HTS Data and *In Silico* Models for High-Throughout Risk Assessment

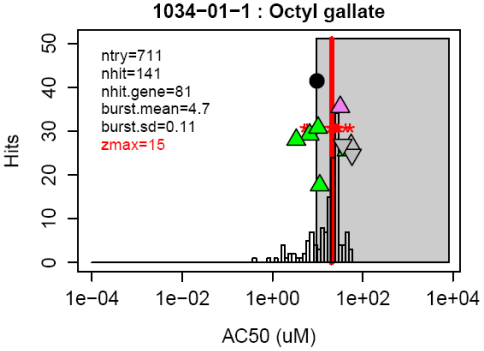
Richard Judson
U.S. EPA, National Center for Computational Toxicology
Office of Research and Development

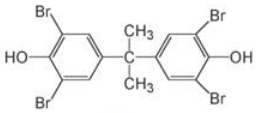
FutureTox, RTP NC January 2014

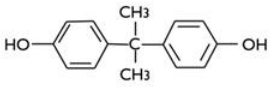
Outline

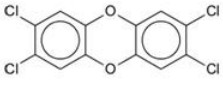
- Types of target activity – “specific” vs. “non-specific”
- Focus on specific
  - Gene-centric: the “Gene Score”
  - Pathway-centric: Estrogen Receptor Pathway
- Focus on non-specific
  - Using in vitro assay and PK data to predict MTD

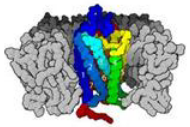
Significance of In Vitro Effects

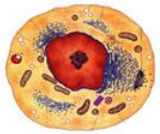












Assay Target Class

Assessment

Molecular Target

EDC
Acetylcholinesterase Inhibition
Ion channel blocker
Genotoxicity

→ AOP Assessment
→ Targeted testing

Cell Stress Mediated

Oxidative stress
Membrane disruption
Nucleophiles
Electrophiles
Energy depletion

→ Estimate MTD
→ Estimate NOEL

No Effect

Non-reactive chemical
Not bioactive
Effects would require high doses

→ Estimate NOEL

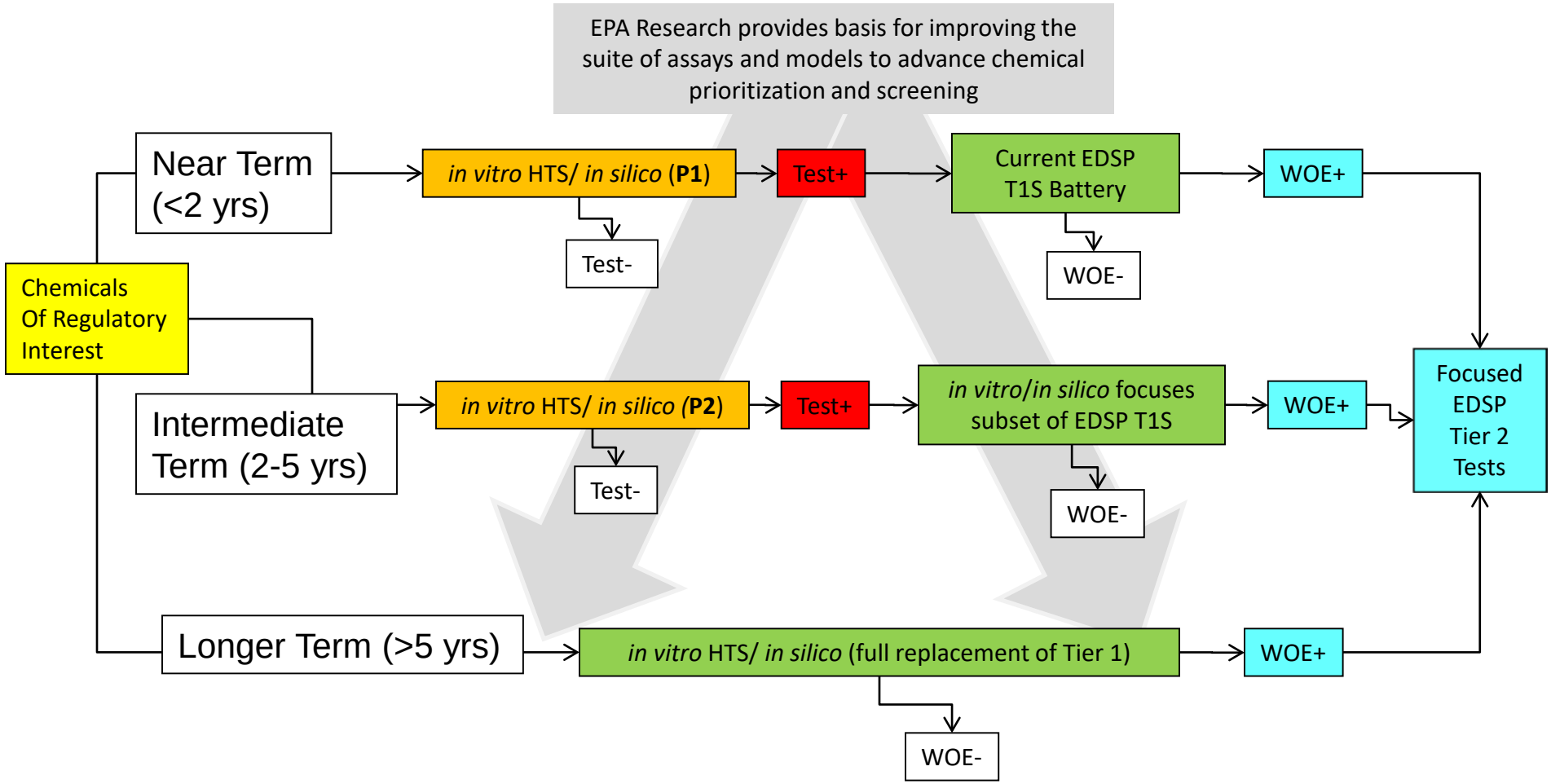
EDSP: A First, Real-World Application of Tox21 HTS

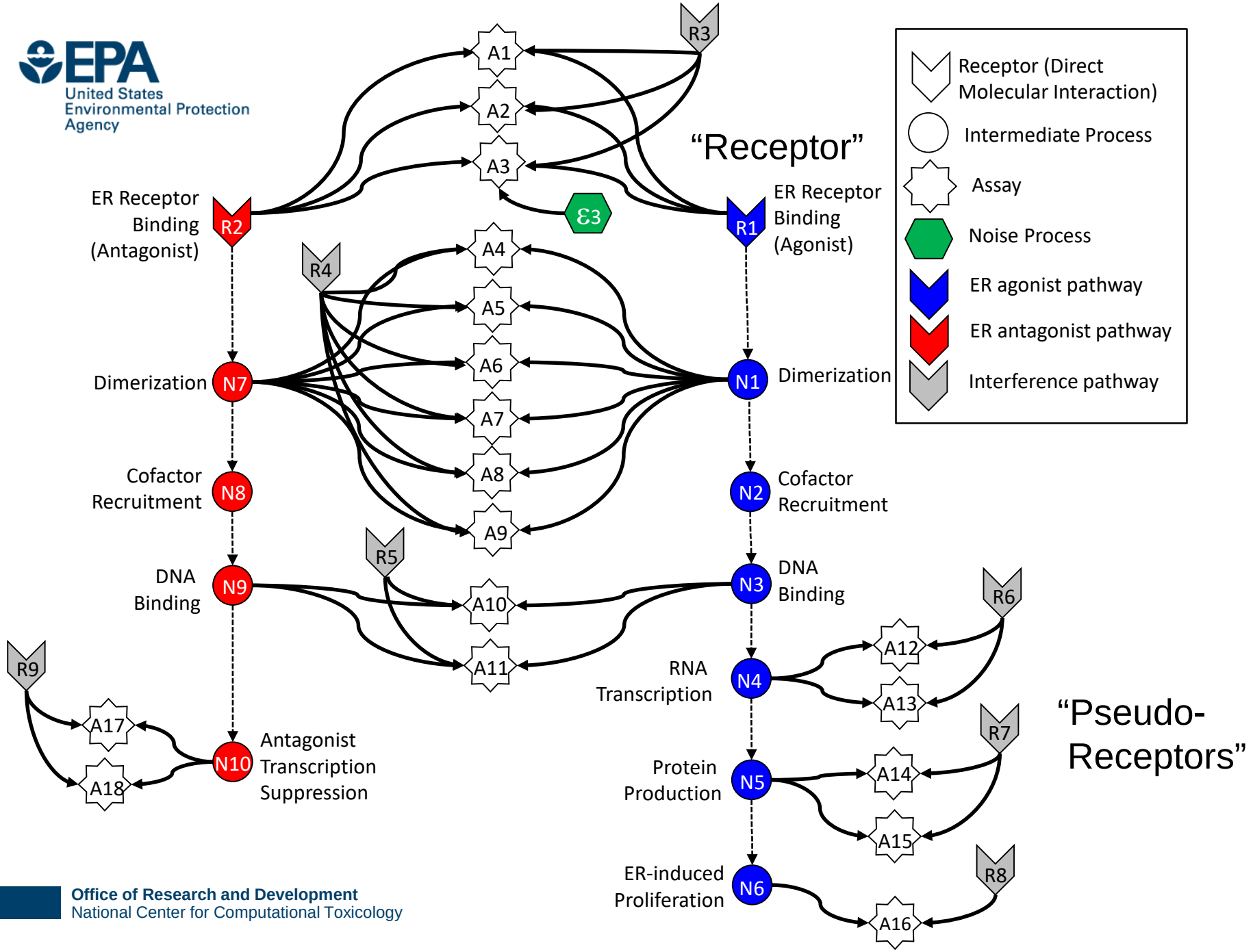
Prioritization for Endocrine Disruptor Screening Program

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

Throughput: ~100 Chemicals per year

Cost: ~$1M per chemical



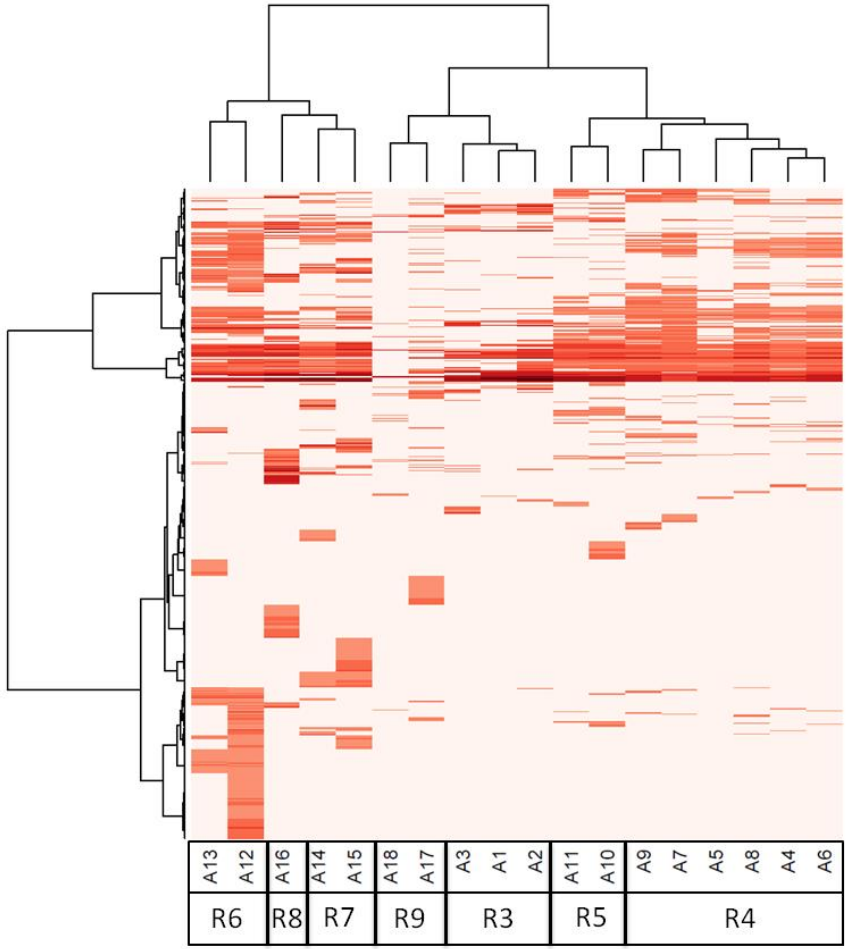

EPA
United States
Environmental Protection
Agency
"Receptor"
"Pseudo-
Receptors"
Receptor (Direct
Molecular Interaction)
Intermediate Process
Assay
Noise Process
ER agonist pathway
ER antagonist pathway
Interference pathway
ER Receptor
Binding
(Antagonist)
ER Receptor
Binding
(Agonist)
Dimerization
Dimerization
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

Major theme – all assays have false positives and negative

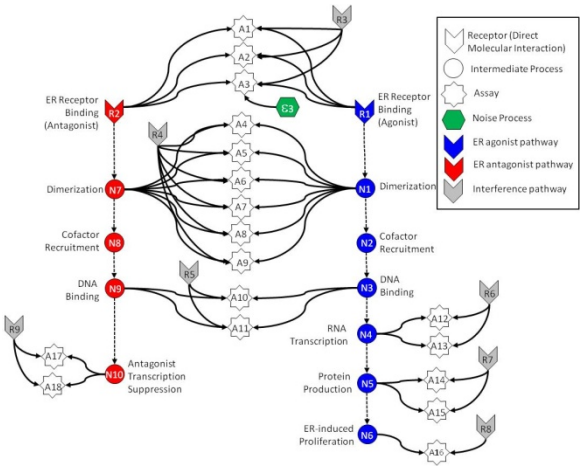
Assays cluster by technology, suggesting technology-specific non-ER activity



Much of this “noise” is reproducible, i.e. it is “assay interference”

Result of interaction of chemical with complex biology in the assay

Our chemical library is only partially “drug-like”
-Solvents
-Surfactants
-Intentionally cytotoxic compounds
-Metals
-Inorganics



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

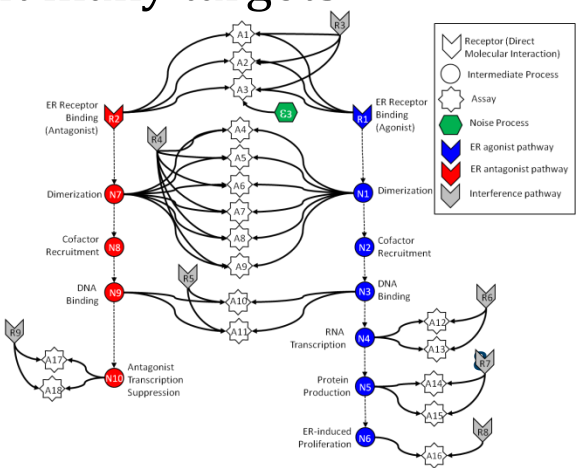
$$A_i^{pred} \in [1,0]$$

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

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

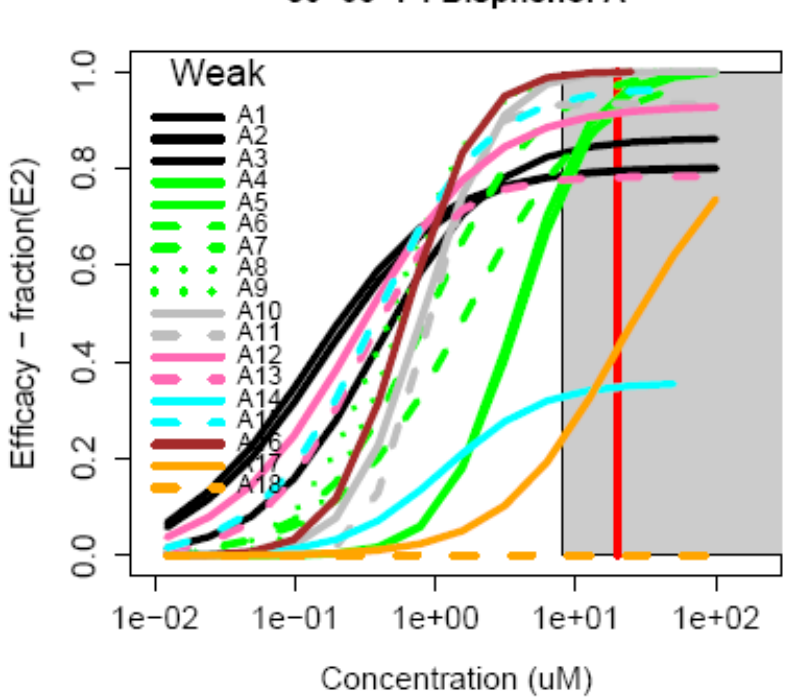
$$AUC_j = \frac{1}{N_{conc}} \sum_{i=1}^{N_{conc}} sign(slope) \times R_j(conc_i)$$

AUC Summarizes results



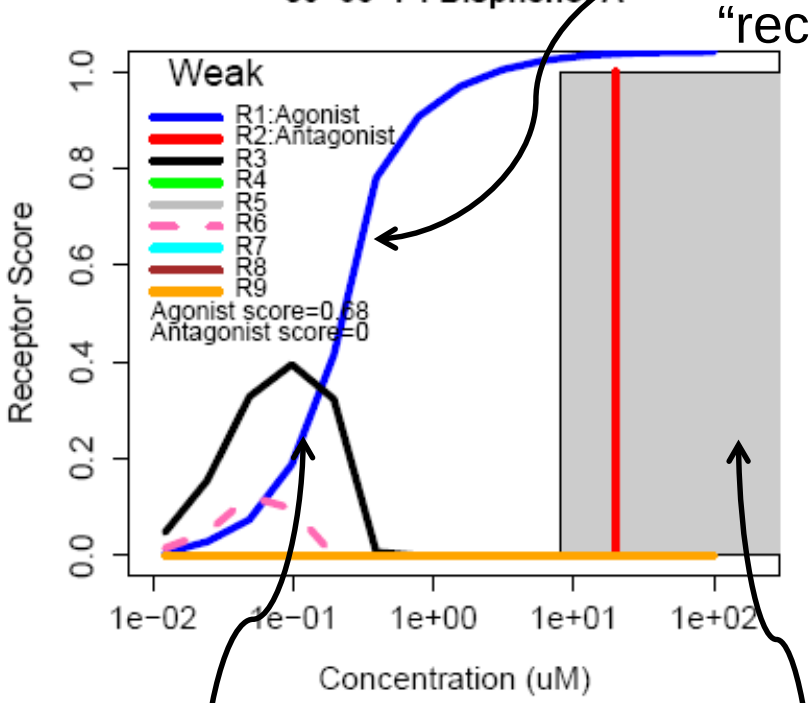
Example 1 – BPA – true agonist (AUC=0.66)

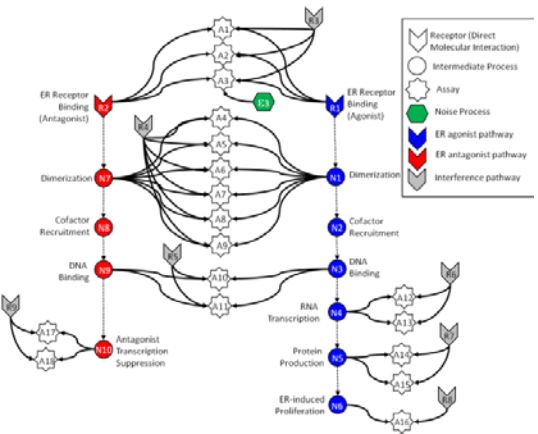
Assays



"Receptors"

Blue: agonist "receptor"





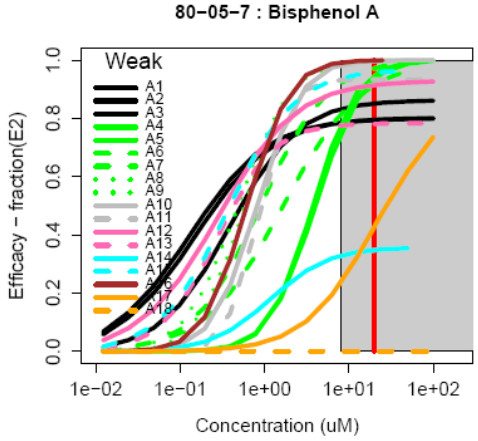
Binding assays active at lowest concentration

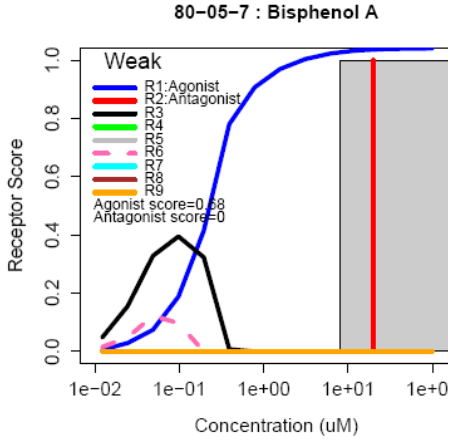
AUC "sign" feature will discount this

Cytotoxicity Region: red line is median cytotox AC50

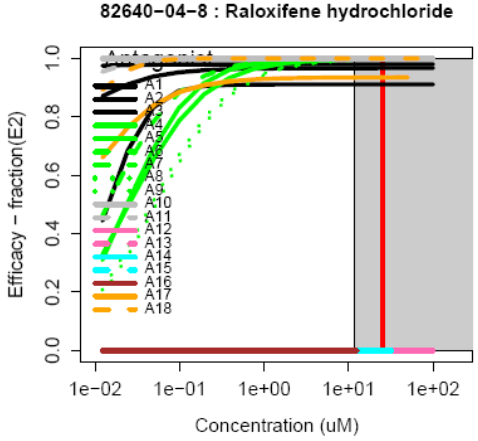
Example curves

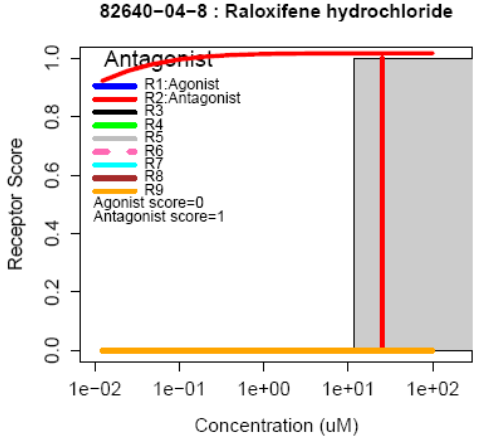
True Agonist



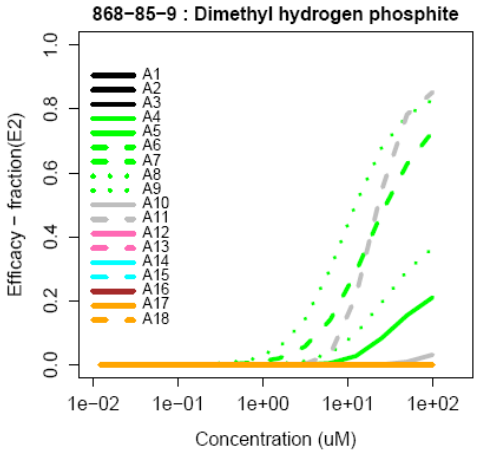


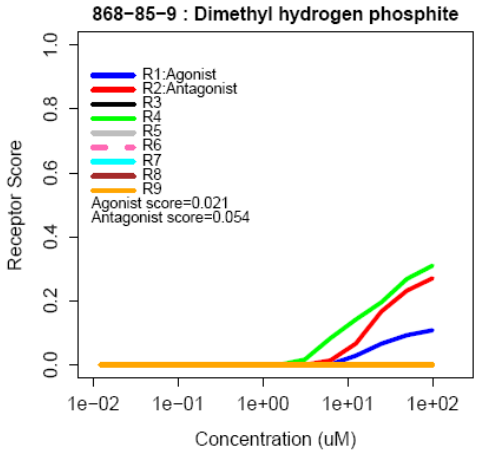
True Antagonist



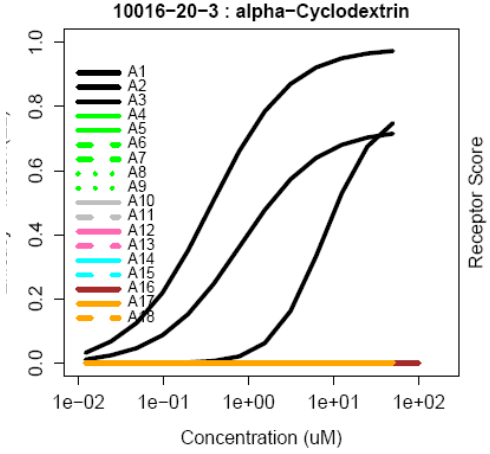


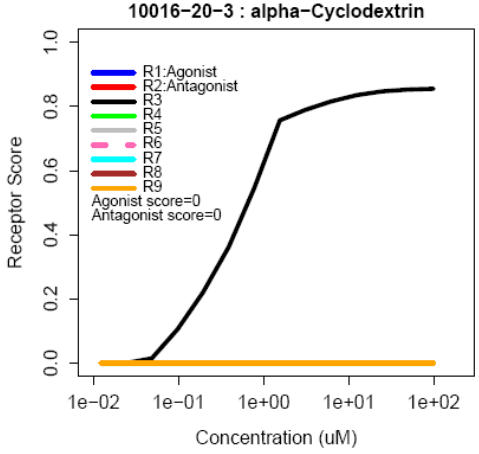
Negative-Broad Assay Interference



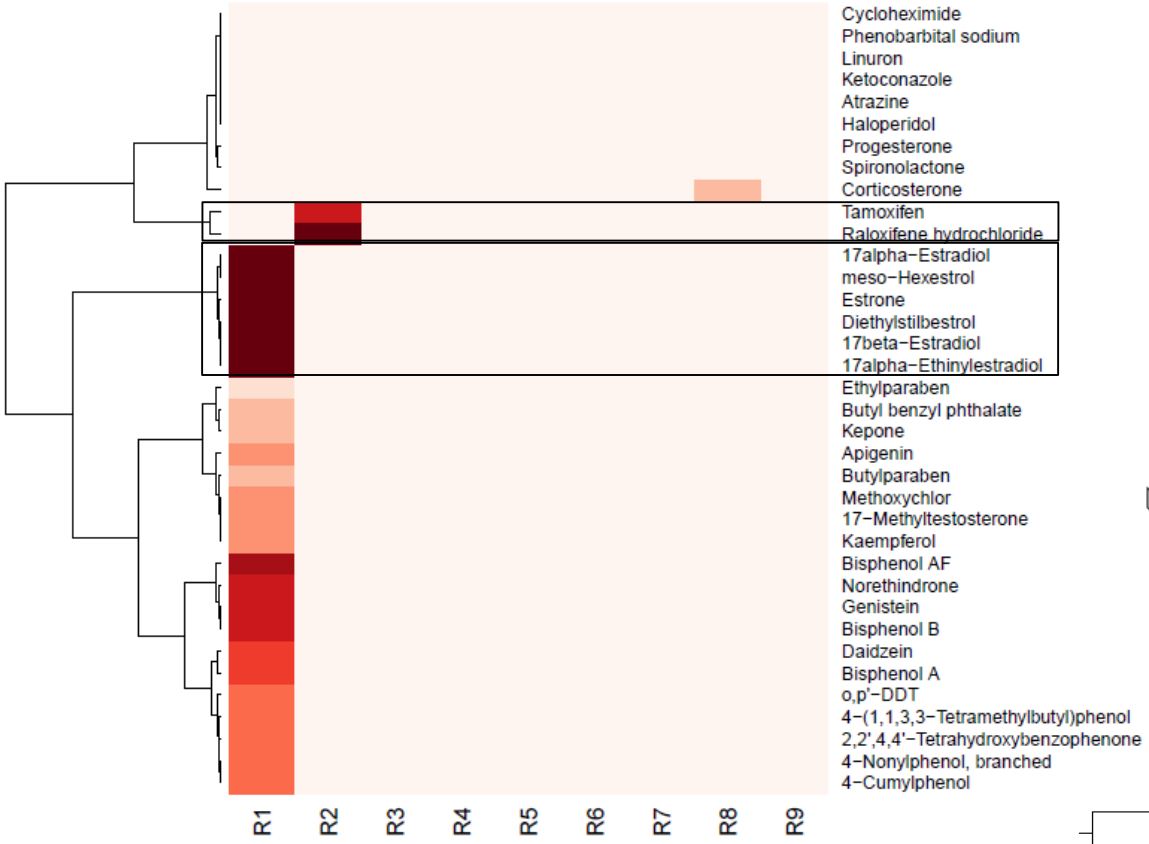


Negative-Narrow Assay Interference

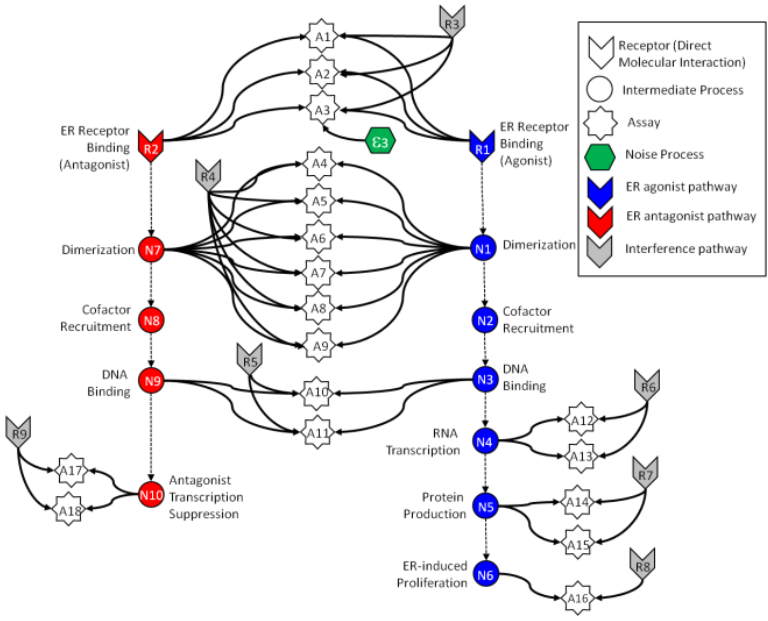


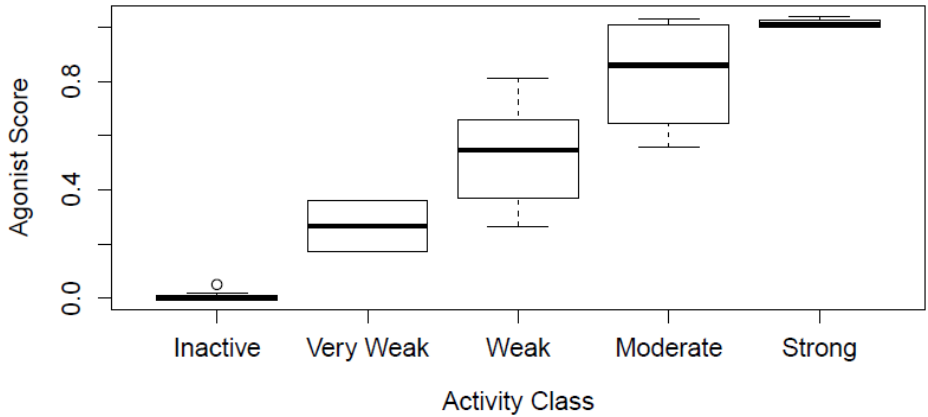


Reference Chemical Classification



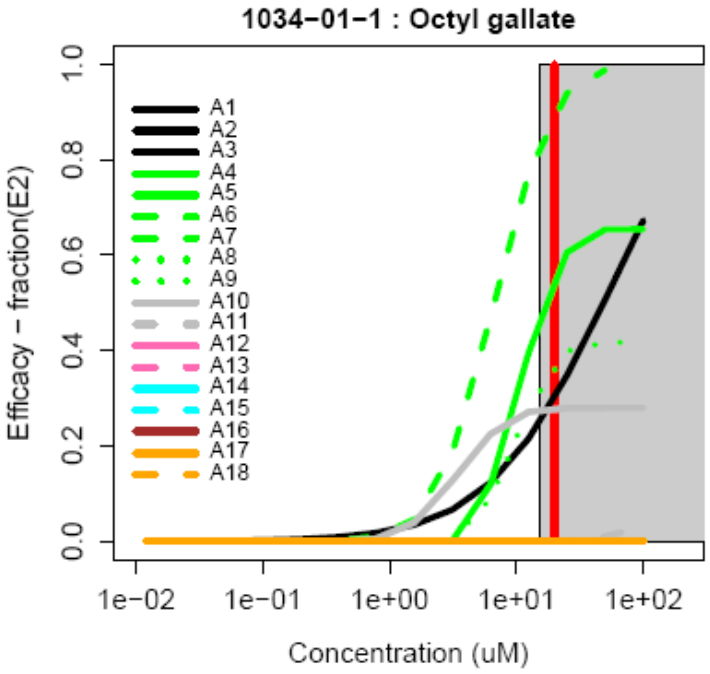
AUC heat map for Reference chemicals



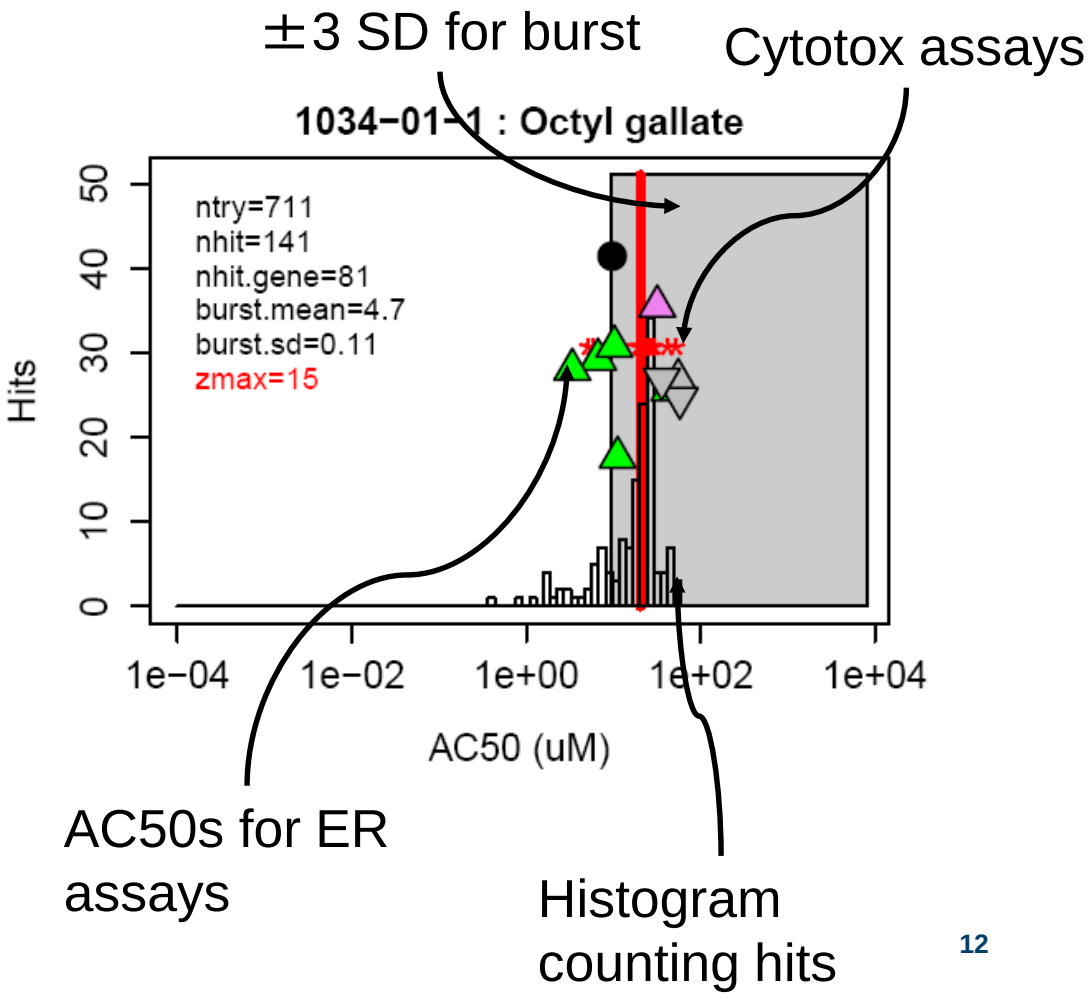


Example illustrating assay data

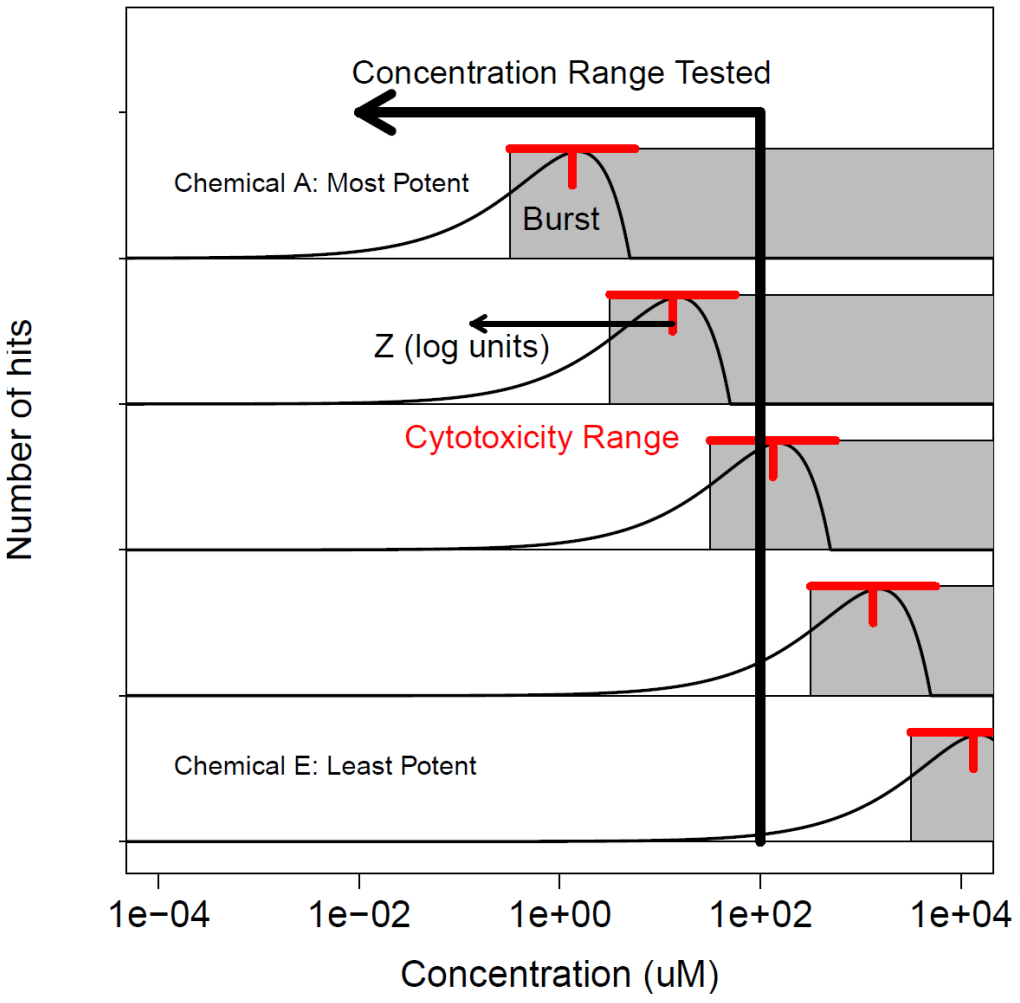
Concentration-response data for single gene (ESR1 / ER)



Histogram of AC50 Values



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

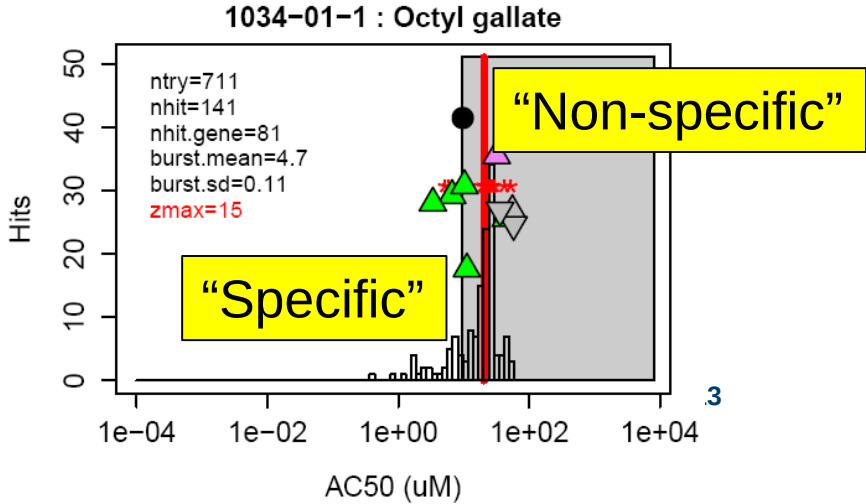


Most chemicals cause activity in many assays near the cytotoxicity threshold

Cell-stress related assay interference

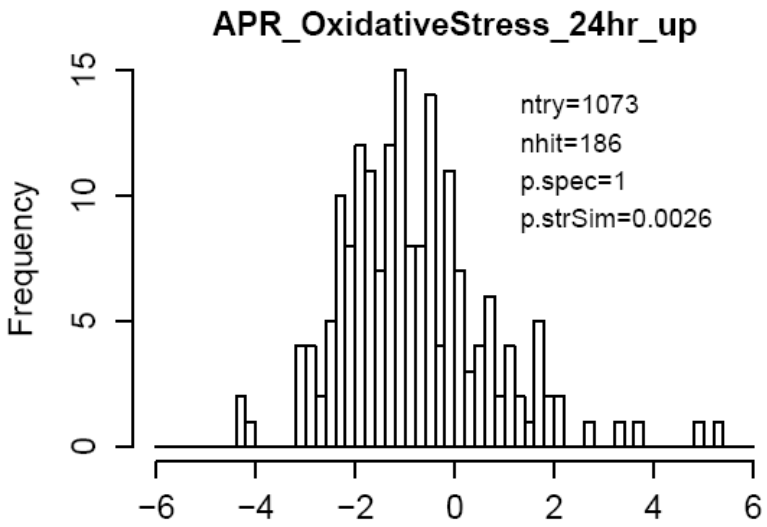
“Hit” (AC50) in burst region is less likely to result from specific activity (e.g. binding to receptor or enzyme)

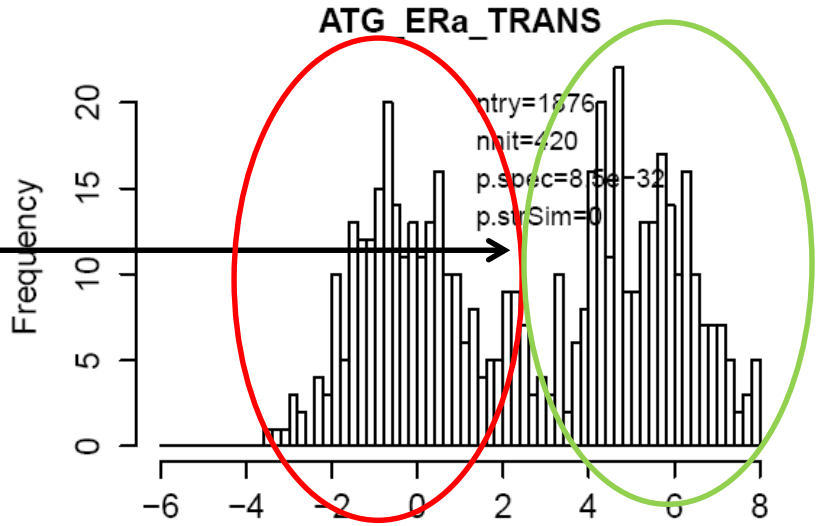
Z-score: # of SD from burst center
-High Z: more likely to be specific
-Low Z: less likely to be specific



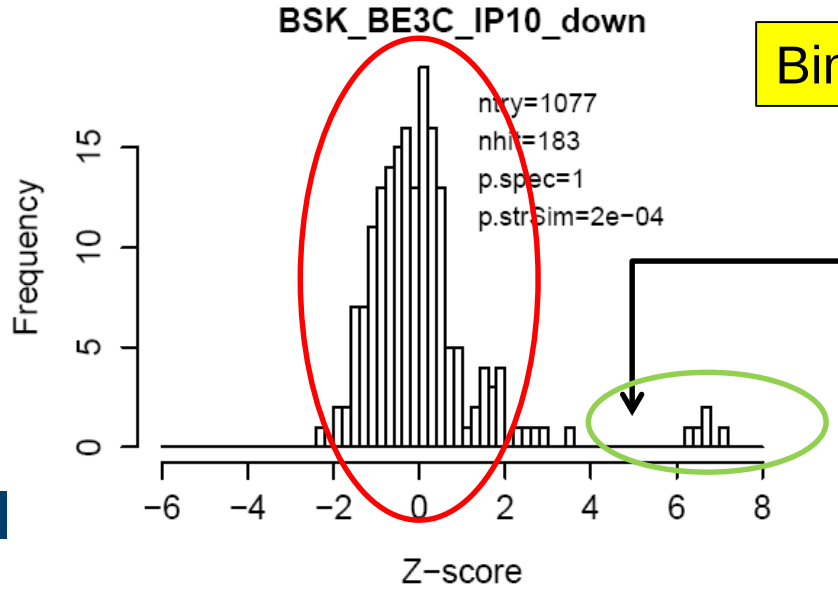
Examine Z-scores by assay

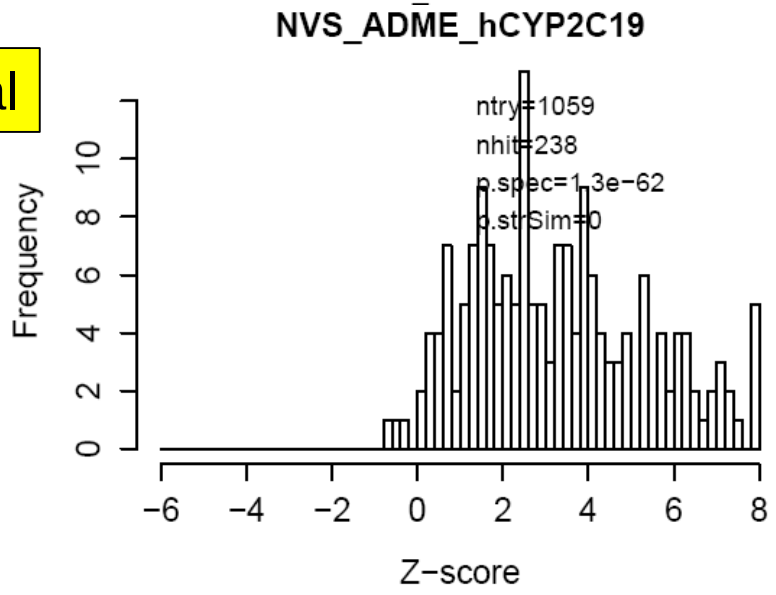
Cytotox / Cell Stress
"True" activity





Bimodal





Gene Score
Combine potency and specificity

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

- Potency: -log(AC50)
- Specificity: Z-score
- Gene score = Potency + Specificity
  - average over assays for gene [-log(AC50) + Z-score]
- Can be used to get quick ranking of chemicals
- Gene Score > 7 are most interesting
  - Z-score=2 and AC50=10 μM
  - 5670 chemical-gene combinations >7 (~1%)
  - 281 Genes (out of 330)
  - 1231 Chemicals (out of 1877)

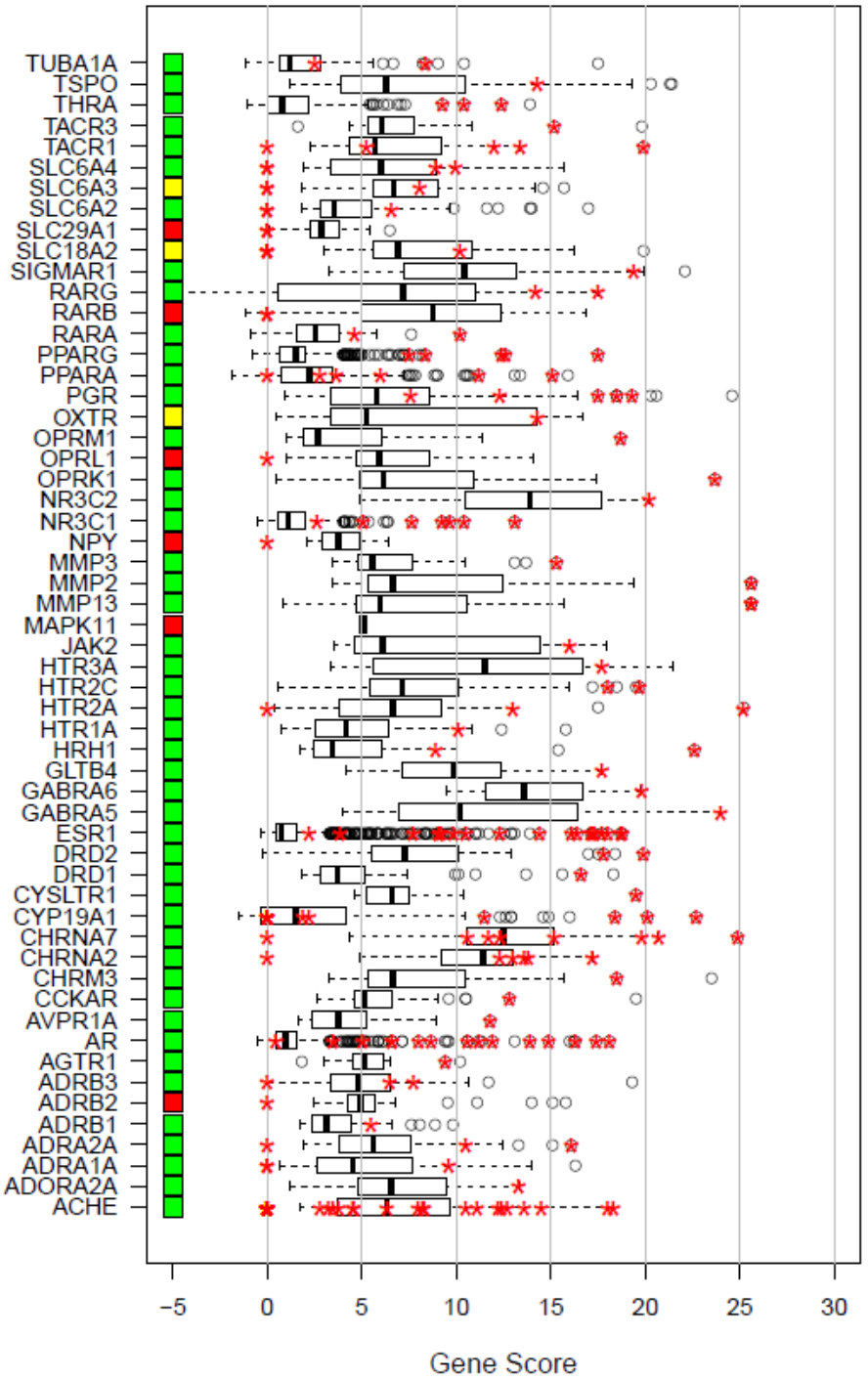
Do Assays Detect Potent Reference Chemicals?

* =Reference chemicals

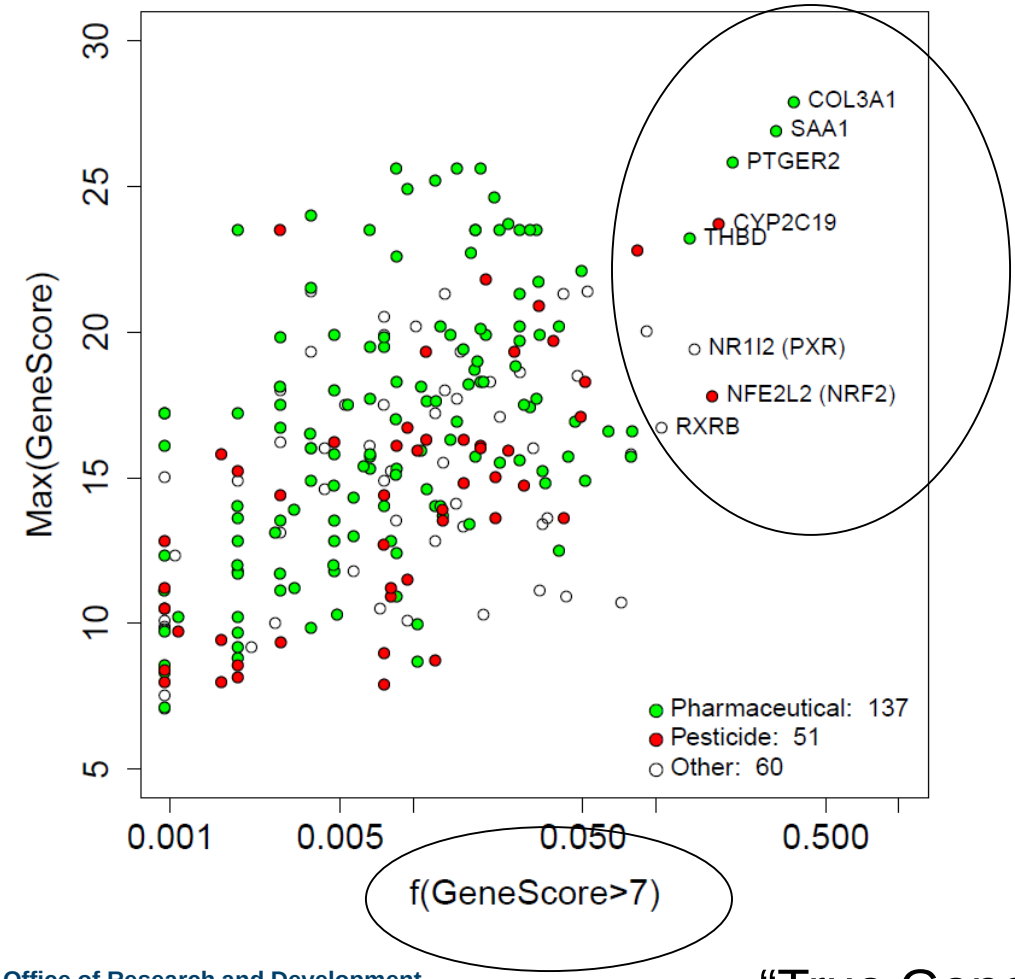
- These chemicals should be near the right of the gene score distribution

- Most assays show reference chemicals to be potent and specific

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



Chemicals with highest Gene Score are often those designed to be bioactive (75%)



Most promiscuous targets (>10% hit) after cytotox / non-specific filtering

"True Gene Promiscuity"

gene	Name	Intended Target	Use Category
COL3A1	Cariporide mesylate	Ion channel Na	Pharmaceutical
SAA1	YM218	AVPR1A	Pharmaceutical
PTGER2	PharmaGSID_47261	HIV nucleocapsid protein	Pharmaceutical
MMP13	CP-544439	ADAMx MMPx [MMP2 MMP3 MMP13]	Pharmaceutical
MMP2	CP-544439	ADAMx MMPx [MMP2 MMP3 MMP13]	Pharmaceutical
HTR2A	Volinanserin	HTR2A	Pharmaceutical
CHRNA7	PHA-00543613	CHRNA7	Pharmaceutical
PGR	Melengestrol acetate	NR3C1	Pharmaceutical
GABRA5	CP-457920	GABARx [GABAR1 GABRA5 GABRA6]	Pharmaceutical
CYP3A5	Malathion	ACHE	Insecticide
CYP2C19	Malathion	ACHE	Insecticide
OPRK1	PharmaGSID_47258	OPRK1	Pharmaceutical
CHRM1	PharmaGSID_48509		Pharmaceutical
CHRM2	PharmaGSID_48509		Pharmaceutical
CHRM3	PharmaGSID_48509		Pharmaceutical
CHRM4	PharmaGSID_48509		Pharmaceutical
EDNRA	MK-547	EDN1	Pharmaceutical
EDNRB	MK-547	EDN1	Pharmaceutical
HRH2	Piragliatin	GCK	Pharmaceutical
PDE4A	FR140423	Opiod receptors	Pharmaceutical
PTPRB	2-Bromo-4-hydroxyacetophenone		microbiocide
THBD	Triamcinolone	NR3C1	Pharmaceutical
CYP2B1	Bromuconazole	Sterol synthesis	Fungicide
H2AFX	Sorbic acid		Fungicide/antimicrobial
CYP19A1	Fadrozole hydrochloride	CYP19A1	Pharmaceutical
HRH1	Diphenhydramine hydrochloride	HRH1	Pharmaceutical
CYP2C6	Malathion	ACHE	Insecticide
SIGMAR1	Volinanserin	HTR2A	Pharmaceutical
CYP4F12	Flufenpyr-ethyl		
PTGS1	Indomethacin	PTGS2	
CYP2A2	Metconazole	Sterol synthesis	
HTR3A	ddI; Didanosine	HIV Reverse Transcriptase	
HDAC6	Acetamide		Solvent/plasticizer
TSPO	C.I. Acid Red 114		Dye

Green: Gene is intended target of the chemical

Promiscuity measures

Calculate the number of genes hit with Gene Score>7

"Hottest" – Most Specific Hits

Category	Nchem	Mean(Hit Ratio)	SD (Hit Ratio)	p-hot
conazole (triazoles)	13	0.045	0.0205	3.19E-06
Pharma Class 4.86	10	0.0484	0.0234	2.39E-05
Pharma Class 4.58	11	0.0555	0.0374	4.65E-05
organometallic	5	0.0576	0.0247	0.00134
Pharma Class 3.292	5	0.0555	0.0405	0.00619

"Coldest" – Fewest Specific Hits

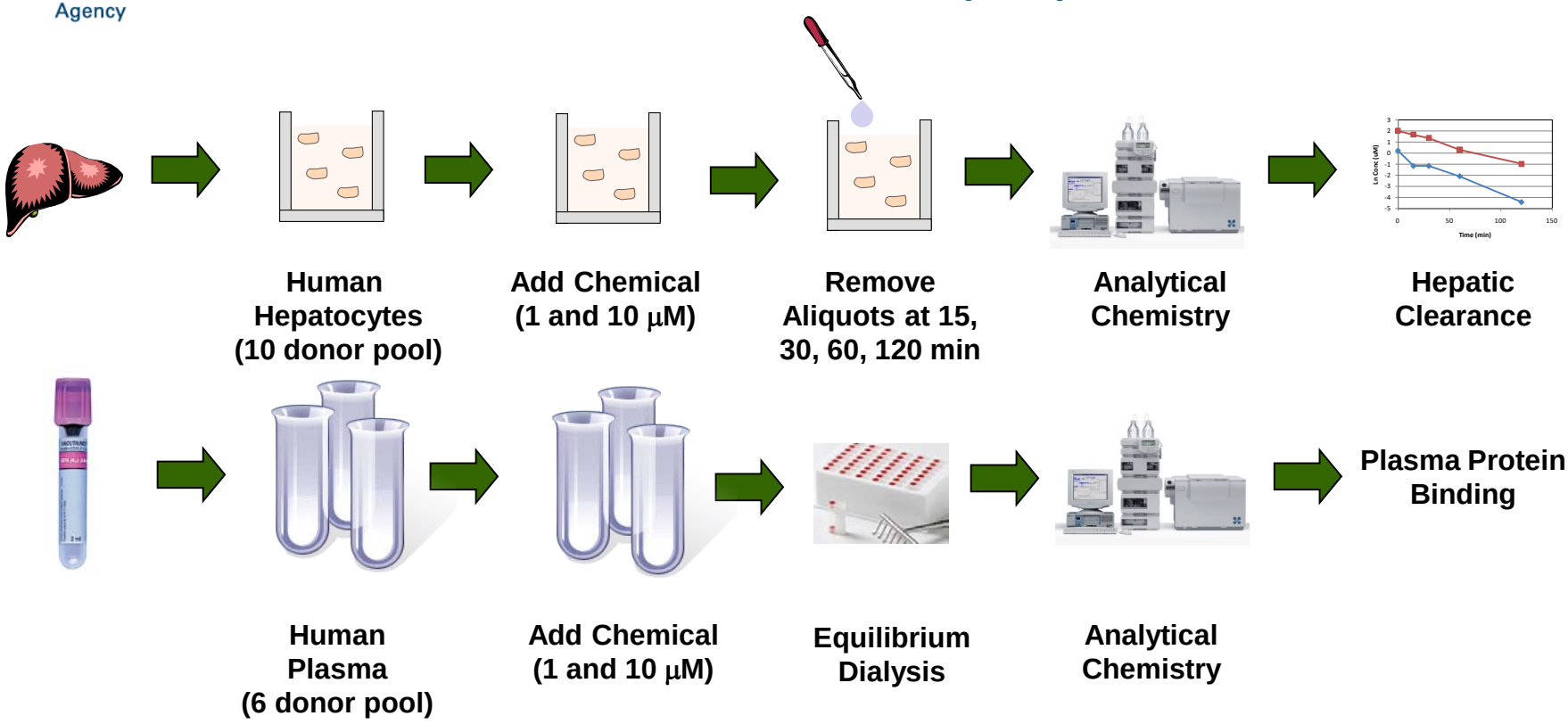
Category	Nchem	mean_HitRatio	SD_HitRatio	p-cold
phthalate	17	0.0061	0.00665	0.000131
alcohol pri	10	0.00447	0.00362	0.000835
carboxylic acid	10	0.00584	0.0056	0.00335
carboxylate	7	0.0044	0.00473	0.00356
carboxylate di	15	0.0078	0.00655	0.00594

Learning from “non-specific hits”

- Hypothesis: In vivo, if a chemical reaches concentrations where cell stress or cytotoxicity occurs, animals will be ill
- Corollary: the cell stress / cytotoxicity level in vivo will be ~ maximum tolerated dose (MTD)

- Testing the hypothesis:
  - Use Reverse Toxicokinetics (RTK) to convert cytotoxicity concentrations (burst region) to dose
  - Compare with MTD

Adding Pharmacokinetics
Reverse ToxicoKinetics (rTK)

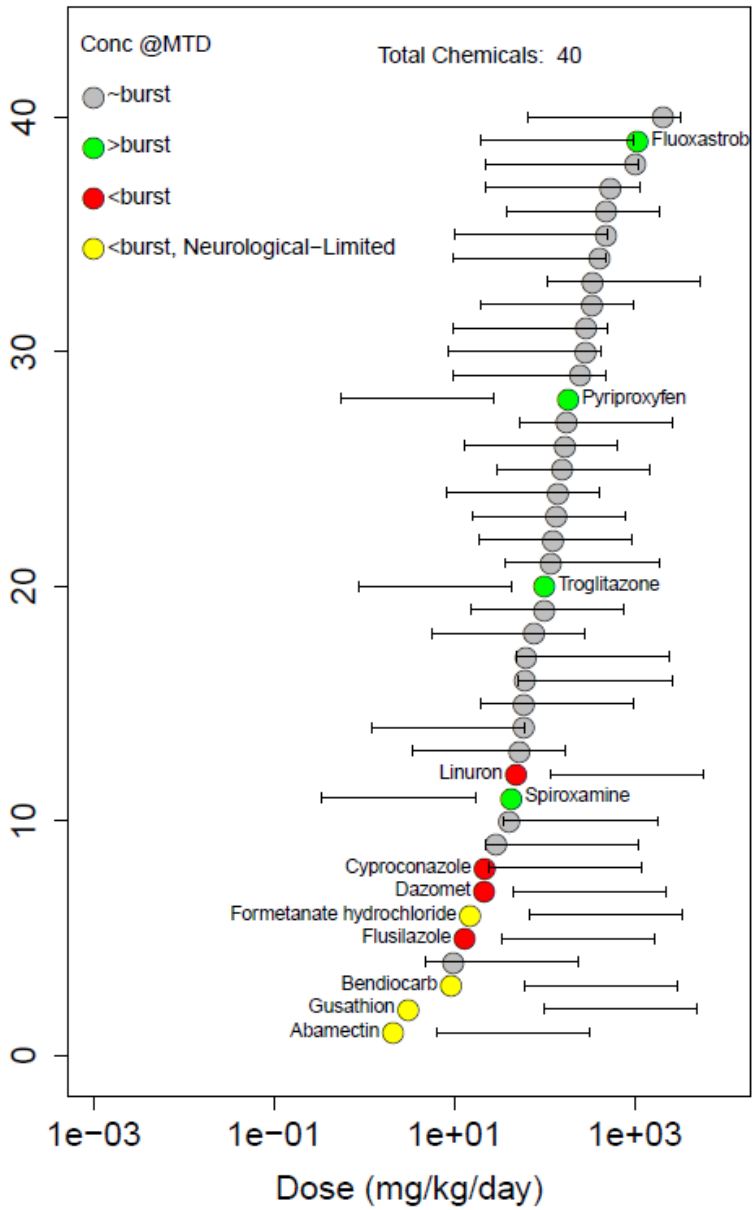


Combine experimental data w/ PK Model to estimate dose / concentration scaling

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

Comparing Burst to MTD



40 chemicals have rat RTK and Rat MTD data (Use MTD from 2-year Chronic/cancer studies)

4 are dose limited due to neurological effects

28/36 have MTD in burst region (78%)

3 show significant deviation – RTK is suspicious
-Pyriproxyfen
-Troglitazone
-Spiroxmine

4 have MTD lower than burst
-Linuron
-Cyproconazole
-Dazomet
-Flusilazole

ToxCast / Tox21 Overall Strategy

- Identify targets or pathways linked to toxicity (AOP focus)
- Develop high throughput assays for these targets or pathways
- Develop predictive systems models
  - *in vitro* → *in vivo*
  - *in vitro* → *in silico*
- Use predictive models:
  - Prioritize chemicals for targeted testing
  - Suggest / distinguish possible AOP / MOA for chemicals
- High Throughput Risk Assessments
- High Throughput Exposure Predictions

EPA National Center for Computational Toxicology Staff
February 5, 2013