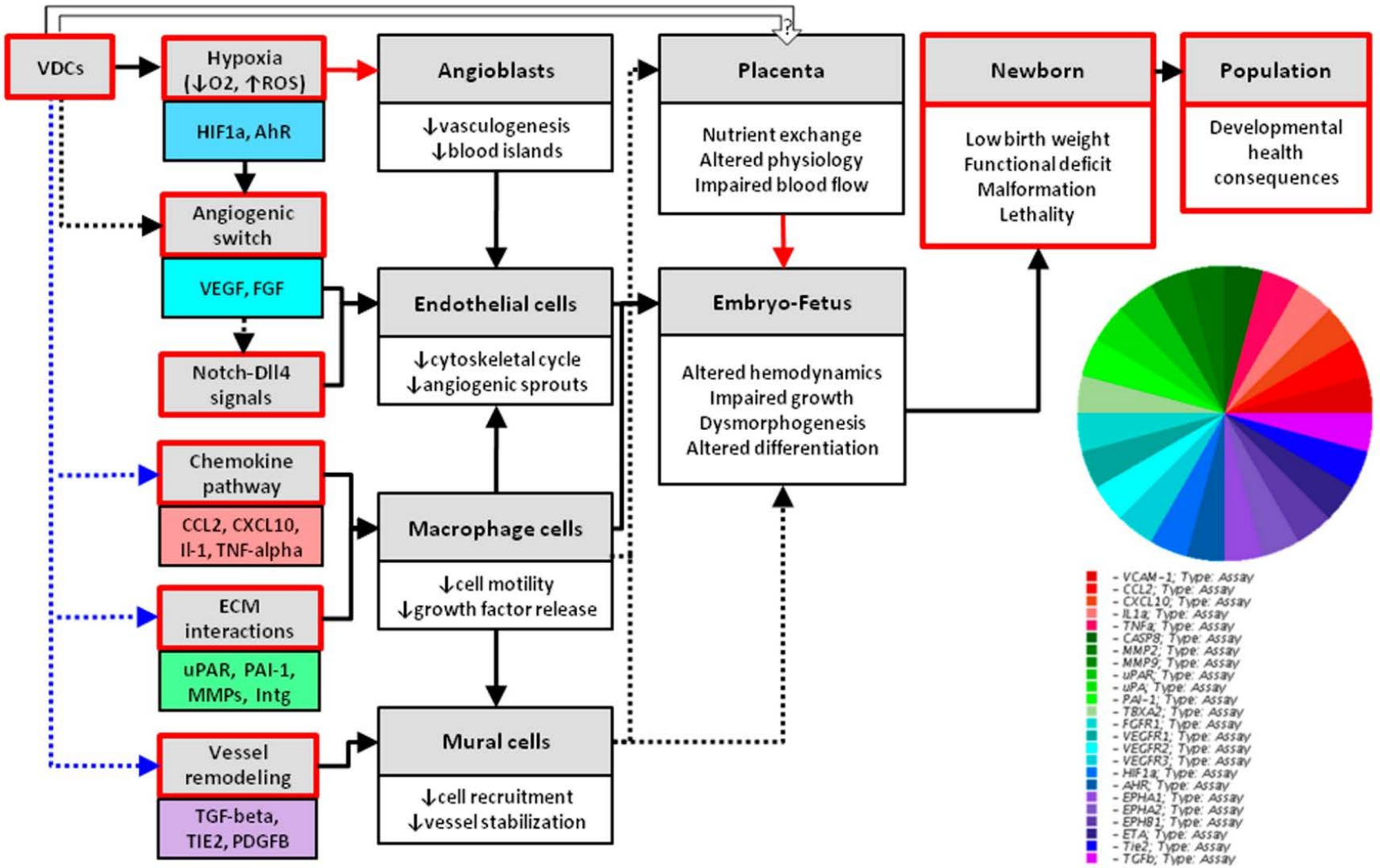


In Vivo Data
ToxRefDB

Adverse Outcome Pathway Approach

United States Environmental Protection Agency

Proposed AOP for developmental changes linked to embryonic vascular disruption



SOURCE: Knudsen and Kleinstreuer (2011) Birth Defects Res. C

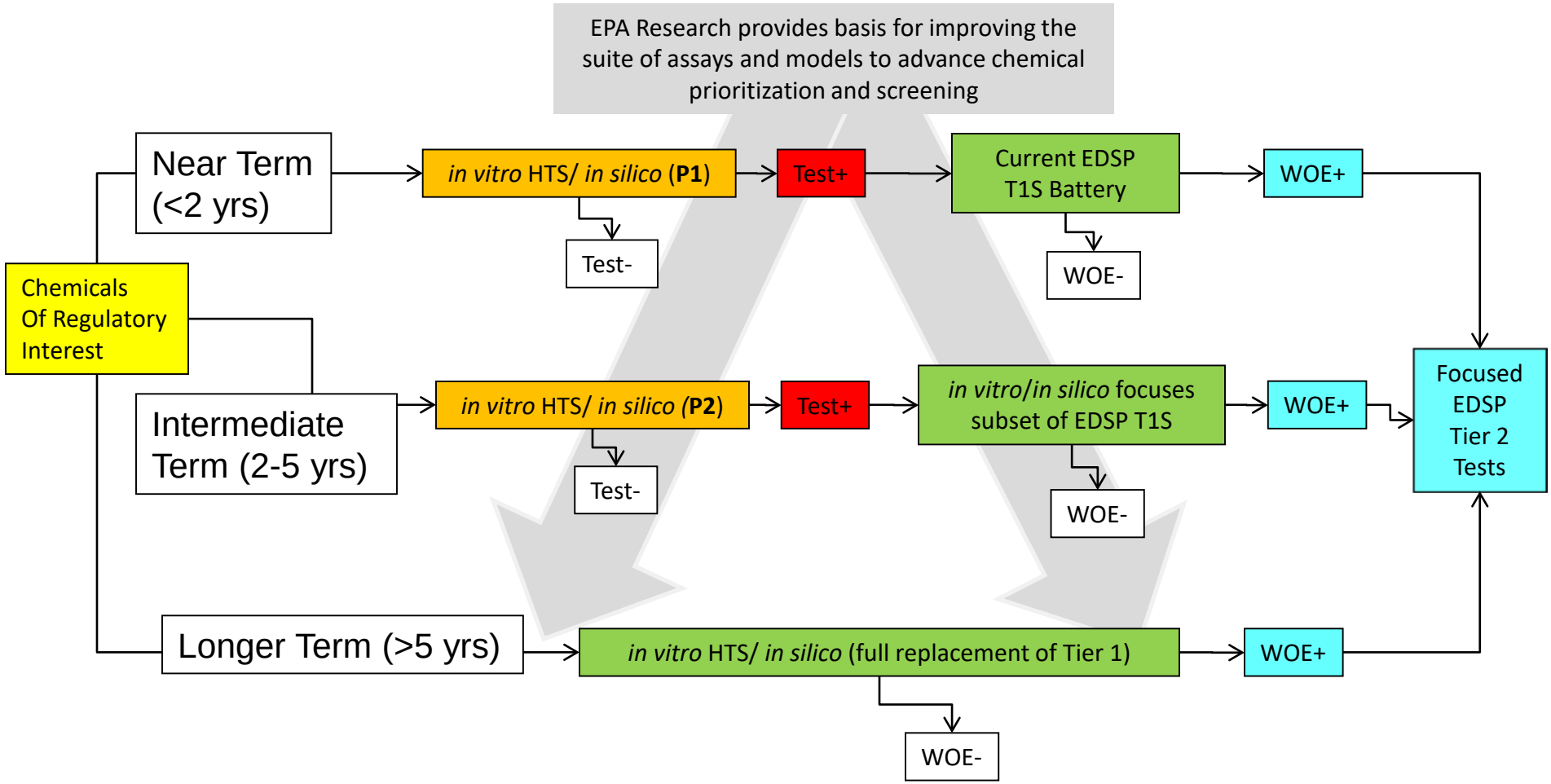
EDSP: A First, Real-World Application of ToxCast

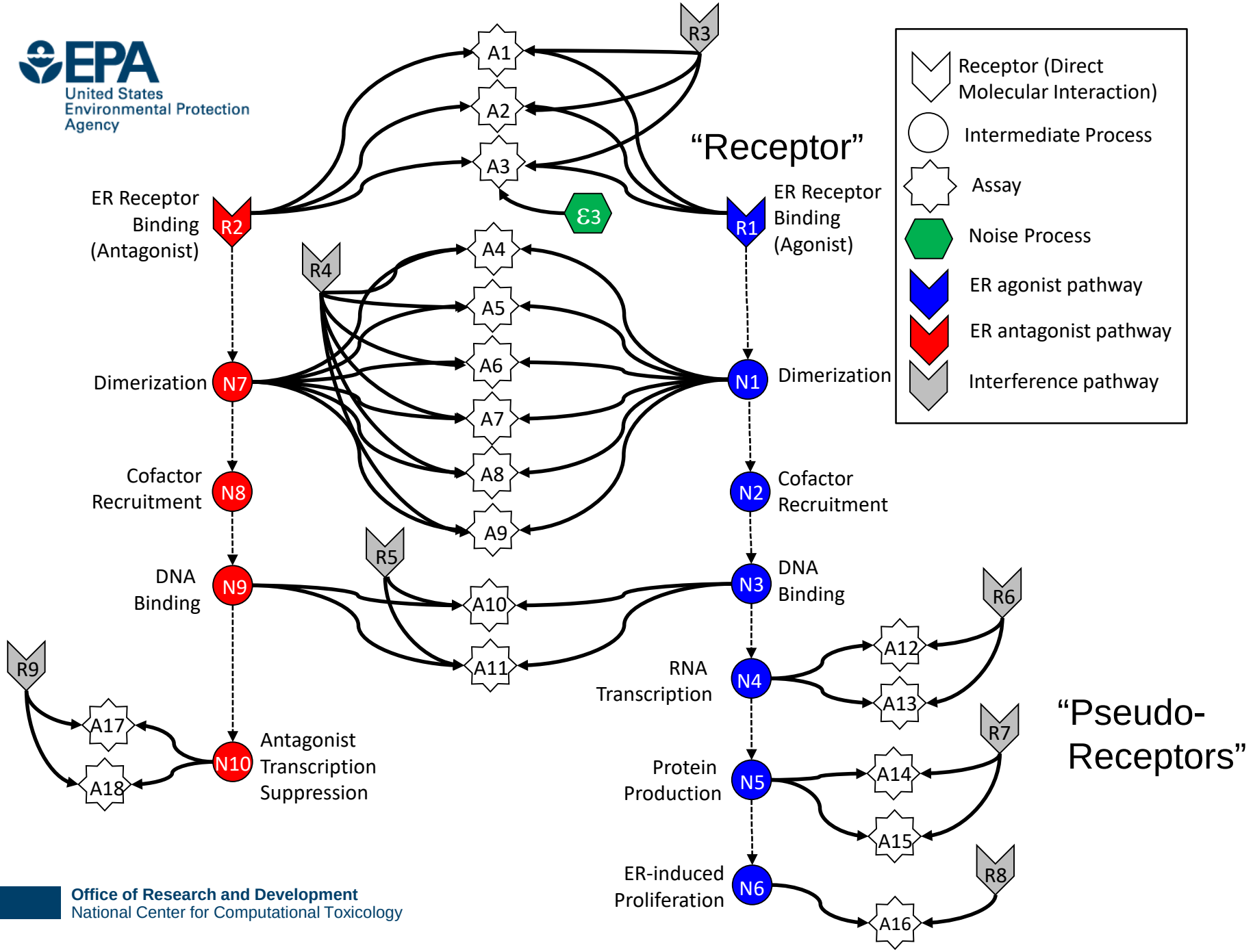
Prioritization for Endocrine Disruptor Screening Program

~5000 Chemicals are required to go through EDSP Tier 1 battery

Throughput: ~100 Chemicals per year

Cost: ~$1M per chemical




EPA
United States
Environmental Protection
Agency
"Receptor"
"Pseudo-
Receptors"
Receptor (Direct
Molecular Interaction)
Intermediate Process
Assay
Noise Process
ER agonist pathway
ER antagonist pathway
Interference pathway
ER Receptor
Binding
(Antagonist)
ER Receptor
Binding
(Agonist)
Dimerization
Dimerization
Cofactor
Recruitment
Cofactor
Recruitment
DNA
Binding
DNA
Binding
RNA
Transcription
Protein
Production
ER-induced
Proliferation
Antagonist
Transcription
Suppression
R1
R2
R3
R4
R5
R6
R7
R8
R9
A1
A2
A3
ε3
A4
A5
A6
A7
A8
A9
A10
A11
A12
A13
A14
A15
A16
A17
A18
N1
N2
N3
N4
N5
N6
N7
N8
N9
N10
Office of Research and Development
National Center for Computational Toxicology

United States Environmental Protection Agency

ID	Assay Name	Source	Gene	Species	Type
1	NVS bovine ER	Novascreen	ESR1	Bos taurus	Receptor Binding
2	NVS human ER	Novascreen	ESR1	Homo sapiens	Receptor Binding
3	NVS mouse ERa	Novascreen	Esr1	Mus musculus	Receptor Binding
4	OT ERa-ERa (8 h)	Odyssey Thera	ESR1	Homo sapiens	Dimerization
5	OT ERa-ERa (24 h)	Odyssey Thera	ESR1	Homo sapiens	Dimerization
6	OT ERa-ERb (8 h)	Odyssey Thera	ESR1, ESR2	Homo sapiens	Dimerization
7	OT ERa-ERb (24 h)	Odyssey Thera	ESR1, ESR2	Homo sapiens	Dimerization
8	OT ERb-ERb (8 h)	Odyssey Thera	ESR2	Homo sapiens	Dimerization
9	OT ERb-ERb (24 h)	Odyssey Thera	ESR2	Homo sapiens	Dimerization
10	OT GFP ERa-ERE (2 h)	Odyssey Thera	ESR1, ERE	Homo sapiens	DNA Binding
11	OT GFP ERa-ERE (8 h)	Odyssey Thera	ESR1, ERE	Homo sapiens	DNA Binding
12	ATG ERa (TRANS)	Attagene	ESR1	Homo sapiens	RNA Reporter Gene
13	ATG ERE (CIS)	Attagene	ESR1	Homo sapiens	RNA Reporter Gene
14	Tox21 ERa BLA Agonist ratio	NCGC	ESR1	Homo sapiens	Reporter Gene
15	Tox21 ERa LUC BG1 Agonist	NCGC	ESR1	Homo sapiens	Reporter Gene
16	ACEA T47D (80 h)	ACEA	ESR1	Homo sapiens	Proliferation
17	Tox21 ERa BLA Antagonist ratio	NCGC	ESR1	Homo sapiens	Reporter Gene
18	Tox21 ERa LUC BG1 Antagonist	NCGC	ESR1	Homo sapiens	Reporter Gene

Computational Model

$$A_i = \sum_j F_{ij} R_j$$

$\boldsymbol{A_i}$ is the efficacy of the assay at a given concentration
$\boldsymbol{R_j}$ is the "true" efficacy which is unobservable
$\boldsymbol{F}$ links receptors to assays

$$\varepsilon^2 = \sum_i (A_i^{pred} - A_i^{meas})^2 + penalty(\vec{R})$$

Solve a constrained least-squares problem to minimize difference between the measured and predicted assay values

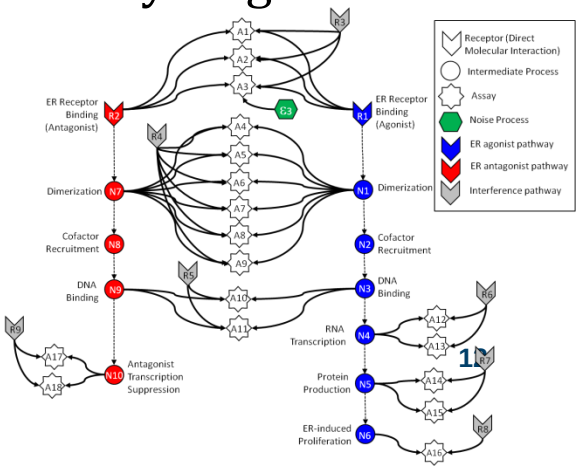
$$A_i^{pred} \in [1,0]$$

$$penalty(\vec{R}) = \alpha \frac{SR^2}{SR^2 + SR_0{}^2}$$

Penalty enforces physical assumption that chemical will not hit many targets simultaneously

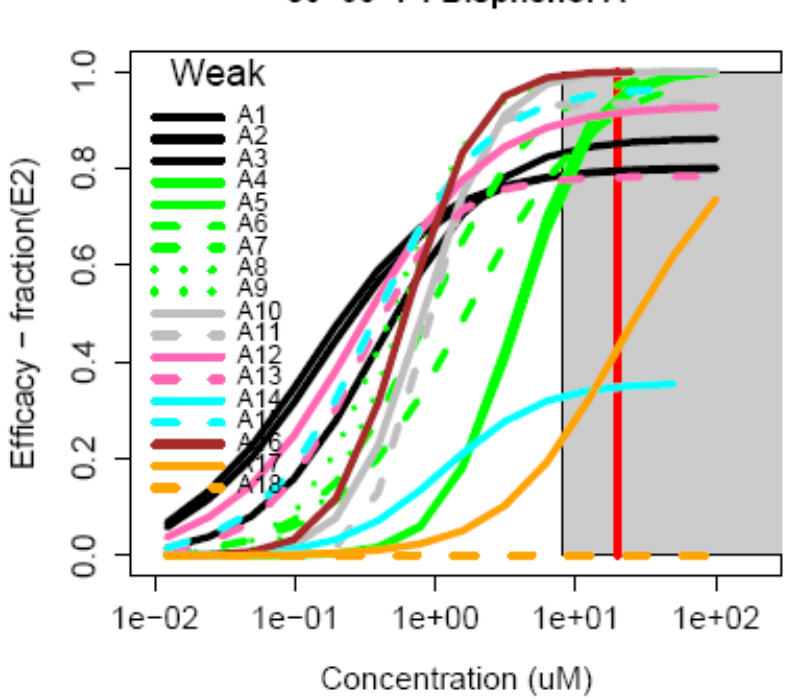
$$AUC_j = \frac{1}{N_{conc}} \sum_{i=1}^{N_{conc}} sign(slope) \times R_j(conc_i)$$

AUC Summarizes results



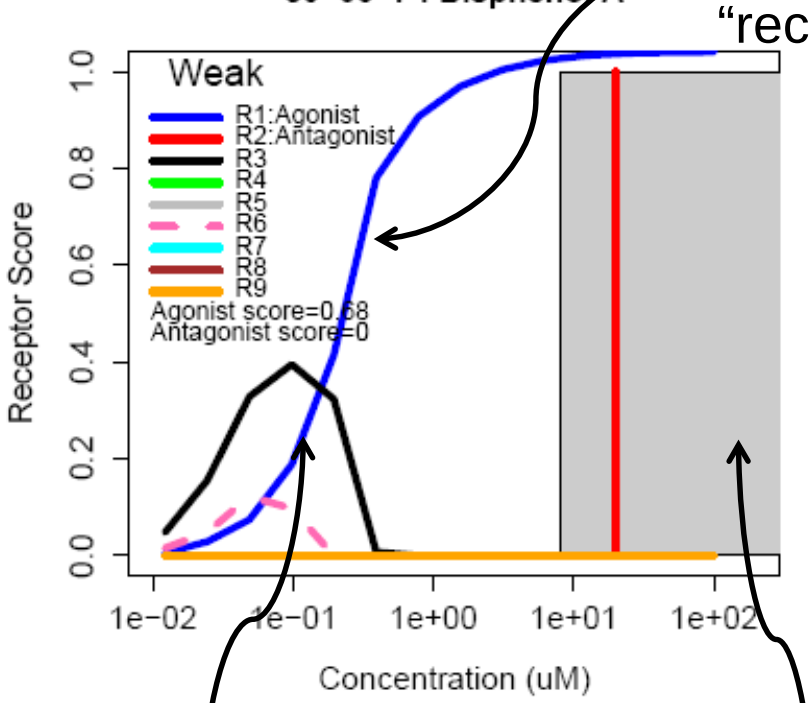
Example 1 – BPA – true agonist (AUC=0.66)

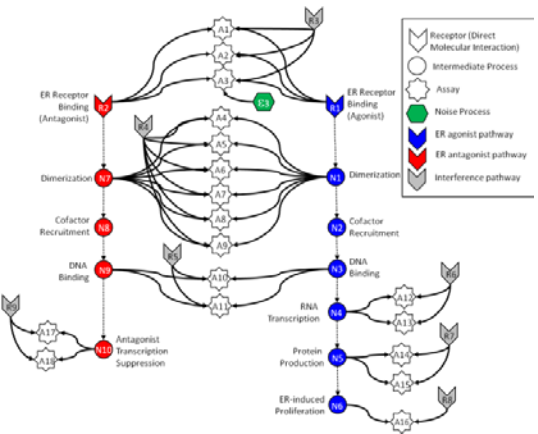
Assays



"Receptors"

Blue: agonist "receptor"





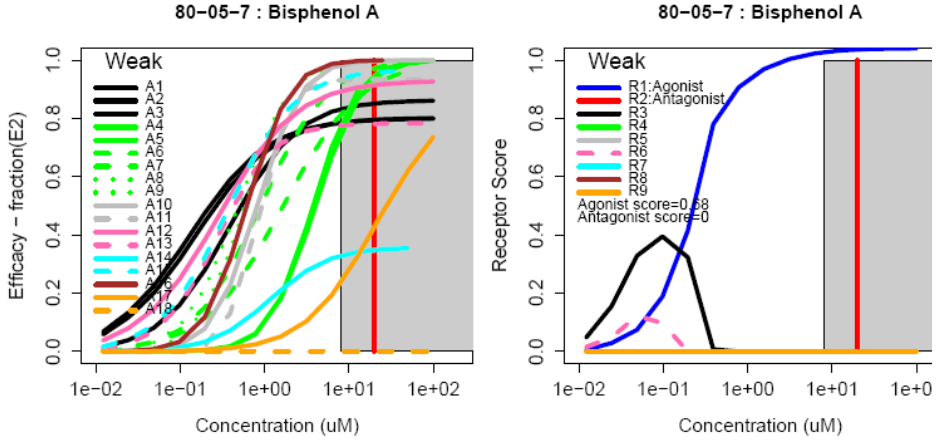
Binding assays active at lowest concentration

AUC "sign" feature will discount this

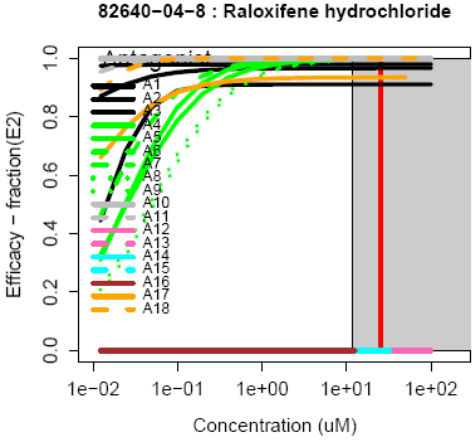
Cytotoxicity Region: red line is median cytotox AC50

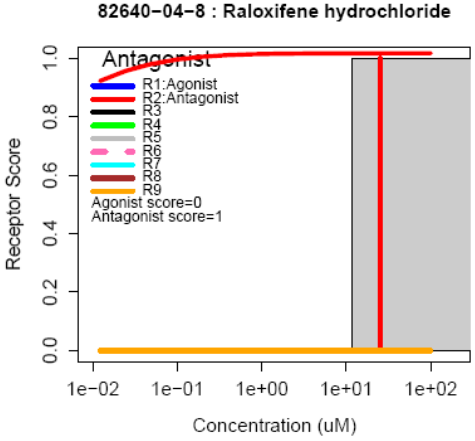
Example curves

True Agonist

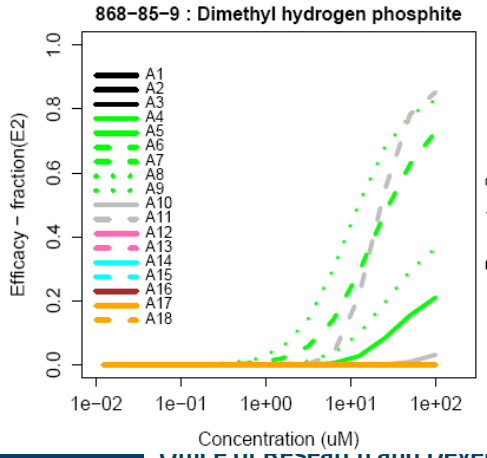


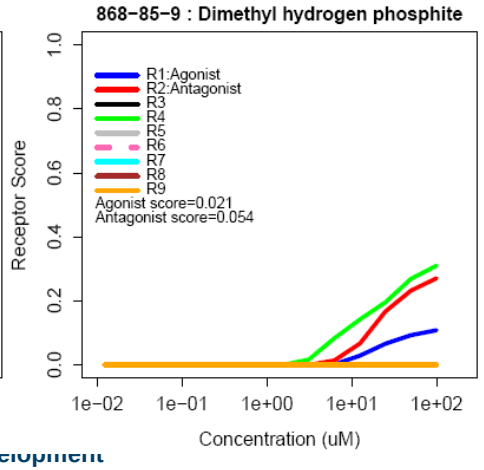
True Antagonist



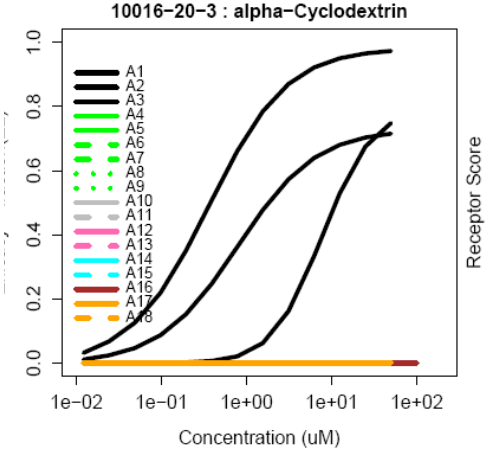


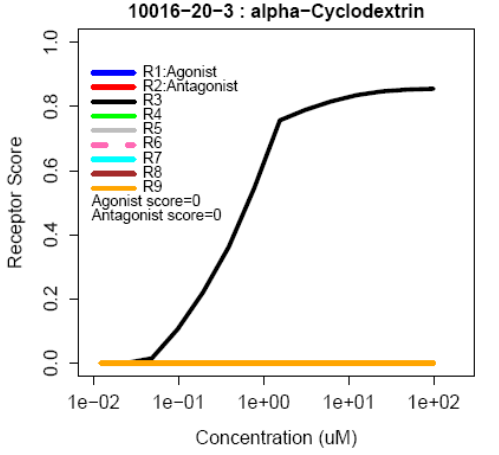
Negative-BAI



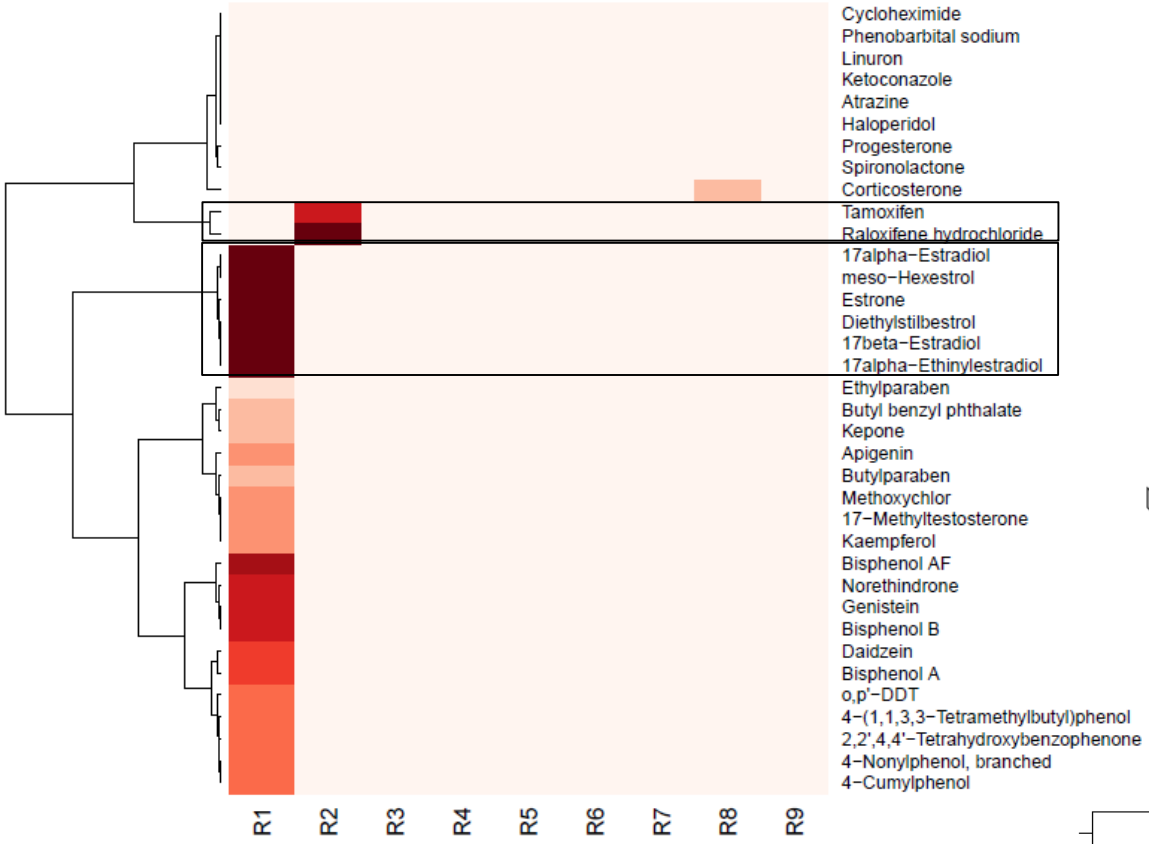


Negative-NAI

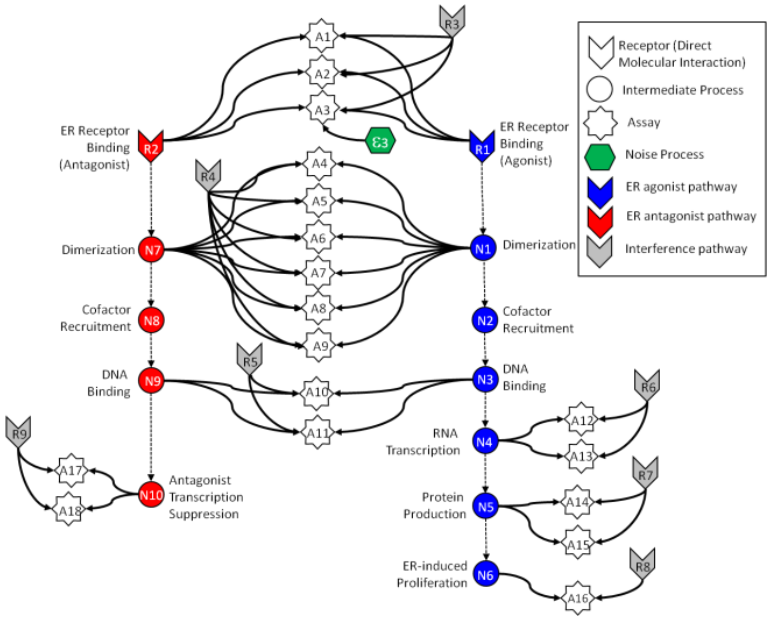


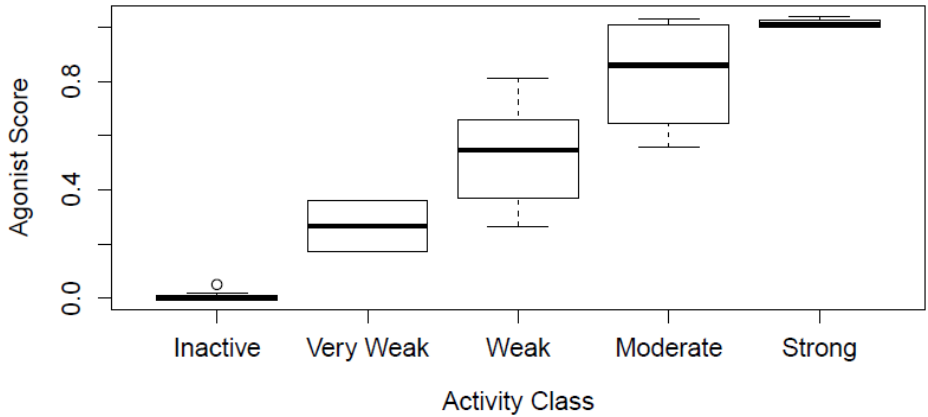


Reference Chemical Classification



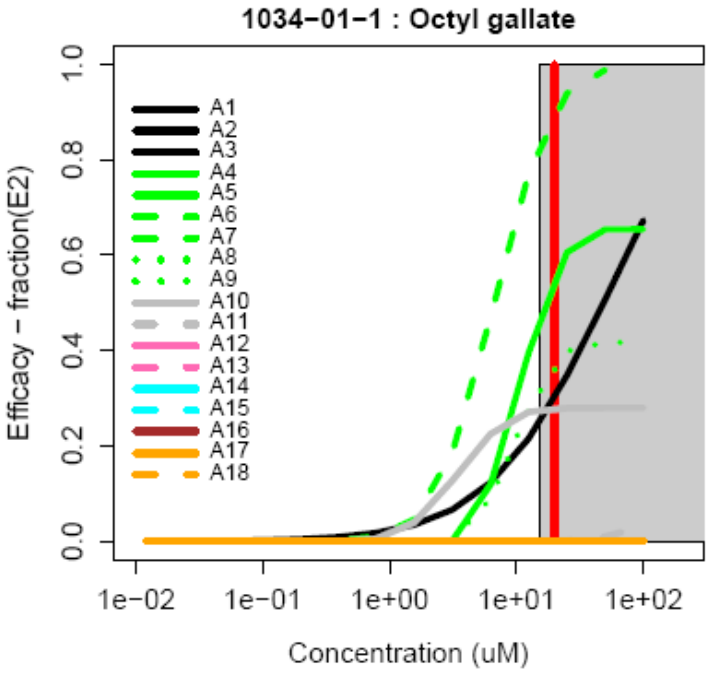
AUC heat map for Reference chemicals





Back to ToxCast: Example illustrating assay data

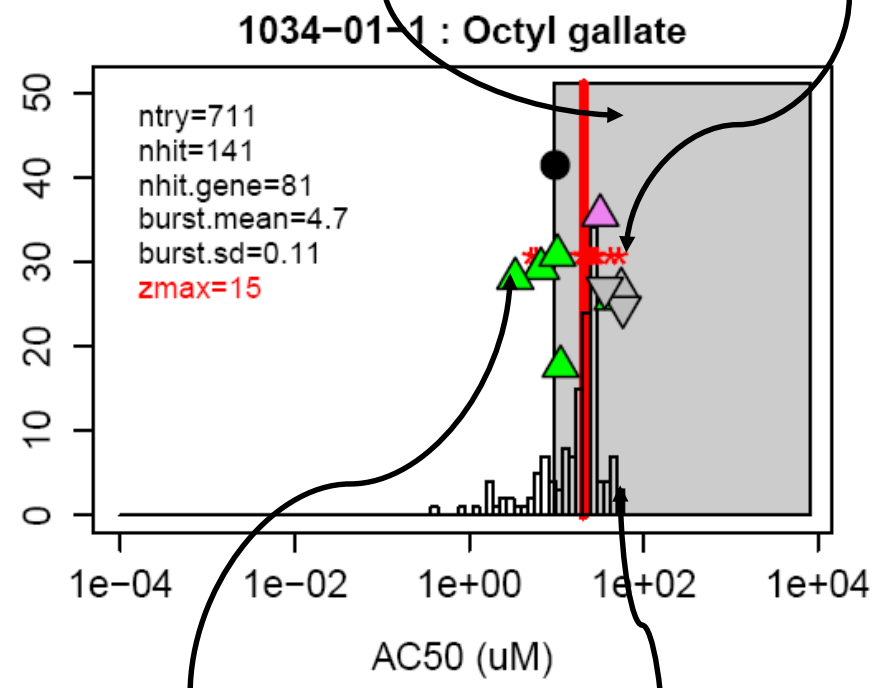
Concentration-response data for single gene (ESR1 / ER)



Histogram of AC50 Values

±3 SD for burst

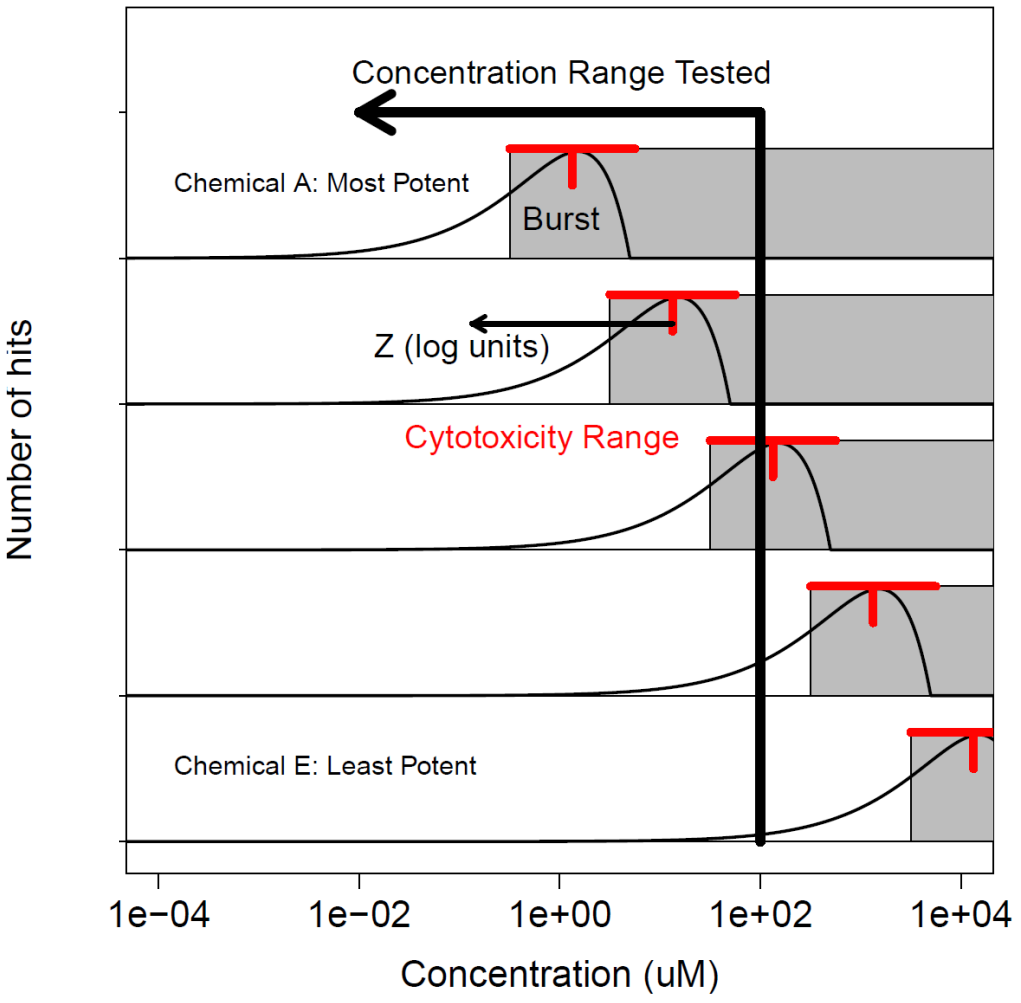
Cytotox assays



AC50s for ER assays

Histogram counting hits

Most chemicals display a “burst” of activity at same concentration as cytotoxicity

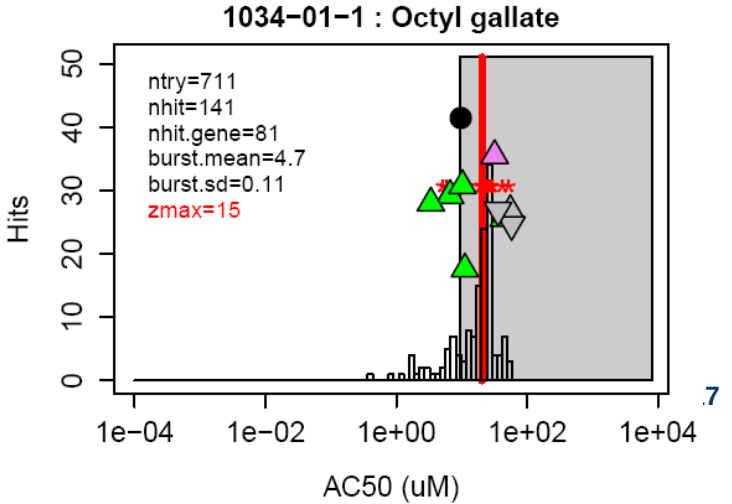


Most chemicals cause activity in many assays near the cytotoxicity threshold

Cell-stress related assay interference

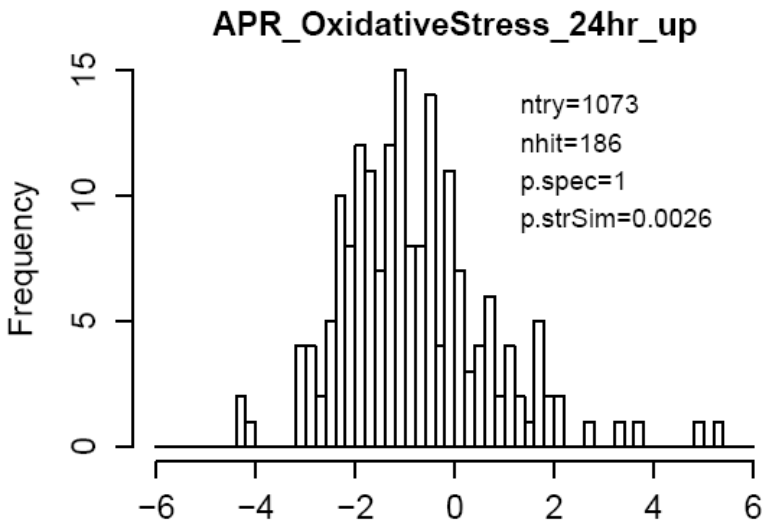
“Hit” (AC50) in burst region is less likely to result from specific activity (e.g. binding to receptor or enzyme)

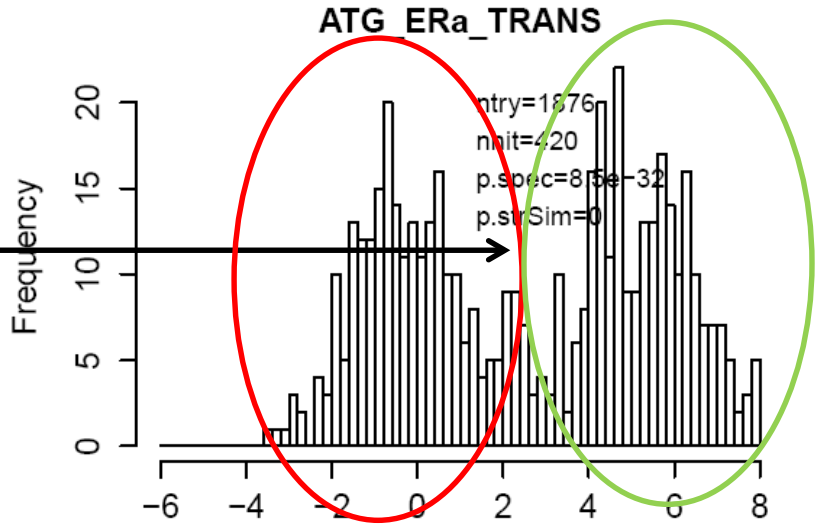
Z-score: # of SD from burst center
-High Z: more likely to be specific
-Low Z: less likely to be specific



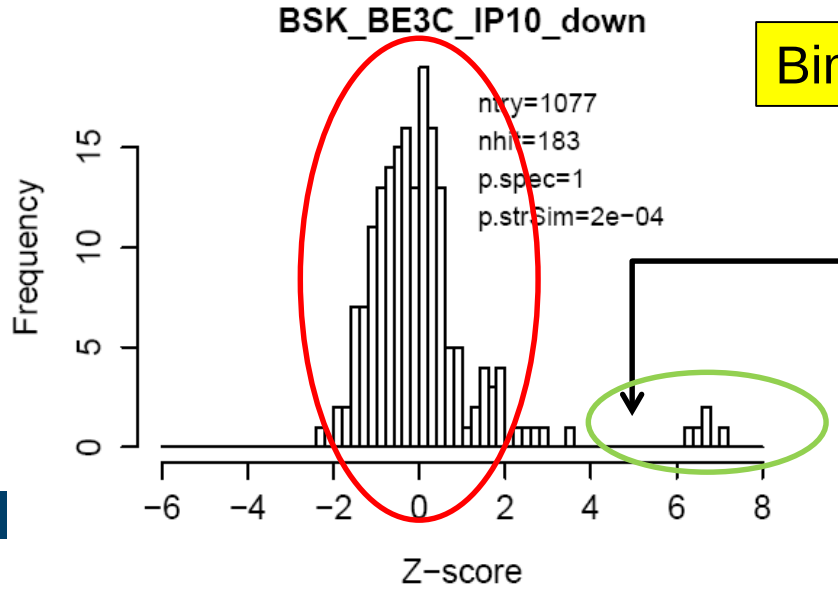
Examine Z-scores by assay

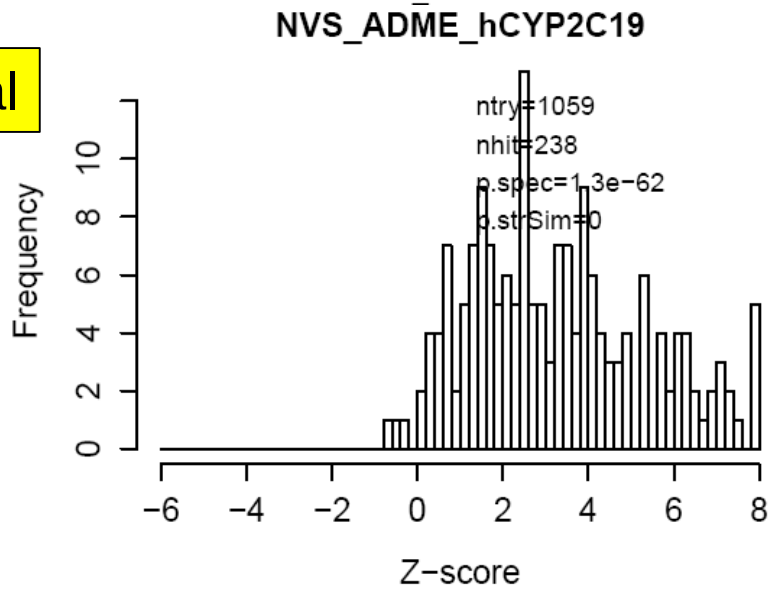
Cytotox / Cell Stress
"True" activity



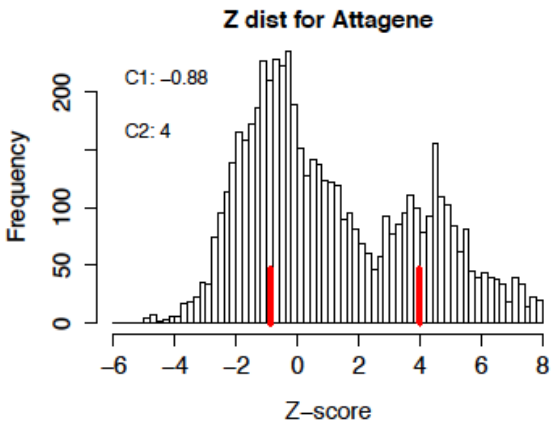


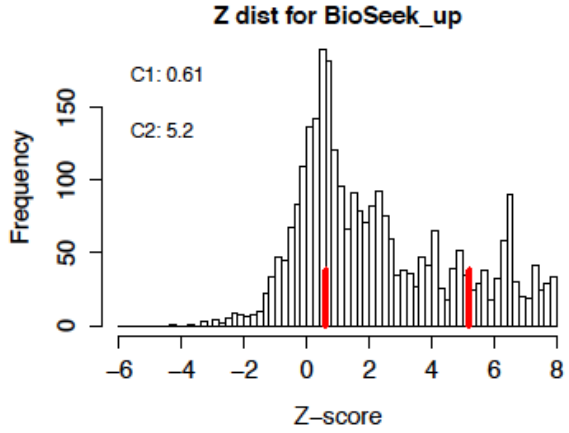
Bimodal

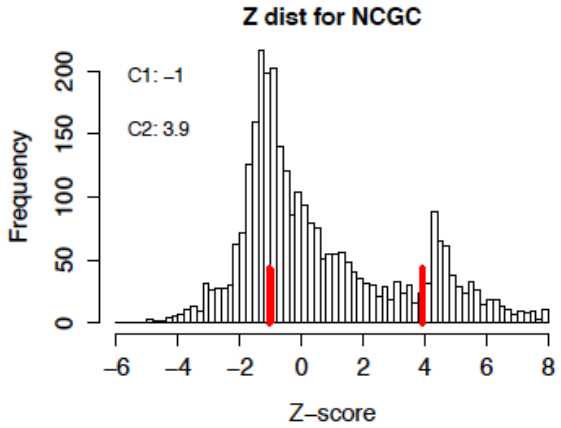


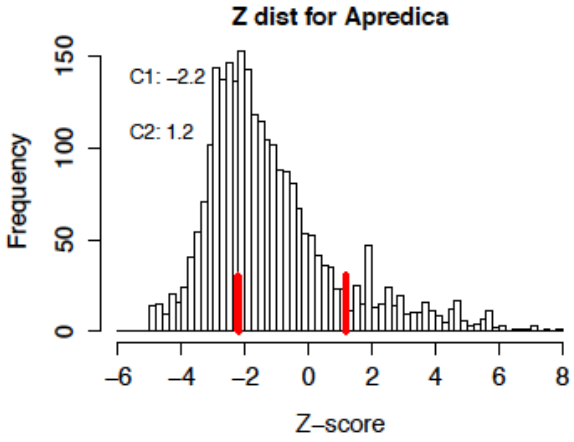


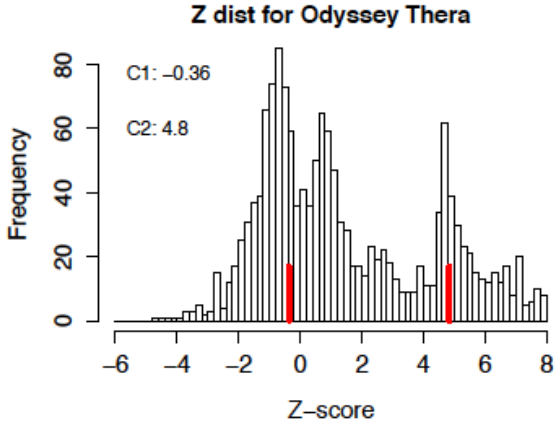
Bimodal is seen in all technologies

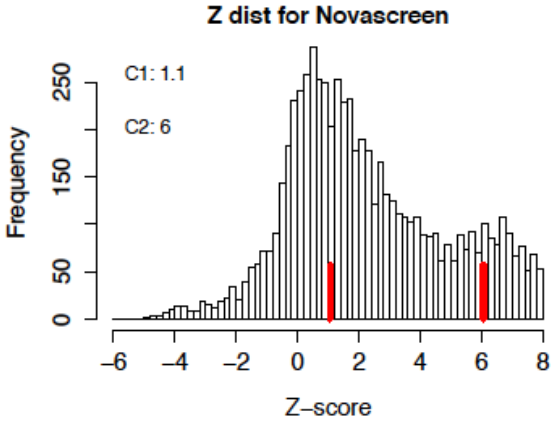




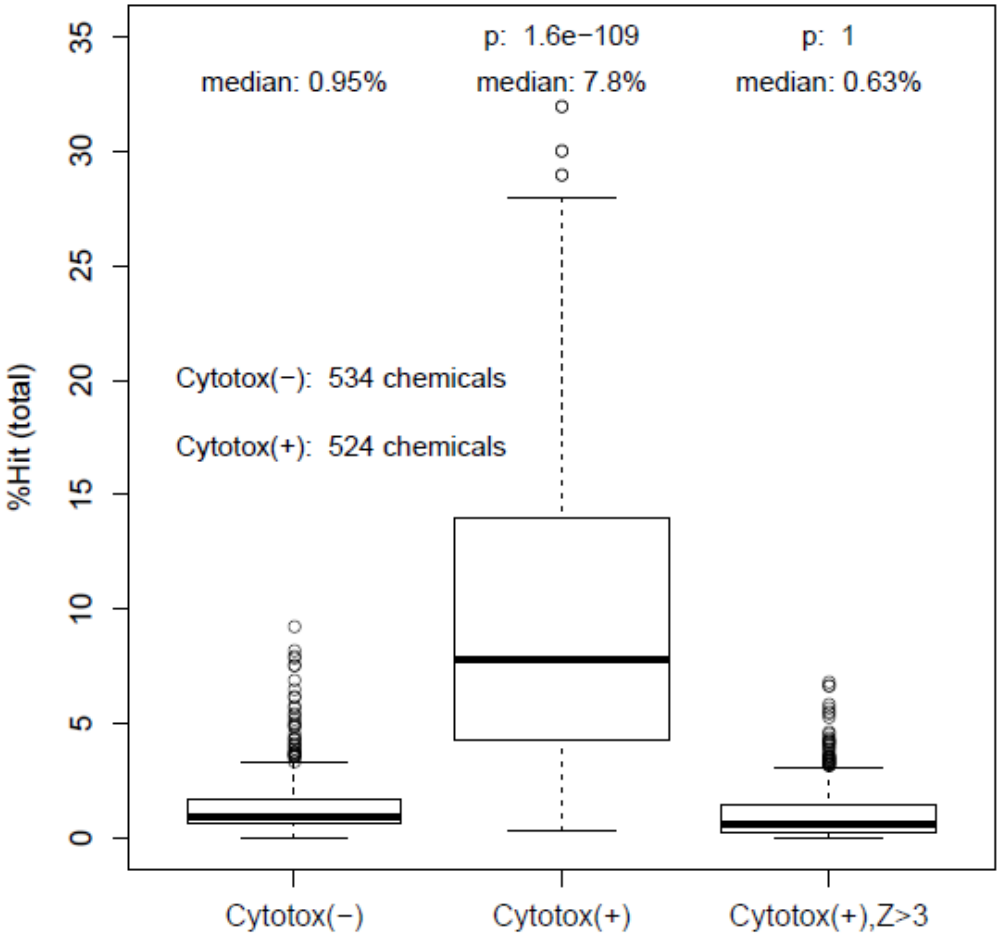








Non-specificity with cytotox is general



Having cytotoxicity <100 uM greatly increases number of hits

Gene Score
Combine potency and specificity

- How to summarize 1000s of chemicals x 100s of assays?

- Potency: -log(AC50)
- Gene score = mean potency across all assays for a gene
- Can be used to get quick ranking of chemicals

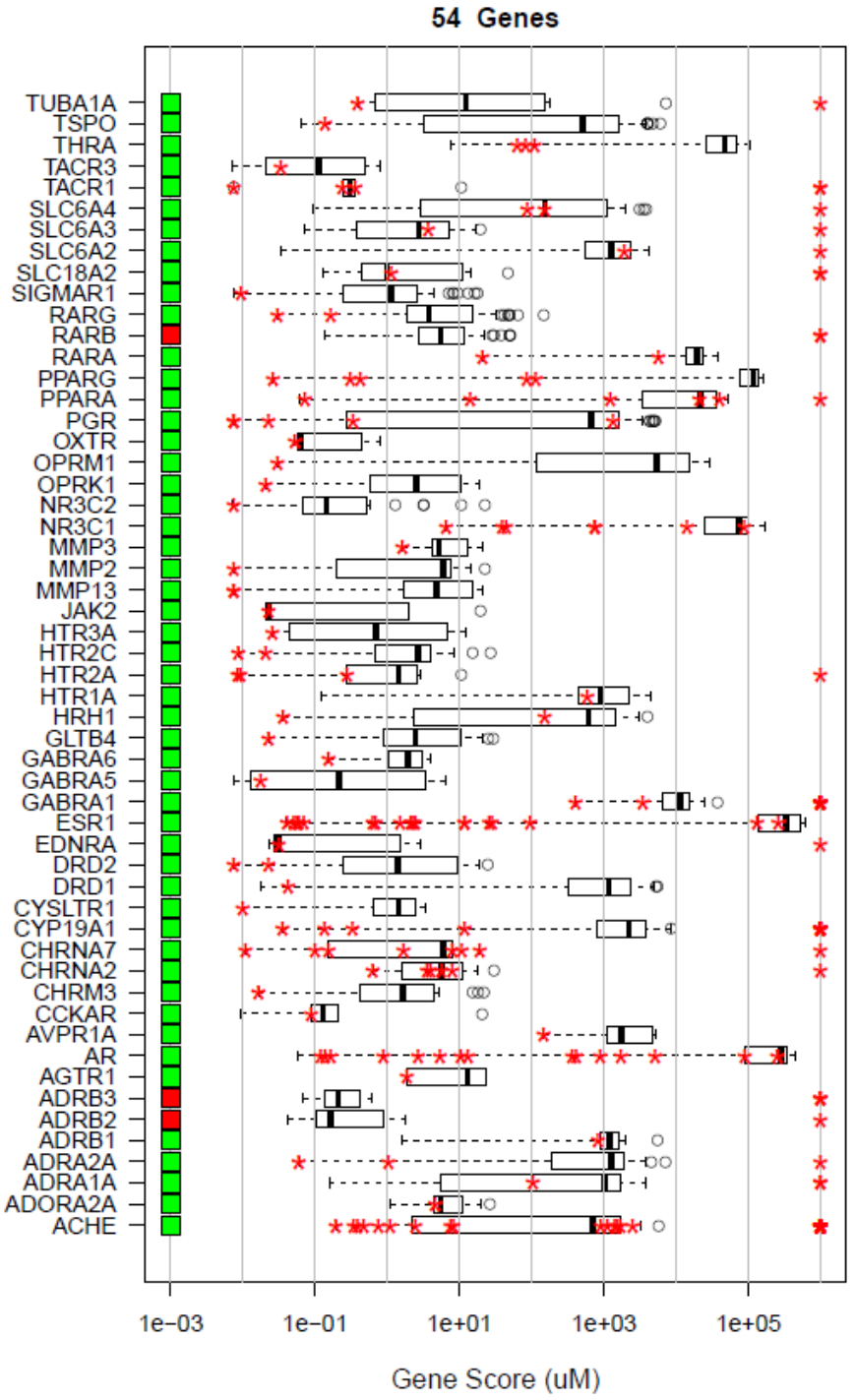
Do Assays Detect Potent Reference Chemicals?

* =Reference chemicals

- These chemicals should be near the left of the gene score distribution

- Most assays show reference chemicals to be potent

- Gives confidence that novel chemicals active in the assay are perturbing that pathway



ShortName	UseCategory	IntendedTarget	Gene	GeneScore
PharmaGSID_48521	Pharmaceutical		SAA1	12.9
Tributyltin methacrylate	Microbicide		H2AFX	12.8
Cladribine	Pharmaceutical	DNA	H2AFX	12.8
Cytarabine hydrochloride	Pharmaceutical	DNA	H2AFX	12.8
Tributyltin chloride	Microbicide		H2AFX	12.6
Mancozeb	Microbicide		SRC	12.6
Ziram	Fungicide		H2AFX	12.3
Sodium dimethyldithiocarbamate	Chelator		H2AFX	12.1
Simvastatin	Pharmaceutical	HMGCR	THBD	12
FR150011	Pharmaceutical	CYSLTR1	COL3A1	11.9
Pentachlorophenol	Wood preservative (previously)		COL3A1	11.7
tert-Butylhydroquinone	Antioxidant		SAA1	11.6
3,3,5,5-Tetraiodothyroacetic acid	Pharmaceutical	THRA	COL3A1	11.5
4-Chloro-1,2-diaminobenzene	Chemical intermediate/dye additive		H2AFX	11.5
2-Aminoanthraquinone	chemical intermediate (dyes and pharmaceuticals)		SAA1	11.5
Dichlorvos	Insecticide	ACHE	SAA1	11.5
Corticosterone	Pharmaceutical	NR3C1	COL3A1	11.4
Tebufenozide	Insecticide	Ecdysone receptor	COL3A1	11.2
Clotrimazole	Fungicide	Yeast 14 demethylase	PTGER2	11.2
Cariporide mesylate	Pharmaceutical	Ion channel Na	COL3A1	11
Triglycidyl isocyanurate	Epoxy hardener		H2AFX	11
Diethyl phthalate	Plastics		COL3A1	10.7
YM218	Pharmaceutical	AVPR1A	SAA1	10.7
Tebuthiuron	Herbicide		COL3A1	10.6
Octhilinone	Fungicide		H2AFX	10.6
PharmaGSID_47261	Pharmaceutical	HIV nucleocapsid protein	PTGER2	10.5
Imazethapyr	Herbicide	ALS	SAA1	10.4
4-Cyclohexylcyclohexanol	Chemical reactant		SAA1	10.3
Cycloate	Herbicide	cyp19a1 (?)	COL3A1	10.2
UK-373911	Pharmaceutical			10.2
3,5,3'-Triiodothyronine	Pharmaceutical			10.2
Cloprop	Herbicide			10.1
FR900409	Pharmaceutical			10.1
Norflurazon	Herbicide			10.1

Highest gene scores are mostly pharmaceuticals and pesticidal active ingredients

Promiscuity measures

Calculate the number of genes hit with Gene Score>7

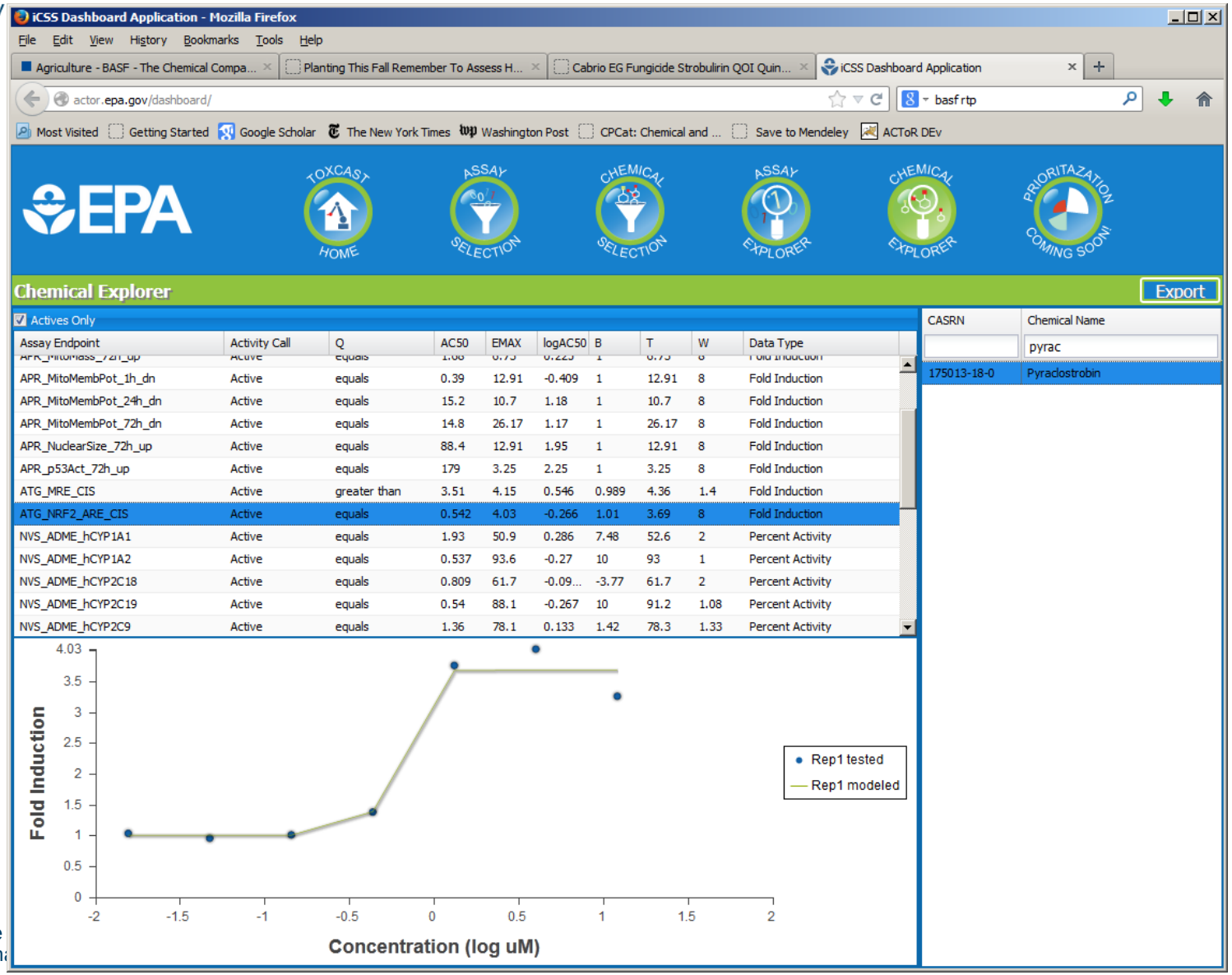
"Hottest" – Most Specific Hits

Category	Nchem	mean_HitRatio	SD_HitRatio	p-hot
conazole (triazoles)	13	0.0343	0.0213	3.50E-06
Pharma Class 4.86	10	0.031	0.015	1.14E-05
Pharma Class 4.58	11	0.0285	0.0169	4.11E-05
conazole (imidazoles)	6	0.0313	0.03	0.003
Pharma Class 3.292	5	0.0385	0.0333	0.00493
steroid P	5	0.0224	0.0162	0.00519
Pharma Class 4.43	7	0.0197	0.0118	0.00673

"Coldest" – Fewest Specific Hits

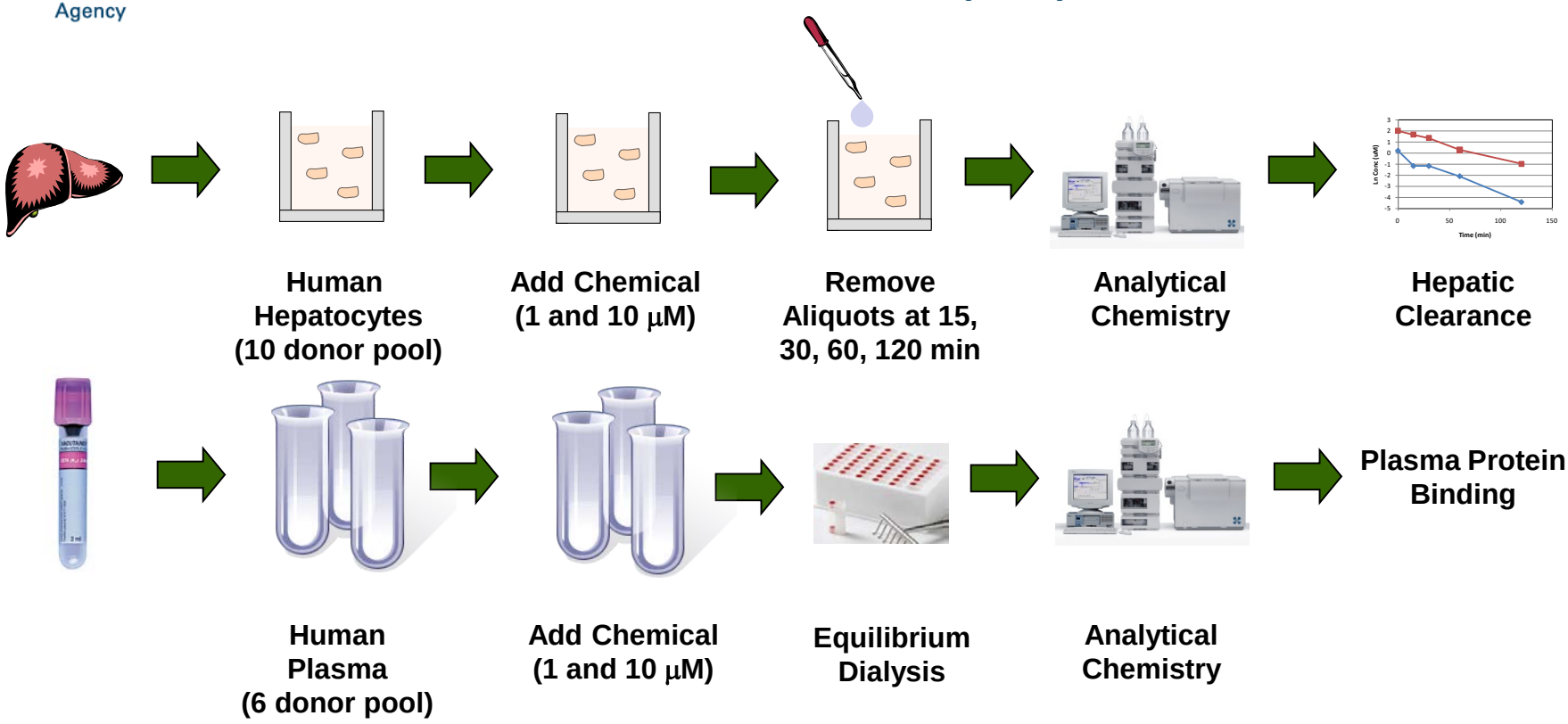
Category	Nchem	mean_HitRatio	SD_HitRatio	p-cold
alcohol pri	10	0.00105	0.00236	0.000214
phthalate	17	0.00324	0.00517	0.000838
carboxylate di	15	0.00281	0.00329	0.00286
carboxylate	7	0.0015	0.00187	0.00422

ToxCast Dashboard: making data public



http://actor.epa.gov/dashboard

Adding Pharmacokinetics
Reverse ToxicoKinetics (rTK)



Combine experimental data w/ PK Model to estimate dose / concentration scaling

RatCast: Same experiment, but with rat hepatocytes and plasma

Collaboration with Thomas et al., Hamner Institutes
Publications: Rotroff et al, ToxSci 2010, Wetmore et al, ToxSci 2012

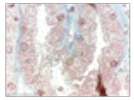
HTRA – High-throughput Risk Assessment

United States Environmental Protection Agency

Pharmacodynamics

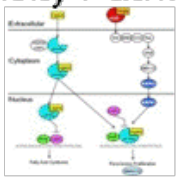
AOP

Adverse Effect

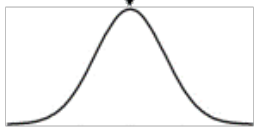


Key Events

Toxicity Pathway

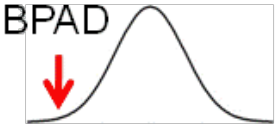


HTS Assays



Biological Pathway Activating Concentration (BPAC) Probability Distribution

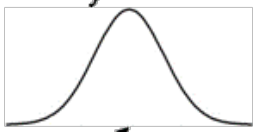
BPAD



Probability Distribution for Dose that Activates Biological Pathway

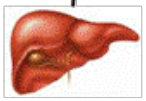
Pharmacokinetics

Dose-to-Concentration Scaling Function (C_{ss}) Probability Distribution



PK Model

Populations

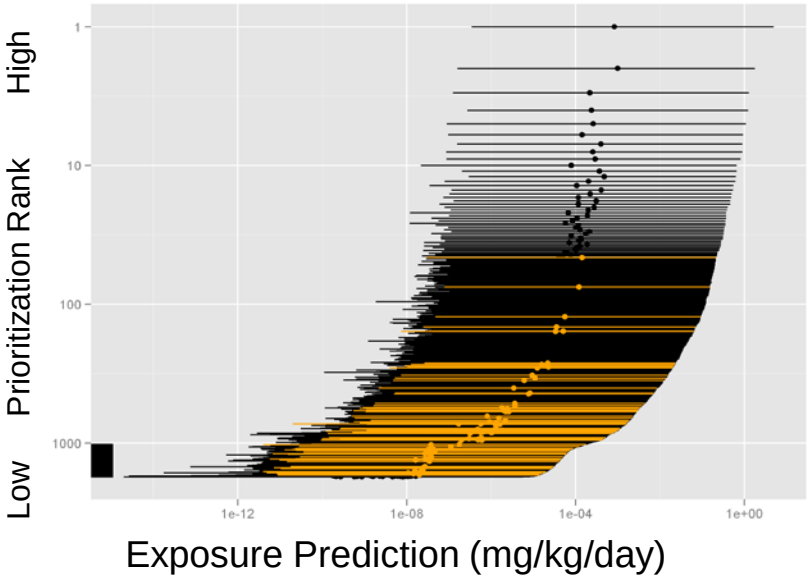


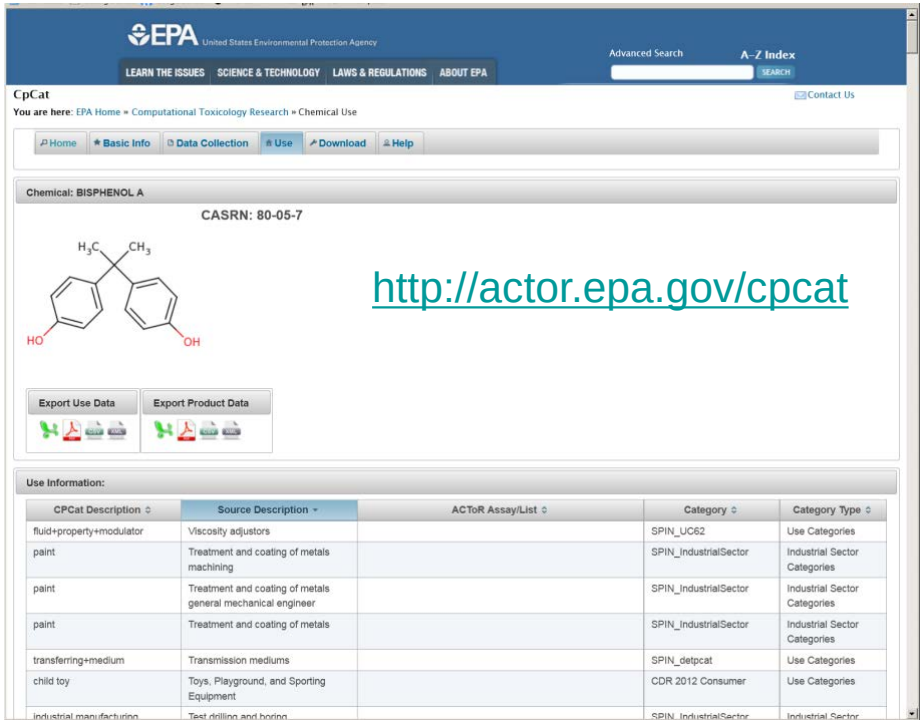
Intrinsic Clearance

Plasma Protein Binding

ExpoCast: High-Throughput Exposure Models

- High-Throughput Exposure Predictions for Risk-Based Prioritization
  - Quantitatively predict exposure for 1000s of chemicals
  - Combine with information on chemical uses and products (CPCat Database)
- Output
  - Quantitative estimates of exposure levels
  - Estimates of variance for the predictions
  - Ranking of chemicals by exposure potential





Wambaugh et al. Environ. Sci. Technol., 2013, 47, pp 8479

ToxCast is a work in progress

- ToxCast is controversial with some audiences
  - Recent presentations / publications have claimed “ToxCast has failed”

- What would success look like?
  - Animal to human extrapolation or even rat to mouse is far from perfect

- Is the data “wrong”?
- Is the data incomplete?
- Are our models (i.e. understanding of biology) incomplete?

- All of these are true to some extent

Understanding Success and Failure

- Why *In vitro* to *in vivo* can work:
  - Chemicals cause effects through direct molecular interactions that we can measure with *in vitro* assays

- Why *in vitro* to *in vivo* does not always work:
  - Pharmacokinetics issues: biotransformation, clearance (**FP**, **FN**)
  - Assay coverage: don't have all the right assays (**FN**)
  - Tissue issues: may need multi-cellular networks and physiological signaling (**FN**)
  - Statistical power issues: need enough chemicals acting through a given MOA to be able to build and test model (**FN**)
  - Homeostasis: A multi-cellular system may adapt to initial insult (**FP**)
  - *In vitro* assays are imperfect (**FP**, **FN**)
  - *In vivo* rodent data is imperfect (**FP**, **FN**)

EPA National Center for Computational Toxicology Staff
February 5, 2013