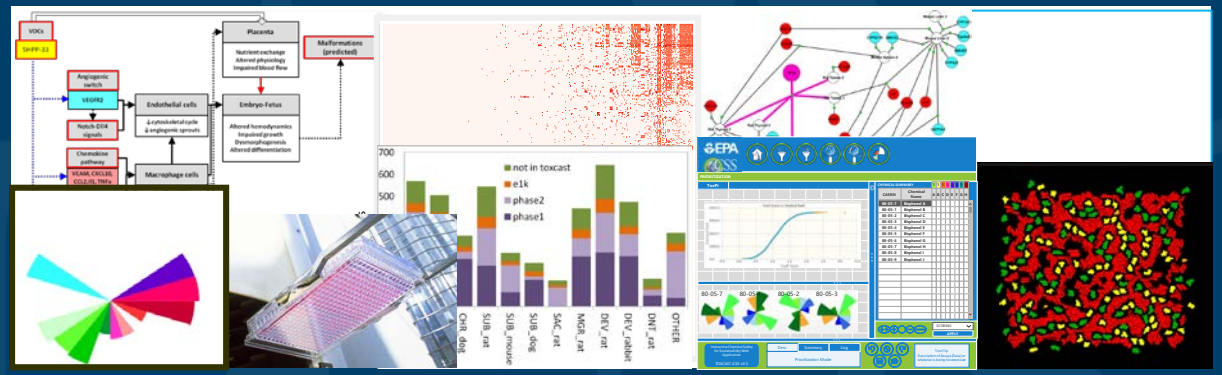
United States Environmental Protection Agency

Predictive Toxicology: Current Status and Future Outlook



EBI-EMBL Industry Programme Workshop
Predictive Toxicology

November 16-17, 2016

Imran Shah

National Center for Computational Toxicology

The views expressed in this presentation are those of the author[s] and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

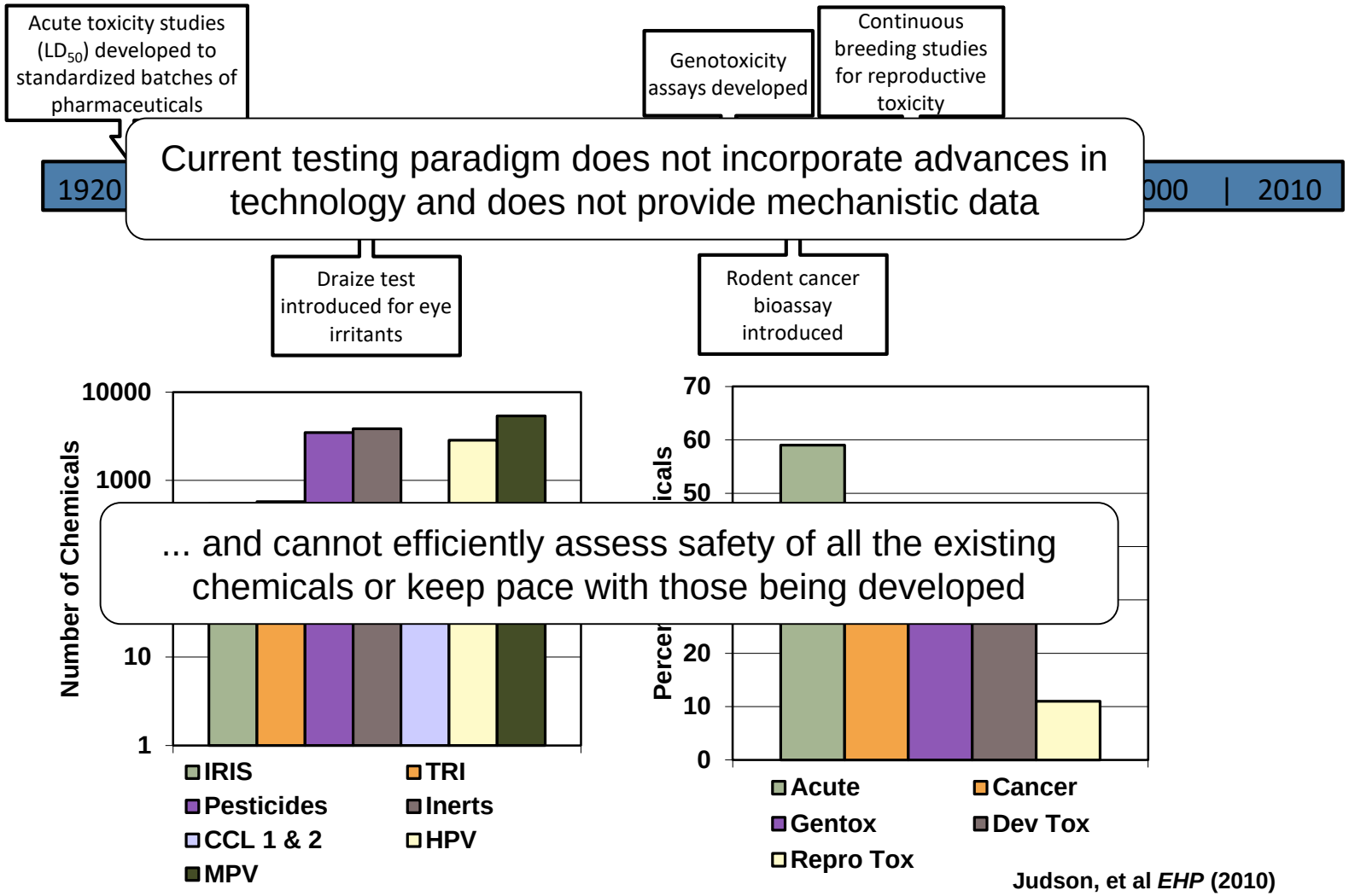
Regulatory Agencies Need to Make A Range of Decisions on Chemicals…

- Multiple drivers shape type of assessment
  - Regulatory scope
  - Economic considerations
  - Multiple applications
- Chemical assessments are “fit-for-purpose”
  - Prioritization (e.g., EDSP, PMN, SNUR)
  - Screening-level assessments (e.g., CCL, GreenChem)
  - Provisional assessments (e.g., PPRTVs)
  - Toxicity assessments (e.g., IRIS)
  - Risk assessments (e.g., MCLs, pesticides)

EPA	Workflow	Throughput	Data
OPPT	New chemicals: Premanufacture Notice (PMN) Existing chemicals: Significant New Use Rule (SNUR)	~1000/yr (90d/chem) ~84,000 total	**III (II)**
	Current Chemical Risk Assessments	~10 total	**I**
	DFE / Green Chemistry	~2500	**I, II, III**
OPP	Pesticide registration (PR)	~10 new/yr ~50 old/yr	**I**
	Pesticide re-registration	~1000/yr 24,576 total	**I**
OW	Chemical Contaminant List	6yr / ~6,000 total	**I,II,III**
	RegDet on CCL	Every 6yr / 90 total	**I**
	Unregulated Contaminant Monitoring	30/5yr	**I**
	Drinking Water Health Advisories		**II, III**
OLEM	**Spills** **Brownfields** **Super Fund**		

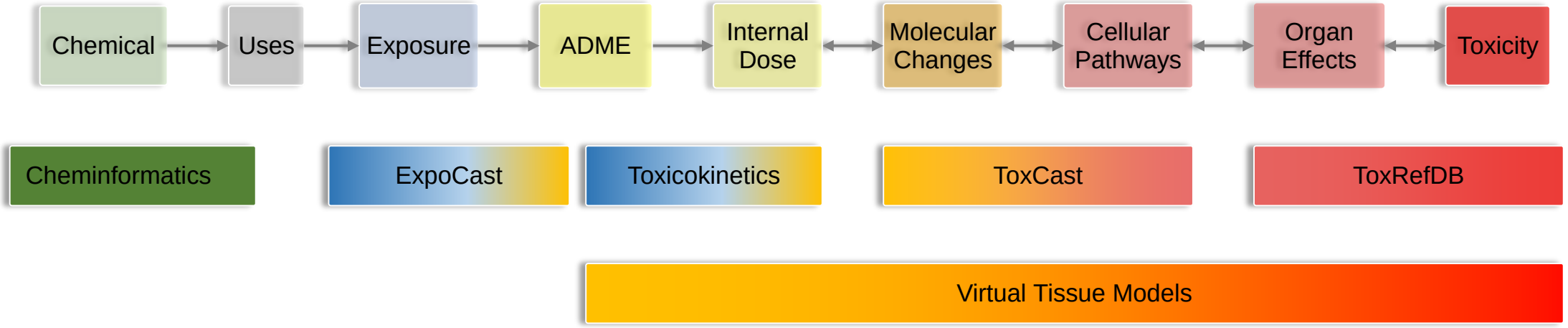
I Guideline animal testing data
II Some *in vitro* bioactivity
III Chemical structure data

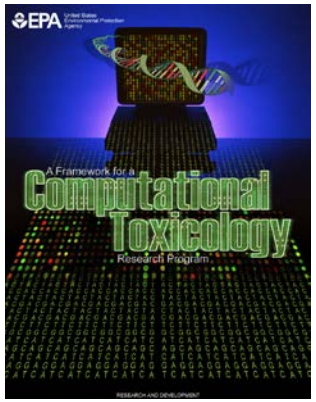
Current System to Evaluate Chemicals is Antiquated and Inefficient



Judson, et al *EHP* (2010)

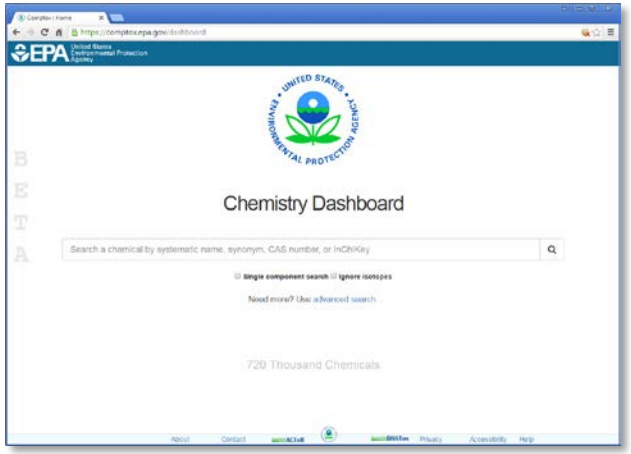
Requires an Integrated and Multidisciplinary Solution



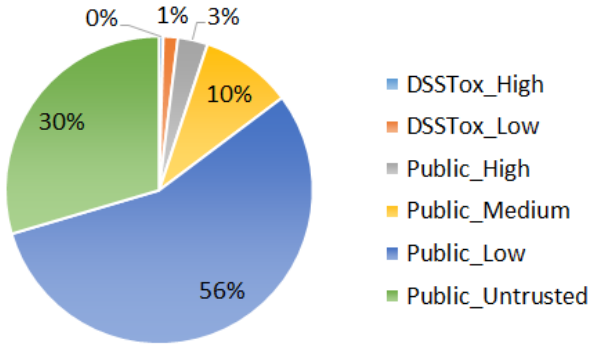


- National Center for Computational Toxicology established in 2005
- Currently staffed by ~60 employees
- Exists within the EPA's Office of Research and Development
- Home of the ToxCast and ExpoCast research efforts
- Key partner in U.S. Tox21 federal consortium

Need to Start with a High Quality Chemistry Foundation

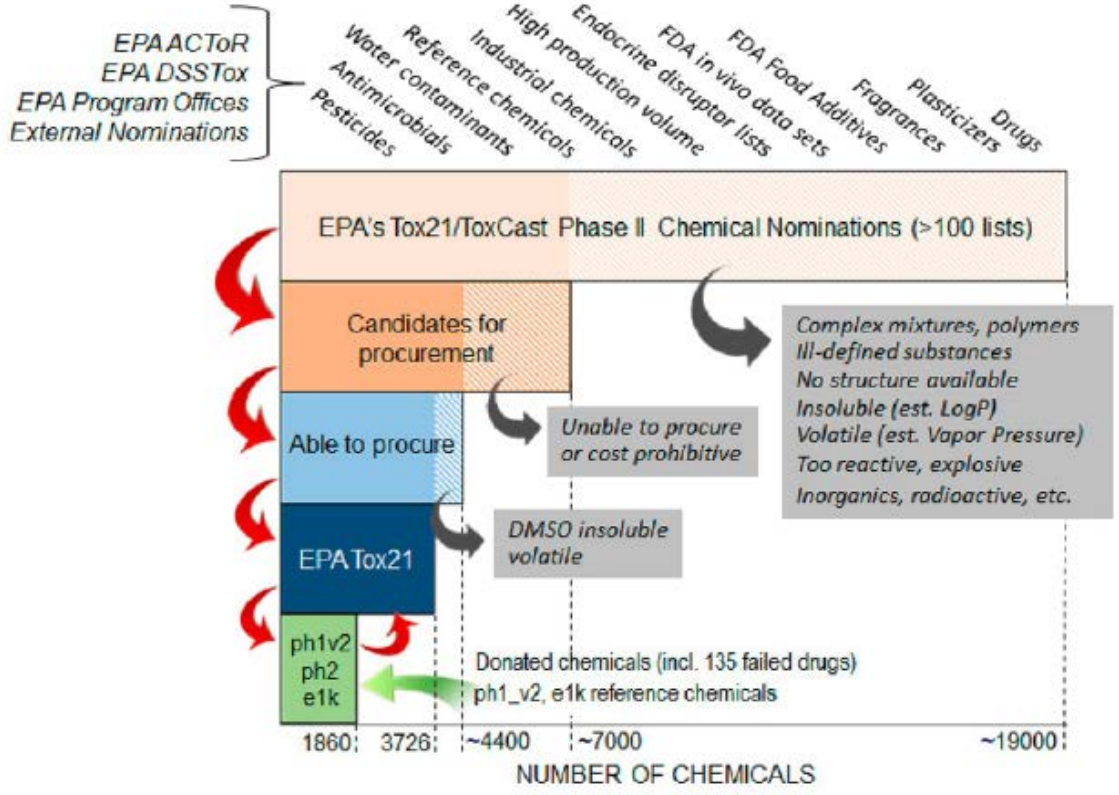


https://comptox.epa.gov



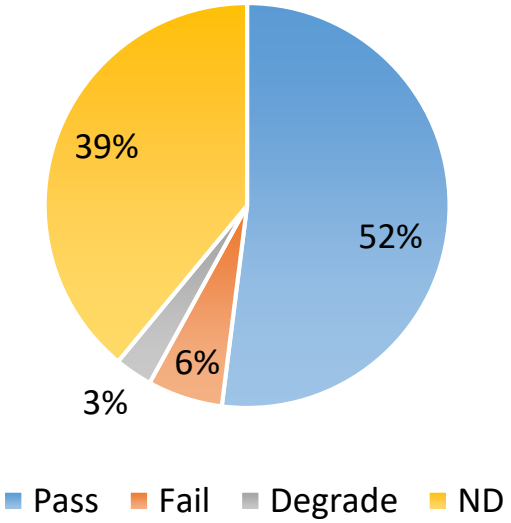
- Developing a centralized resource for curated chemical structure, identifier, and physical chemical properties of >700K unique substances with data quality flags
- Expand and curate training sets for QSAR models for phys-chem, environmental fate, and toxicological properties
- Use the centralized chemical resource as the foundation for an integrated hazard, bioactivity, pharmacokinetics, and exposure information

Need to Start with a High Quality Chemistry Foundation



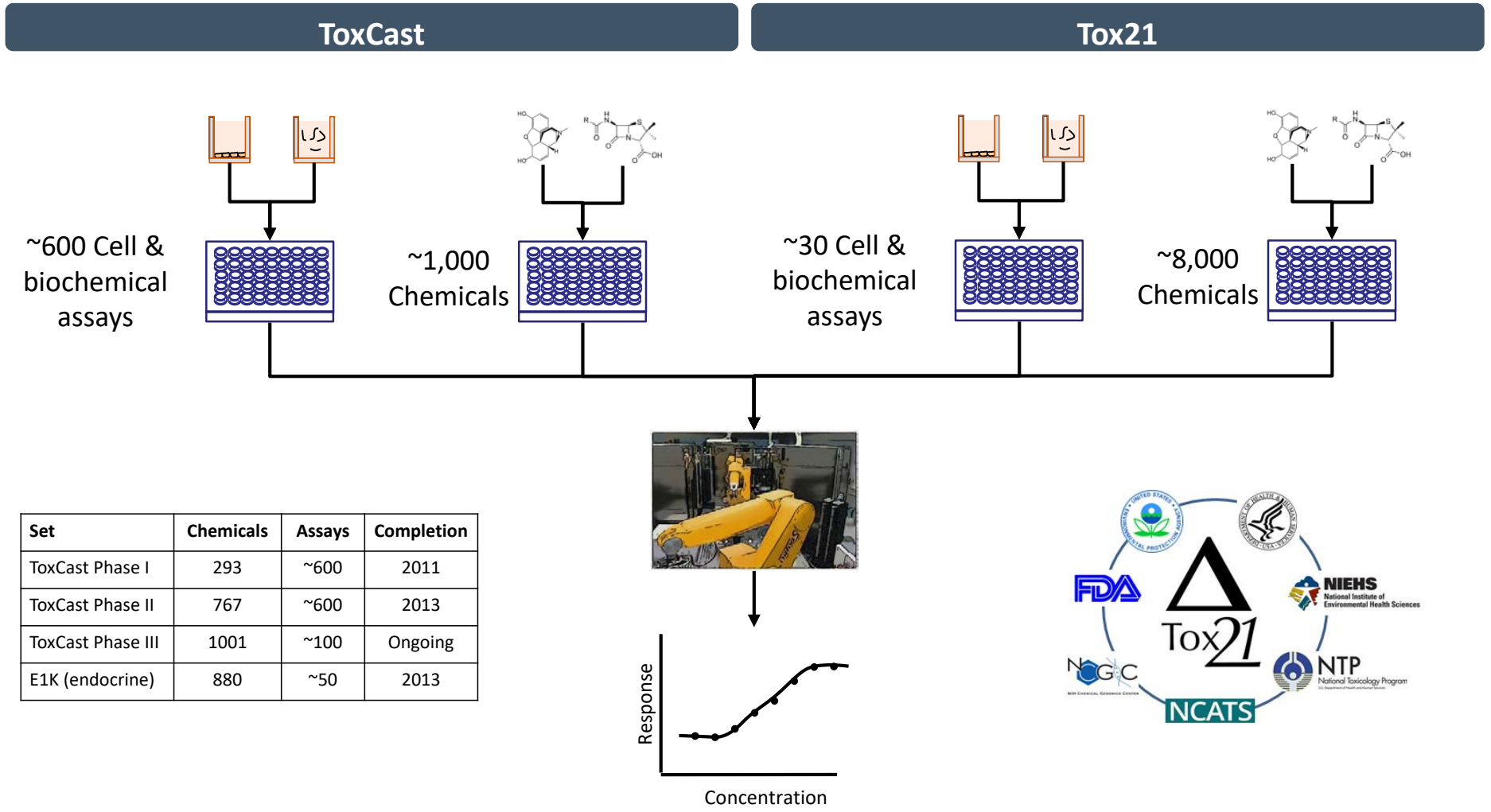
Richard et al., Chem Res Toxicol, 2016

Analytical QC of Chemical Library



Pass = C (75%) or greater
Fail = D, F, Ac, Bc, Cc

High-Throughput Bioactivity Screening



Set	Chemicals	Assays	Completion
ToxCast Phase I	293	~600	2011
ToxCast Phase II	767	~600	2013
ToxCast Phase III	1001	~100	Ongoing
E1K (endocrine)	880	~50	2013

ToxCast Incorporates a Diverse Array of High-Throughput *In Vitro* Assays

United States Environmental Protection Agency

Assay Provider
- ACEA
- Apredica
- Attagene
- BioReliance
- BioSeek
- CeeTox
- CellzDirect
- Tox21/NCATS
- NHEERL MESC
- NHEERL Zebrafish
- NovaScreen (Perkin Elmer)
- Odyssey Thera
- Vala Sciences

Biological Response
- cell proliferation and death
- cell differentiation
- Enzymatic activity
- mitochondrial depolarization
- protein stabilization
- oxidative phosphorylation
- reporter gene activation
- gene expression (qNPA)
- receptor binding
- receptor activity
- steroidogenesis

Target Family
- response Element
- transporter
- cytokines
- kinases
- nuclear receptor
- CYP450 / ADME
- cholinesterase
- phosphatases
- proteases
- XME metabolism
- GPCRs
- ion channels

Assay Design
- viability reporter
- morphology reporter
- conformation reporter
- enzyme reporter
- membrane potential reporter
- binding reporter
- inducible reporter

Readout Type
- single
- multiplexed
- multiparametric

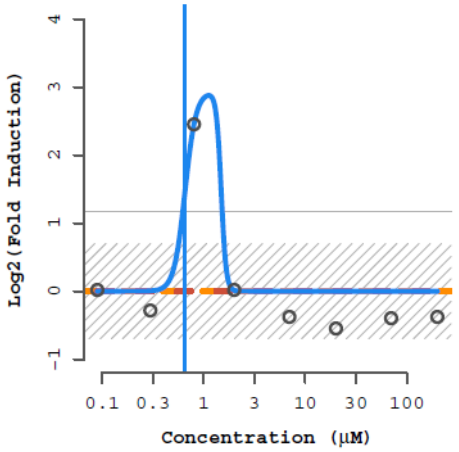
Cell Format
- cell free
- cell lines
- primary cells
- complex cultures
- free embryos

Species
- human
- rat
- mouse
- zebrafish
- sheep
- boar
- rabbit
- cattle
- guinea pig

Tissue Source
- Lung
- Liver
- Skin
- Cervix
- Uterus
- Intestinal
- Bladder
- Pancreas
- Inflammatory
- Breast
- Vascular
- Kidney
- Testis
- Brain
- Spleen
- Ovary
- Prostate
- Bone

Detection Technology
- qNPA and ELISA
- Fluorescence & Luminescence
- Alamar Blue Reduction
- Arrayscan / Microscopy
- Reporter gene activation
- Spectrophotometry
- Radioactivity
- HPLC and HPEC
- TR-FRET

Efforts to Ensure HTS Data Quality and Increase Transparency

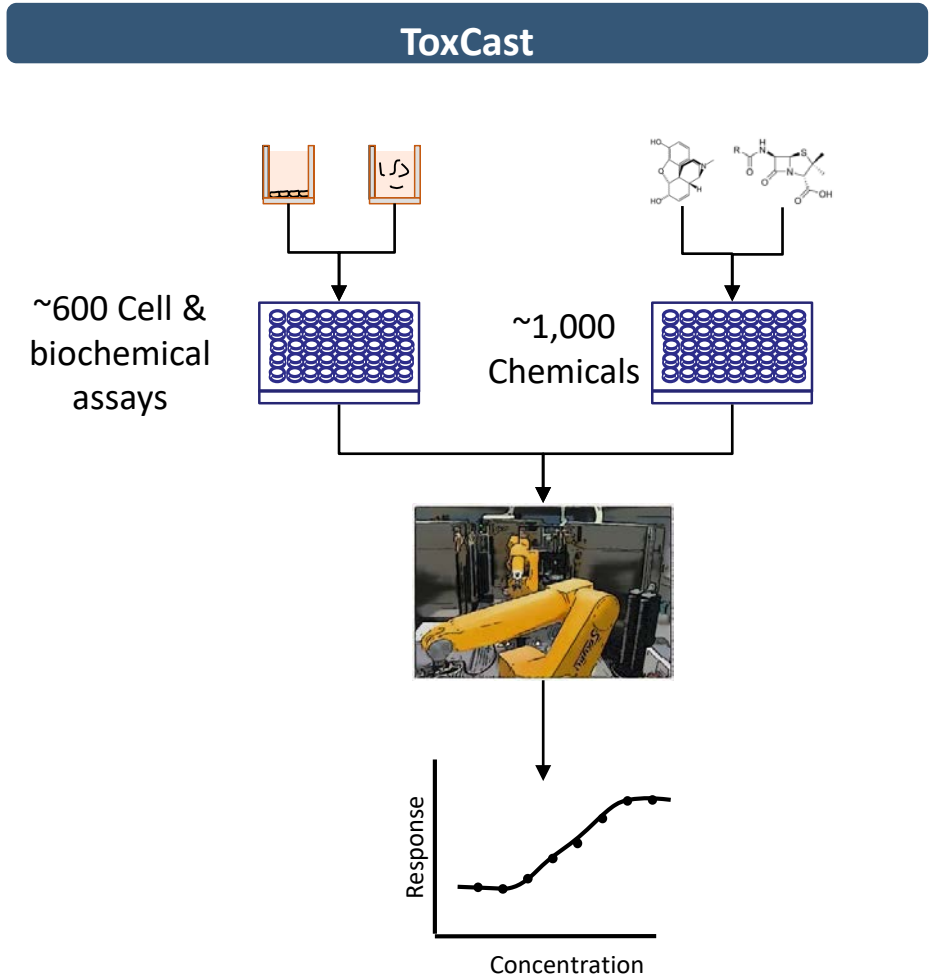


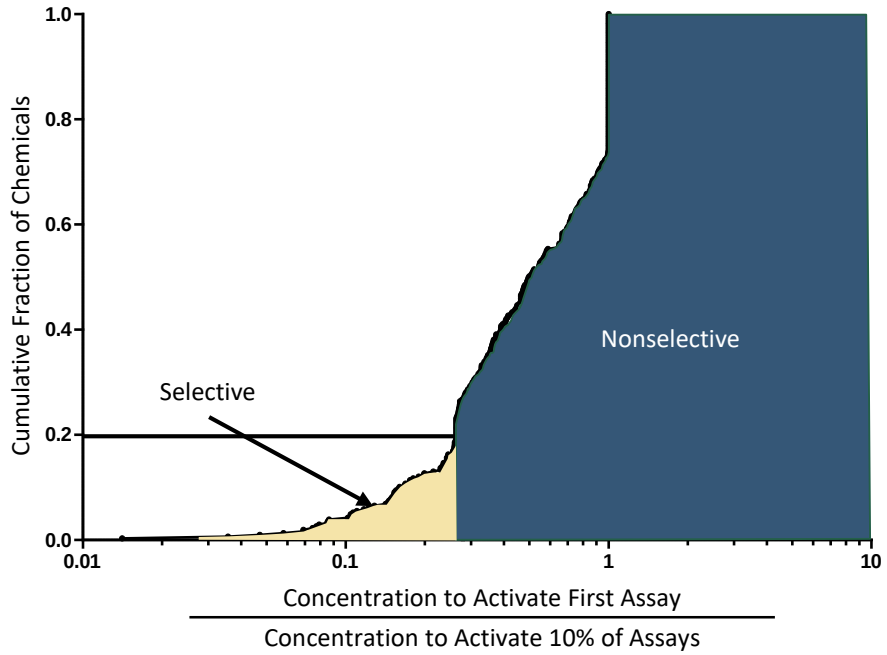

```
ASSAY:   AEID117 (ATG_ERa_TRANS)

NAME:    Thioglycolic acid
CHID:    26141   CASRN: 68-11-1
SPID(S): TX007664
L4ID:    420385

HILL MODEL (in red):
        tp        ga        gw
val:  3.1e-11   -2.15     0.416
sd:   NaN       NaN       NaN

GAIN-LOSS MODEL (in blue):
        tp        ga        gw        la        lw
val:  2.93      -0.184    8         0.173     18
sd:   3.56      0.334     9.48      5.82      814

        CNST        HILL        GNLS
AIC:  20.14       26.14       17.79
PROB: 0.23        0.01        0.76
RMSE: 0.92        0.92        0.32

MAX_MEAN: 2.45      MAX_MED: 2.45      BMAD: 0.233

COFF: 1.17   HIT-CALL: 1    FITC: 50   ACTP: 0.77
```

```
FLAGS:
Only one conc above baseline, active
Borderline active
```

- Public release of Tox21 and ToxCast data on PubChem and EPA web site (raw and processed data)
- Transparent ToxCast data analysis pipeline
  - Data quality flags to indicate concerns with chemical purity and identity, noisy data, and systematic assay errors
  - Publicly available as an R package
- Tox21 and ToxCast chemical libraries have undergone analytical QC and results publicly available
- Public posting of ToxCast procedures
  - Chemical Procurement and QC
  - Data Analysis
  - Assay Characteristics and Performance
- External audit on ToxCast data and data analysis pipeline

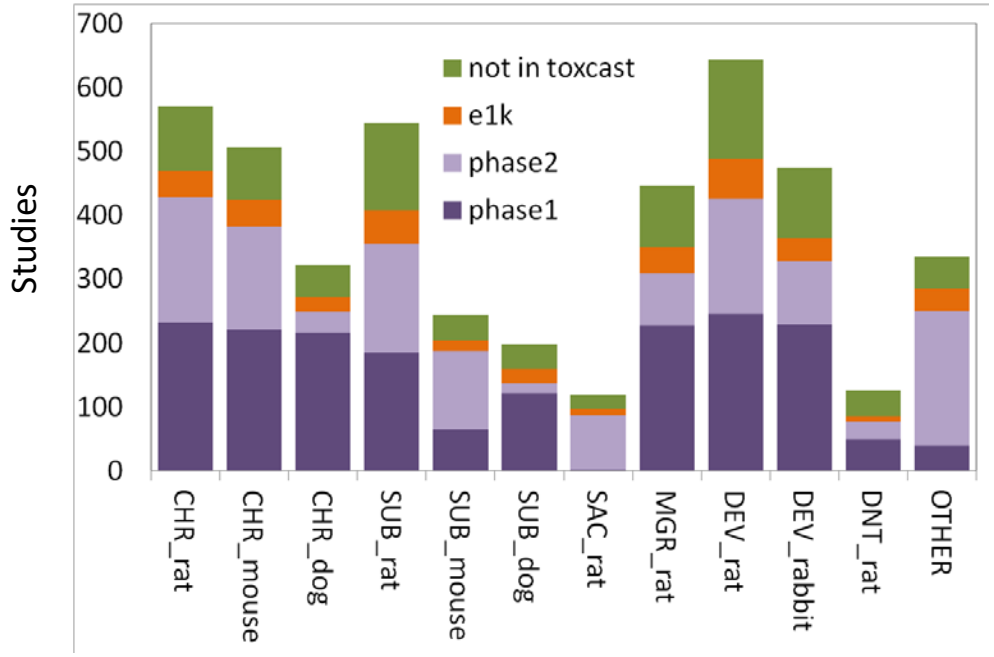
Most Environmental Chemicals are Nonselective for Biological Targets





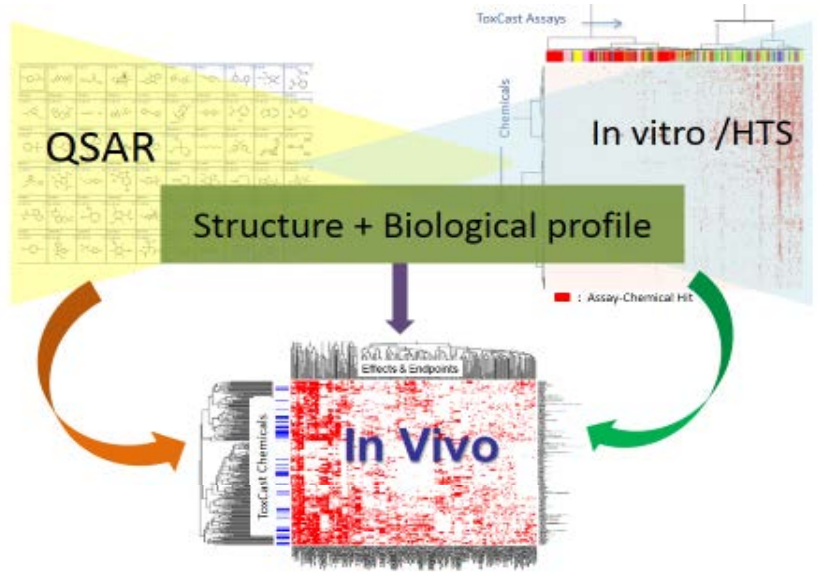
ToxRefDB: Digitizing Legacy *in vivo* Toxicology Data

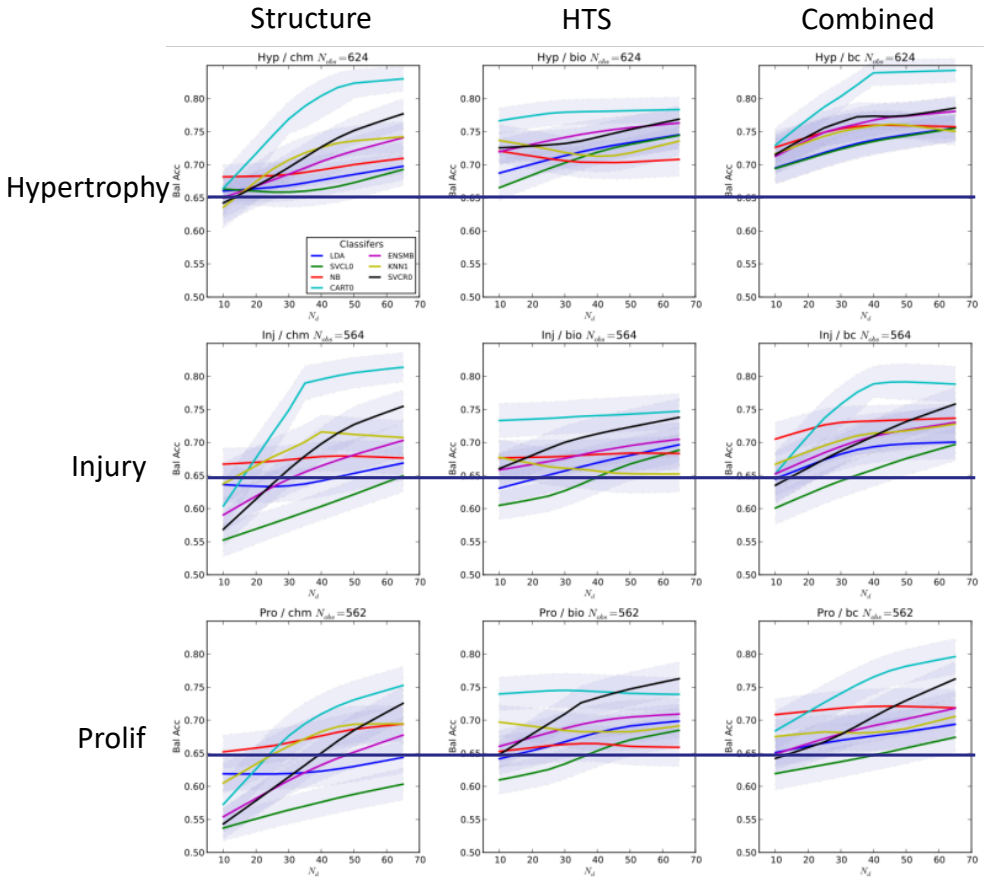
- ToxRefDB holds in vivo endpoint data from animal toxicology studies (DERs, NTP, open literature, pharma)
- Currently at 5567 studies on 1049 unique chemicals



Data Source	Study Count
EPA OPP DER	3279
Open Literature	731
National Toxicol Program	666
Sanofi_Pharma	222
Unpublished Submissions	50
GSK Pharma	38
Health Canada PMRA DER	23

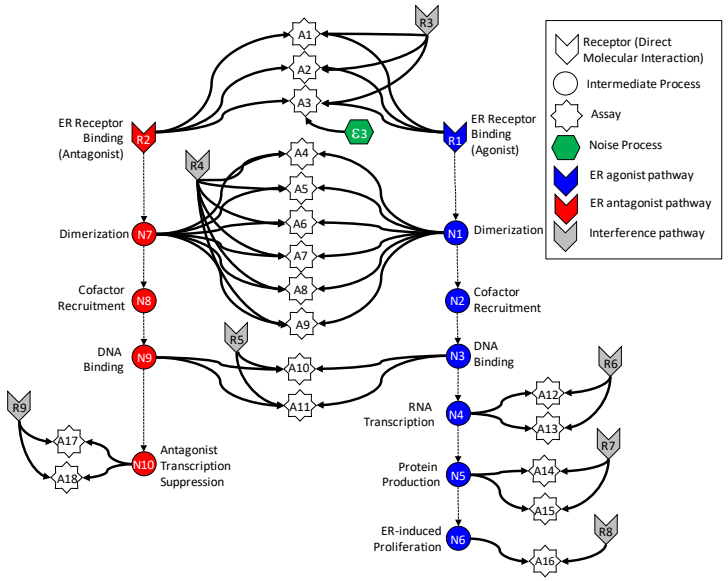
Predicting Target Organ Toxicities by Machine Learning





Liu *et al.*, CRT 2015

Developing a Pathway Model to Predict Endocrine Activity



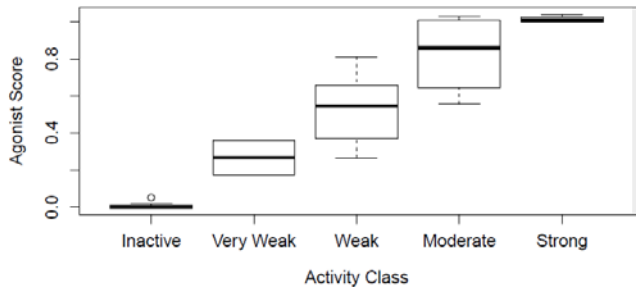
ER Pathway Model Integrating 18 *In Vitro* Assays

Performance for *In Vitro* Reference Chemicals

# True Pos	28
# True Neg	12
# False Pos	0
# False Neg	4
PPV	1.0
NPV	0.75
BA	0.94
Sensitivity	0.88
Specificity	1.0

Performance for *In Vivo* Uterotrophic Studies

# True Pos	28
# True Neg	12
# False Pos	1
# False Neg	1
PPV	0.97
NPV	0.92
BA	0.95
Sensitivity	0.97
Specificity	0.92

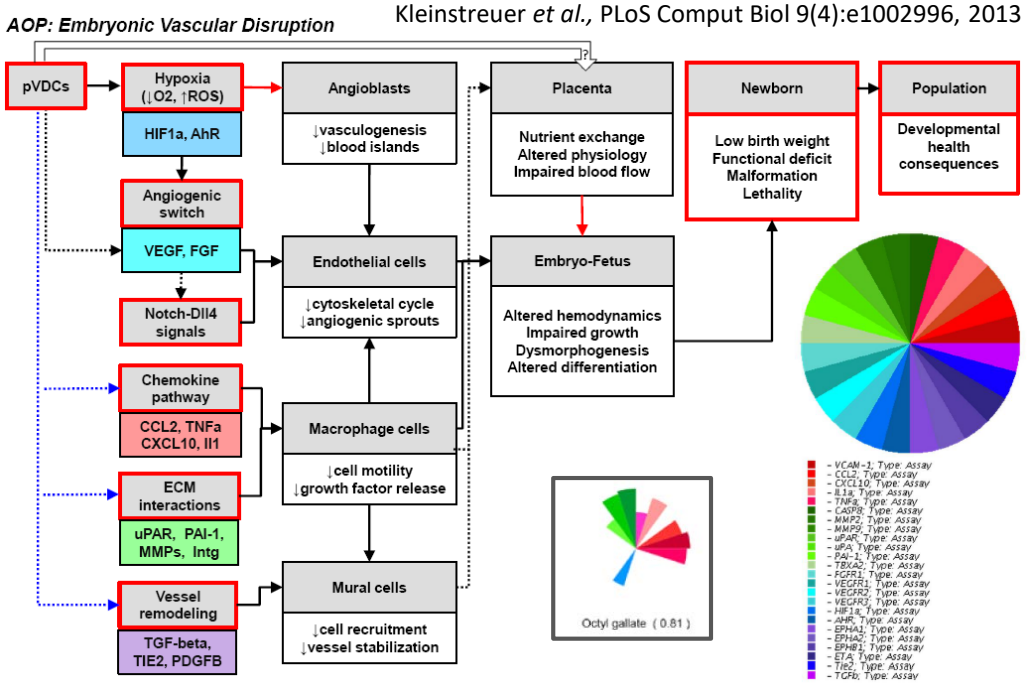


Judson *et al.,* ToxSci (in press)

ER Active Hit Rate

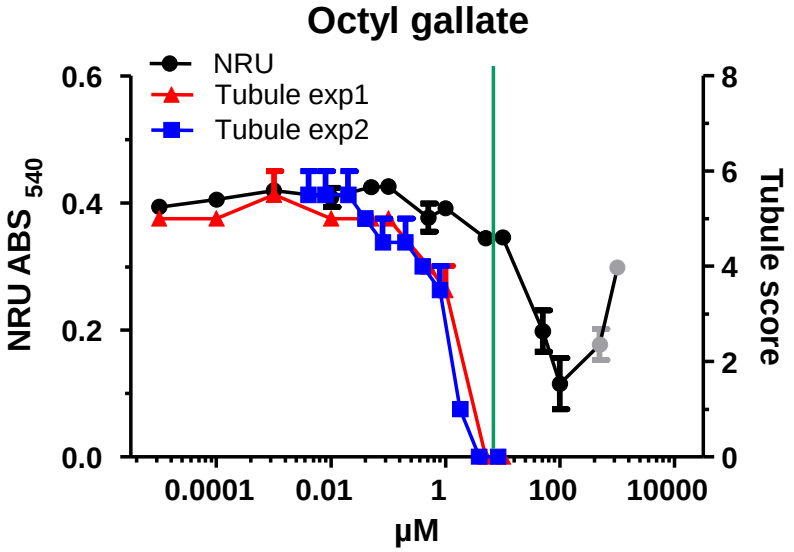
1431 EDSP chemicals run *in vitro*
71 (5%) have a significant ER score

Systems Biology Models To Scale Targets to Pathways and Networks – Virtual Tissues



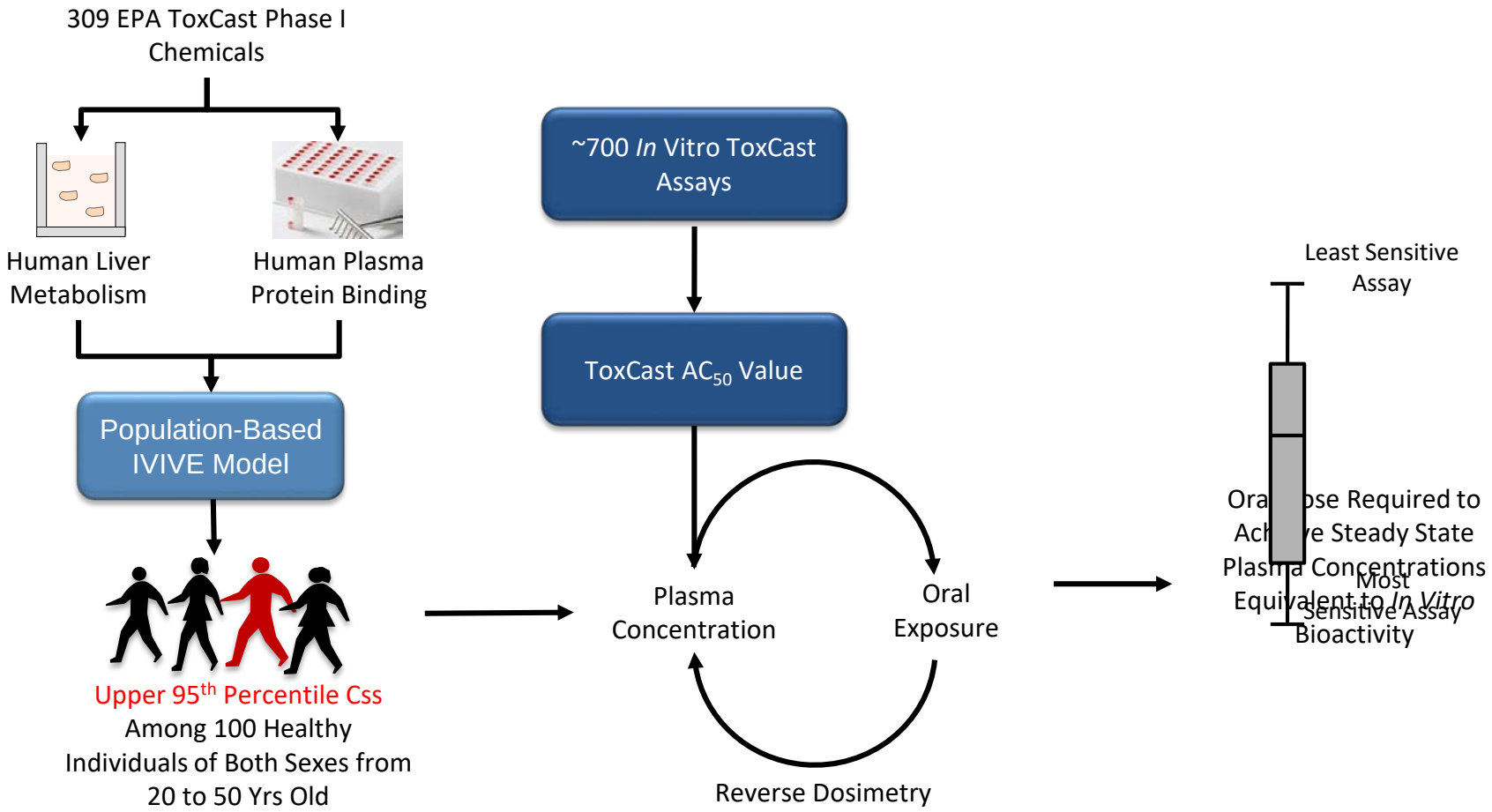
Kleinstreuer *et al.,* PLoS Comput Biol 9(4):e1002996, 2013

Human Tubulogenesis Assay (FICAM: T Heinonin)



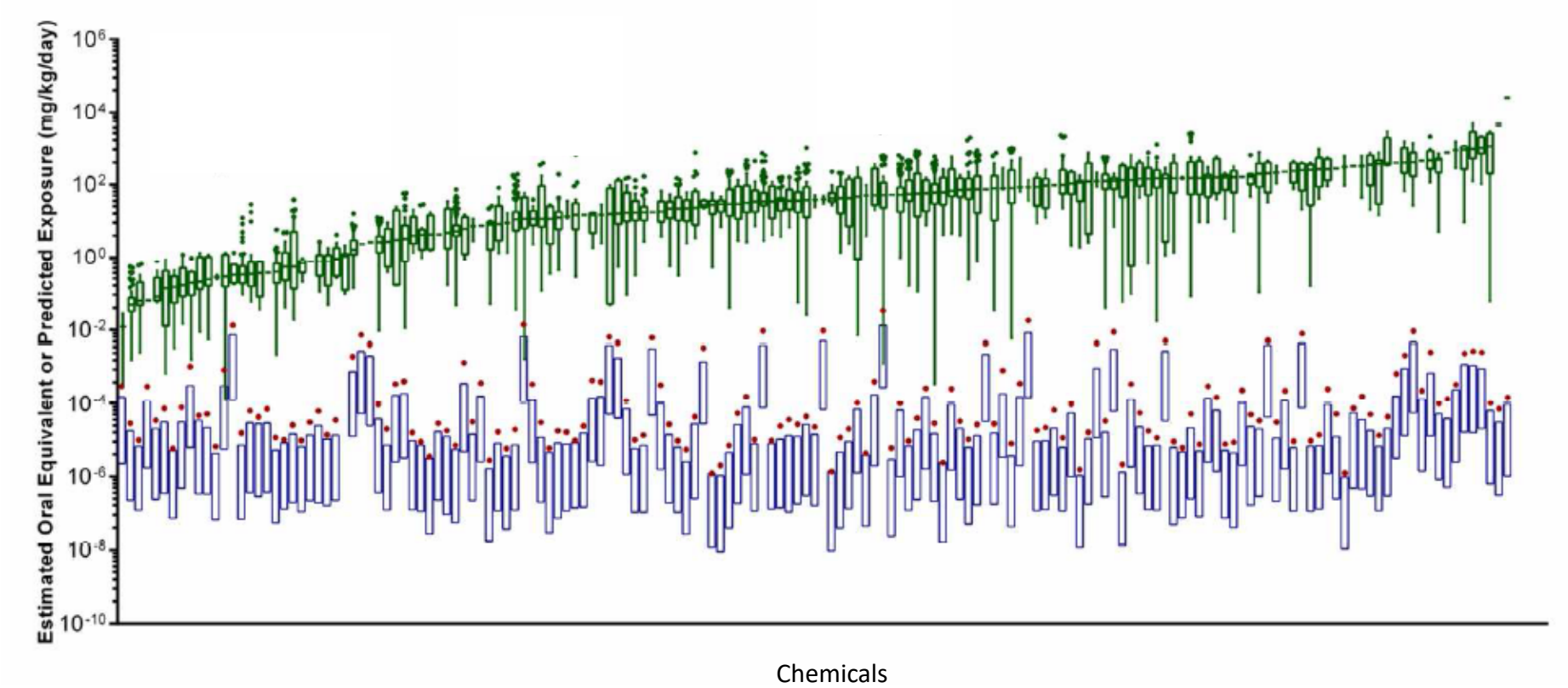
Cell ABM of Octyl Gallate (NCCT: G Nagaraj)

Providing Context by Incorporating Toxicokinetics and Exposure



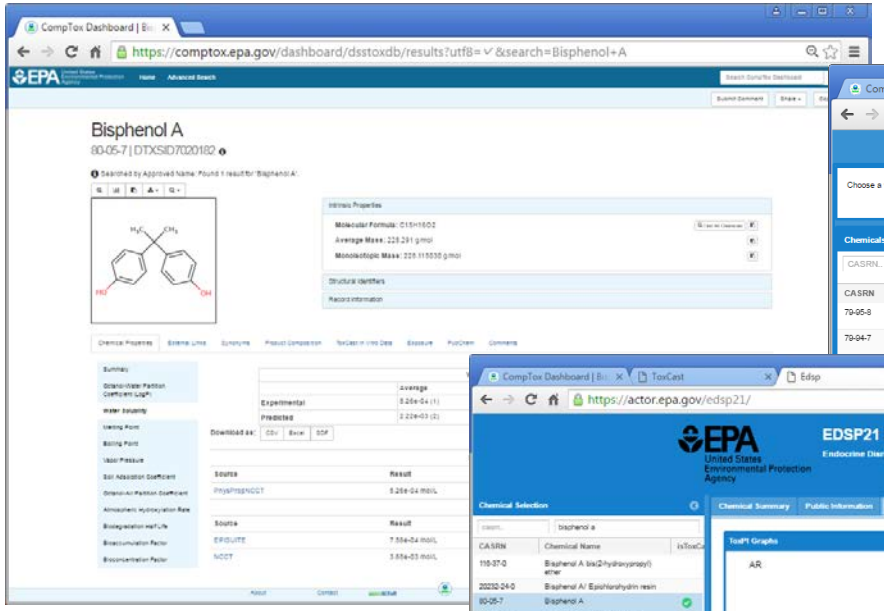
Rotroff *et al., Tox Sci.*, 2010
Wetmore *et al., Tox Sci.,* 2012

Comparing Bioactivity with Exposure Predictions for Risk Context

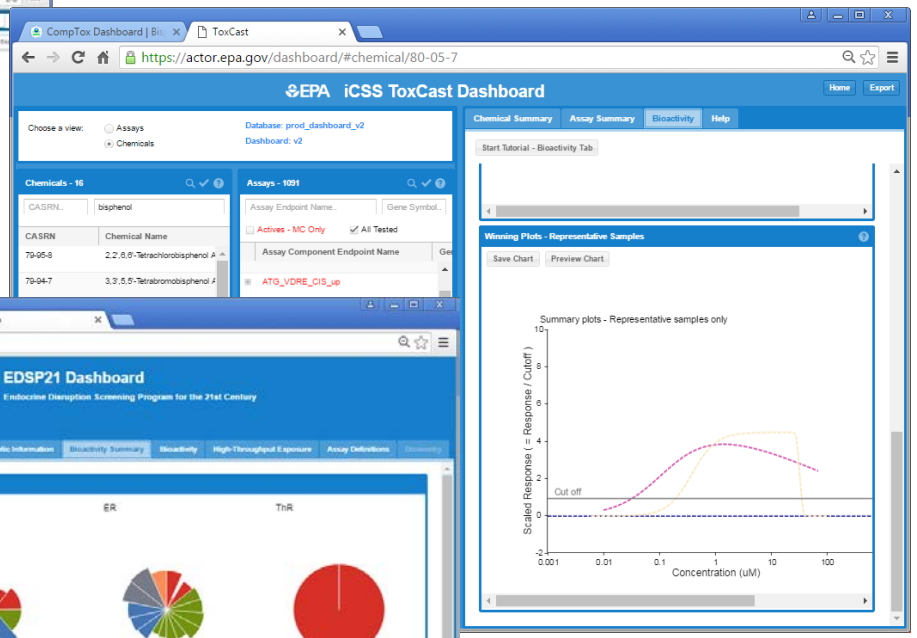


Wetmore *et al., Tox Sci.,* 2015

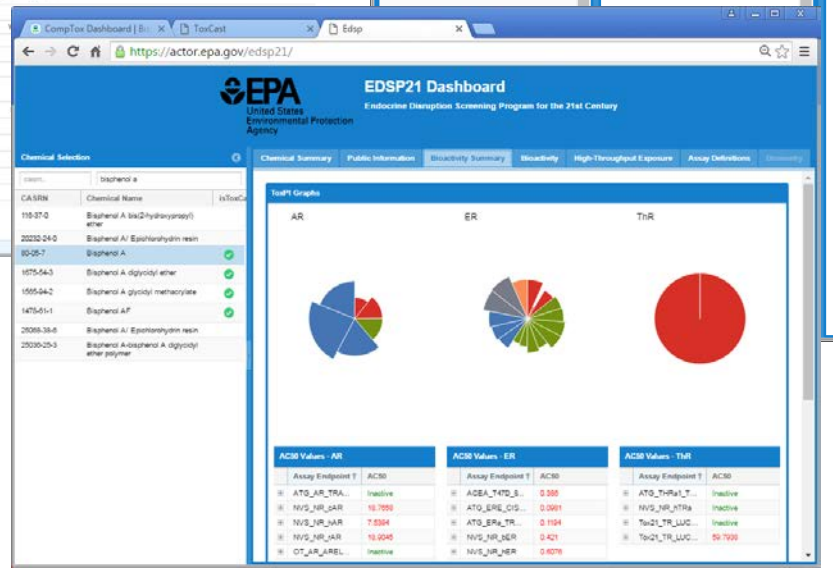
Delivering Data to Stakeholders and Scientific Community



Chemistry Dashboard (https://comptox.epa.gov/dashboard)

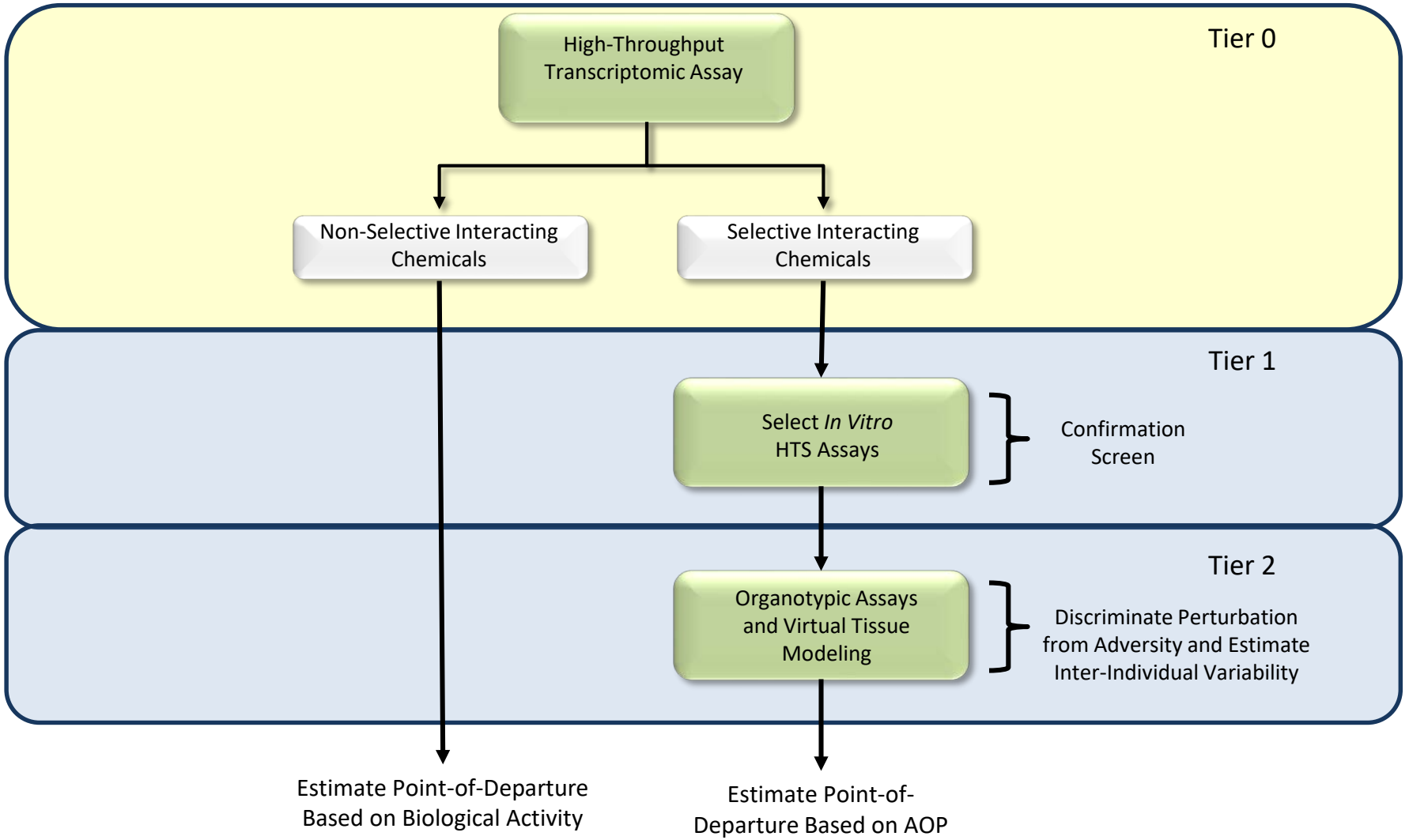


ToxCast Dashboard (https://actor.epa.gov/dashboard)



EDSP21 Dashboard (https://actor.epa.gov/edsp1)

Developing a Broad Hazard Screening Platform

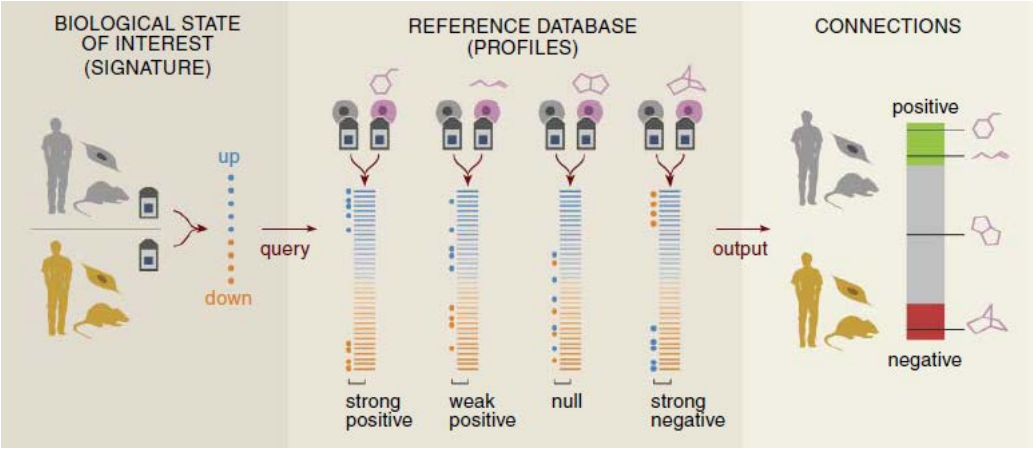


How Would a HT Transcriptomic Platform be Deployed?

High-Throughput Transcriptomic Assay

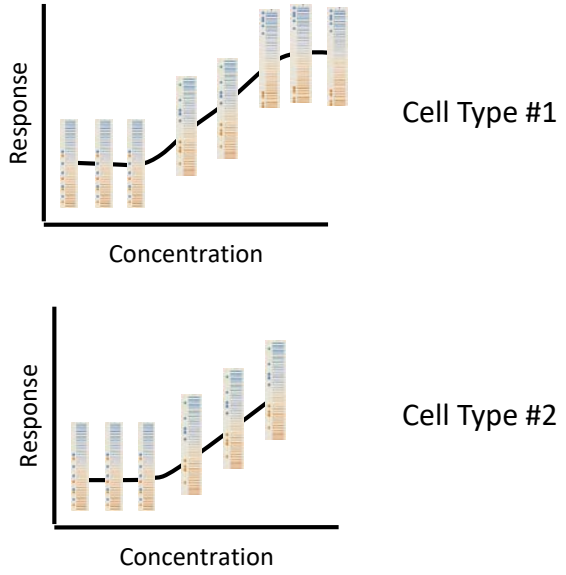
Tier 0

- **Identify predominant mechanisms as a function of concentration**
- Group chemicals by similar mechanism/bioactivity
- Identify a concentration that results in no transcriptional effects



Lamb et al. *Science* (2006)

Broad CMAPdb: 7,000 profiles; 1,309 compounds
NIH LINCs CMAPdb: 9,000 shRNAs, 3,000 over expression ORFs, and 4,000 compounds in 20 cell types/lines (cell lines and primary cells)



Cell Type #1

Cell Type #2

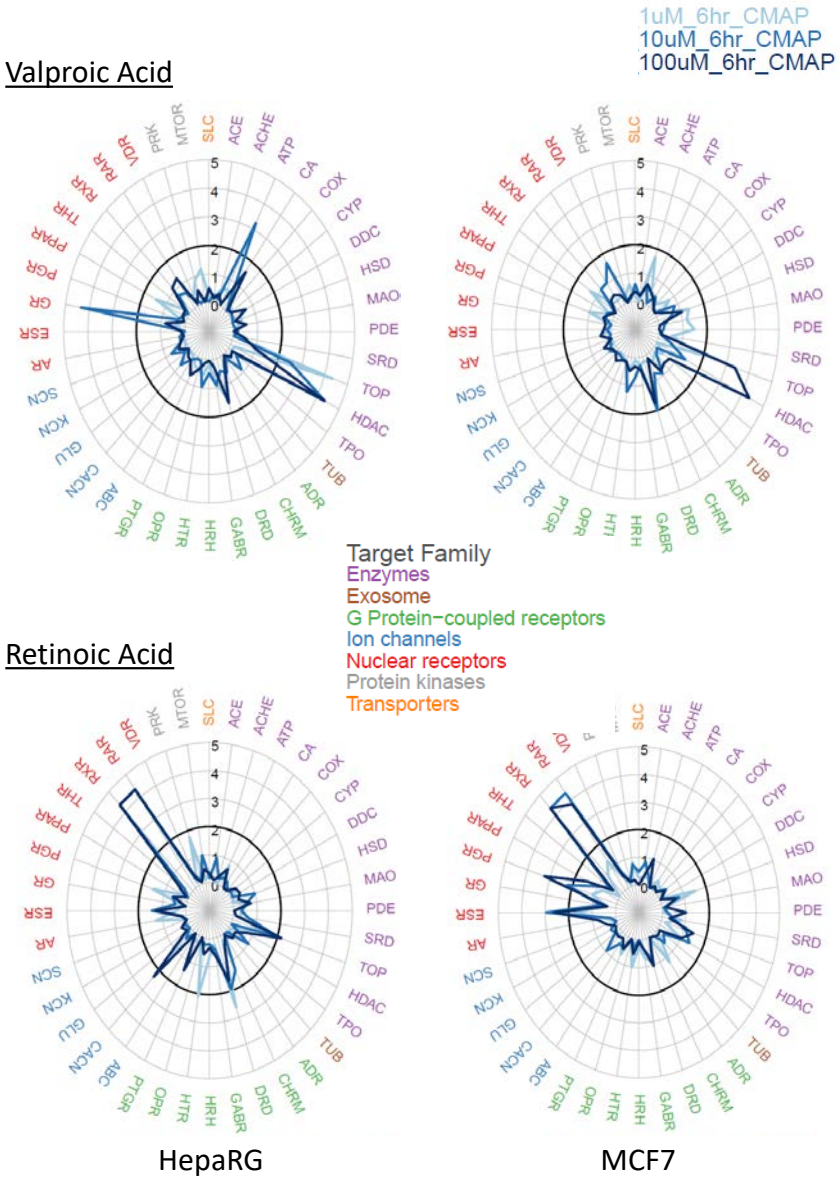
Cell Type #3

...

Using HT Transcriptomics to Identify Mode-of-Action

Target Family	Total Profiles	Target Genes	Chemicals	Cell Lines
Cytokine receptors	3	1	1	3
Enzymes	336	40	112	5
Exosome	14	1	4	4
G protein-coupled receptors	585	16	192	4
Ion channels	194	8	65	3
Nuclear receptors	227	10	71	5
Protein kinases	19	8	6	4
Transporters	102	2	35	3

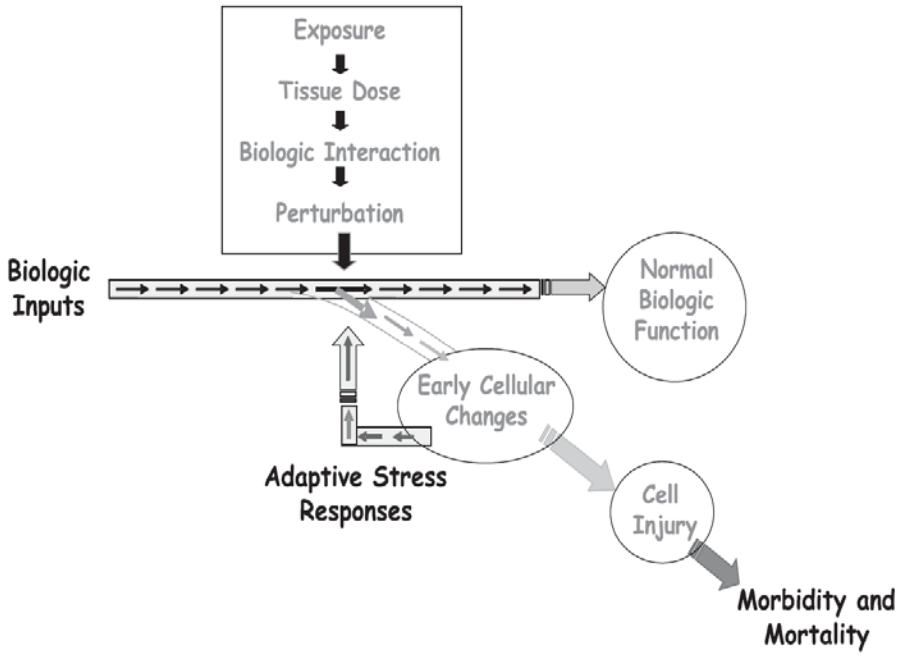
- Developed local database of Broad's CMAP data (~3,000 profiles)
- Annotated targets using KEGG (1,571 profiles)
- Significant genes identified using a z-score cutoff of 2
- Incorporated "JG" scoring method (Jiang and Gentleman 2007)
- Determine significance using a permuted rank approach across target family



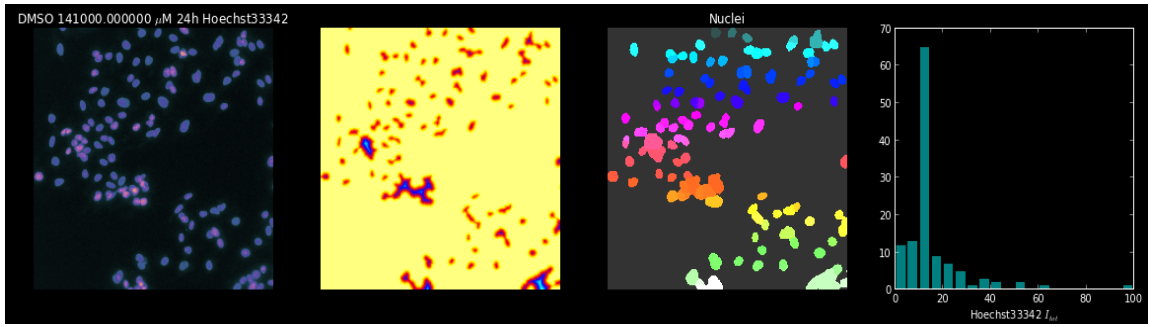
Martin and Karmaus, Unpublished

Distinguishing Adaptation from Adversity

- **Tipping Point**: Threshold between adaptation and adversity



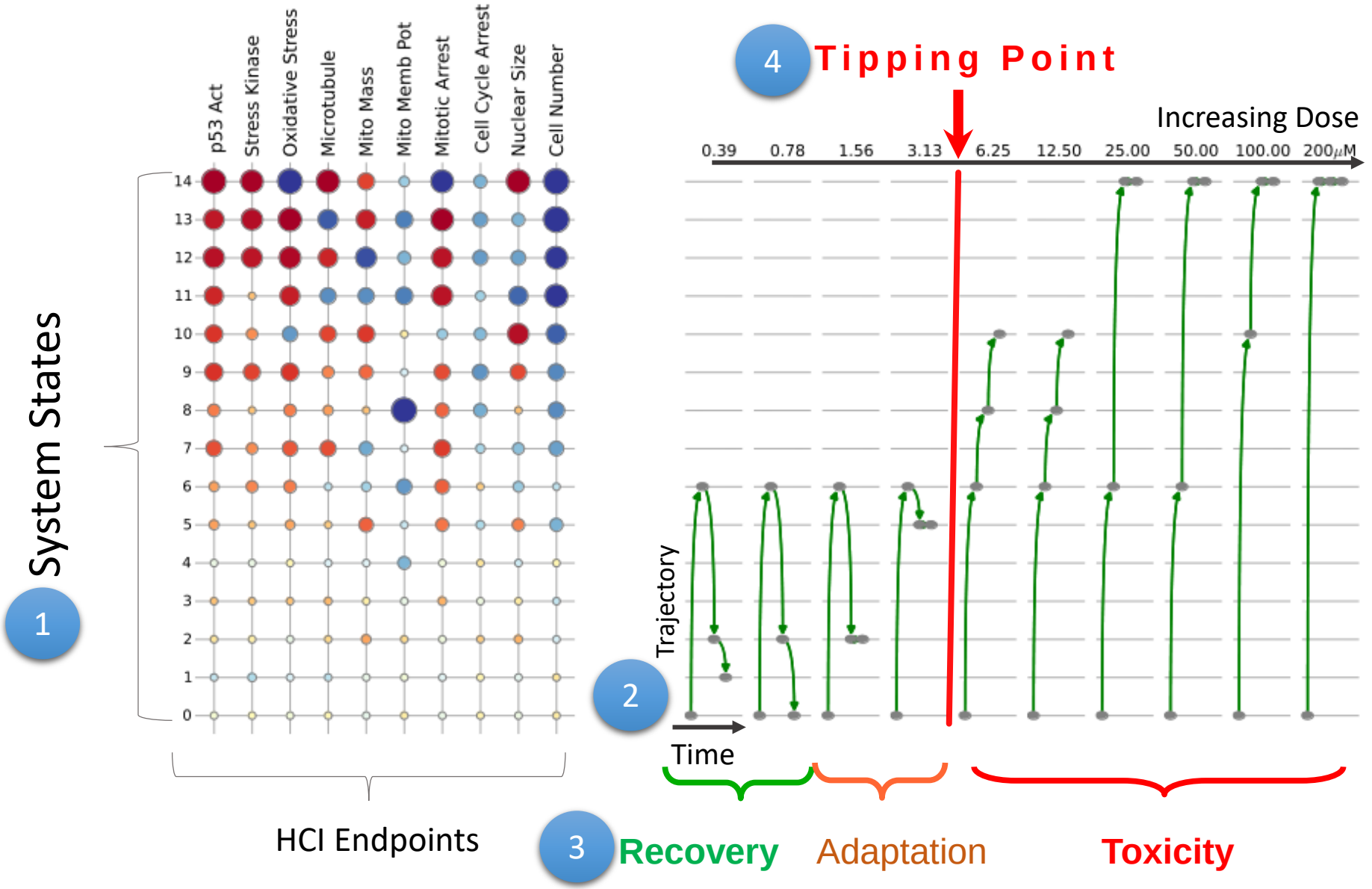
- Use ToxCast High Content Imaging (HCI) data to identify Tipping Points



- 967 chemicals (ToxCast)
- HepG2 cells culture
- 10 concentrations
- 3 Time points
- 10 HCI Assays
- 400 plates
- 100,000 wells
- 2,400,000 images

- Can we use **Tipping Point** to define a point of departure (PoD) for risk assessment ?

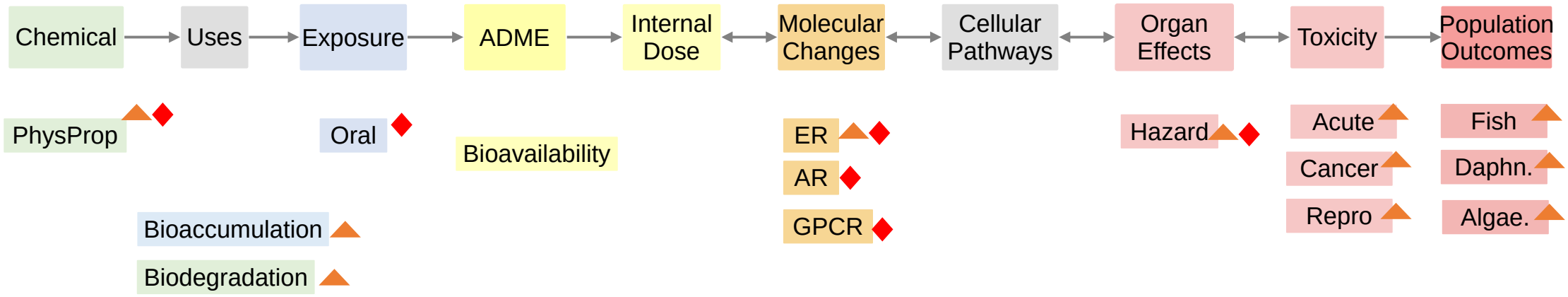
Tipping Point Analysis



Shah et al EHP 2016

Thousands of chemicals have limited data!

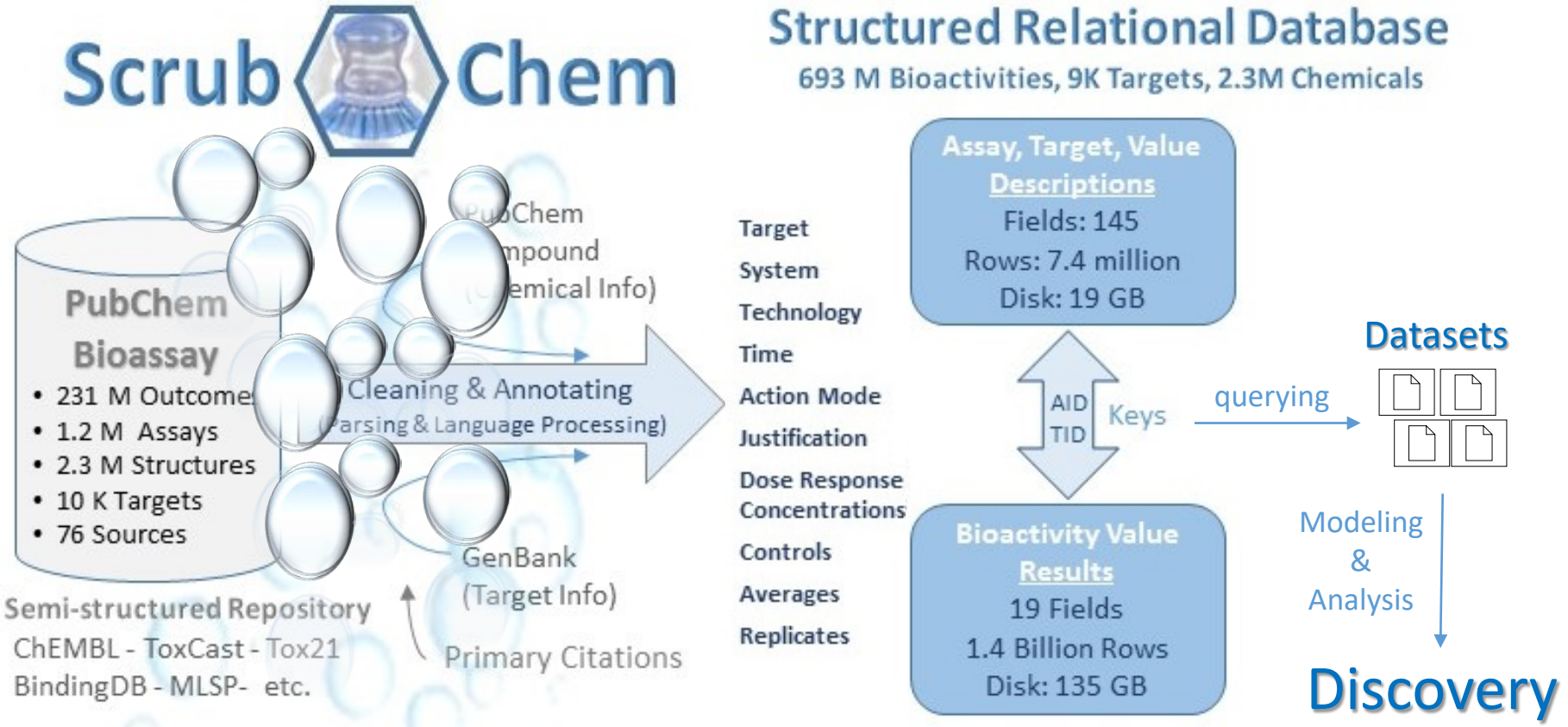
- Tens of thousands of environmental chemicals have very limited exposure/biological data
  - → Need more effective tools to describe chemical properties, effects and linkages
  - → Need predictive models to fill data gaps



▲ existing/legacy tools: TIMES, LeadScope, ECOSAR, EPIWIN

◆ developed internally: Read-across/GenRA, Machine Learning: classification and regression

Structuring PubChem Data for Analysis



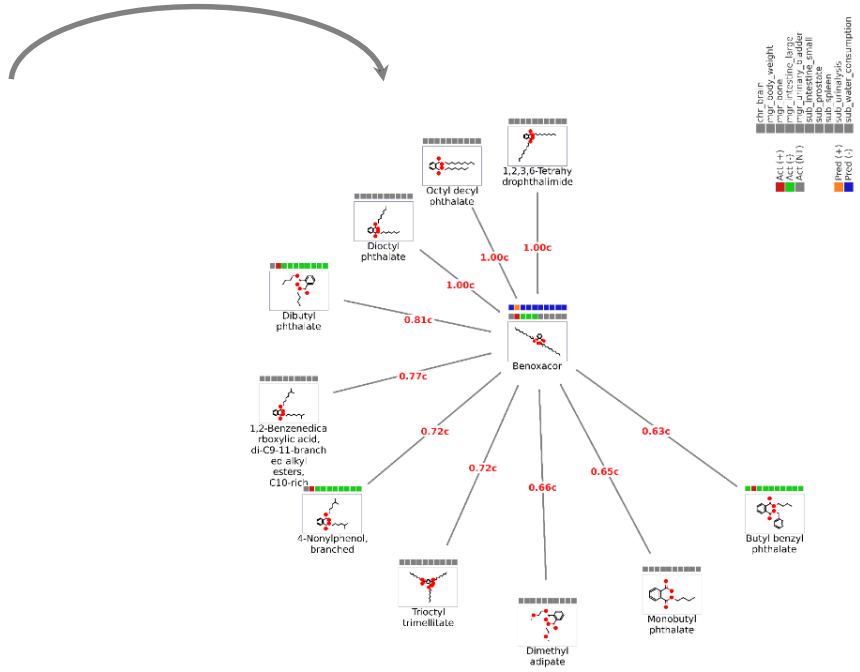
Jason Harris

Generalised Read-across (GenRA)

Chemicals are clustered on the basis of chemical descriptors to identify local neighbourhoods

The Read-across toxicity prediction is a similarity-weighted activity of nearest neighbours based on chemistry and bioactivity descriptors

Uncertainties can be evaluated across the local neighbourhoods



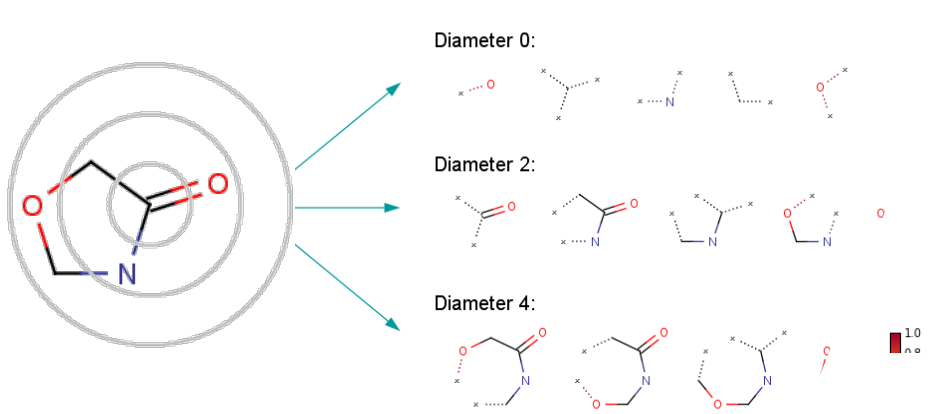
Chem/Bio Similarity

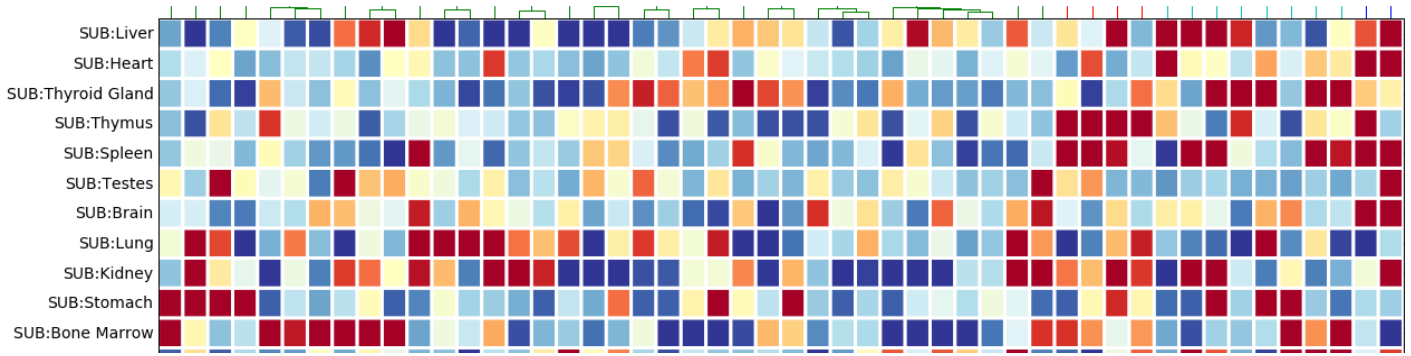
Number of Analogs

Chem/Bio Similarity

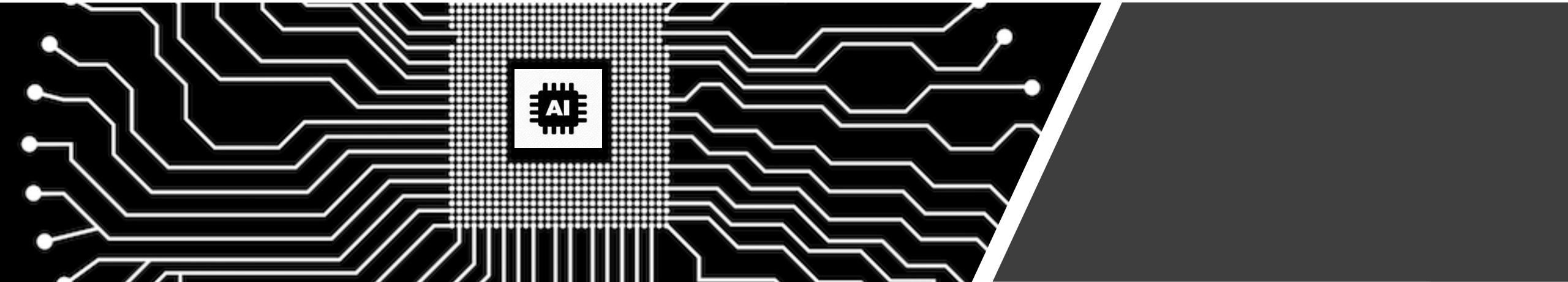
Number of Analogs

Shah, et al. RTP 2016





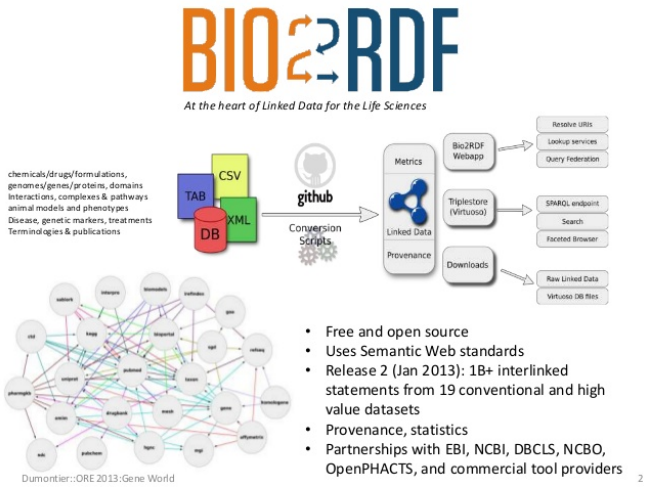
Machine Learning to Predict Chemical Effects



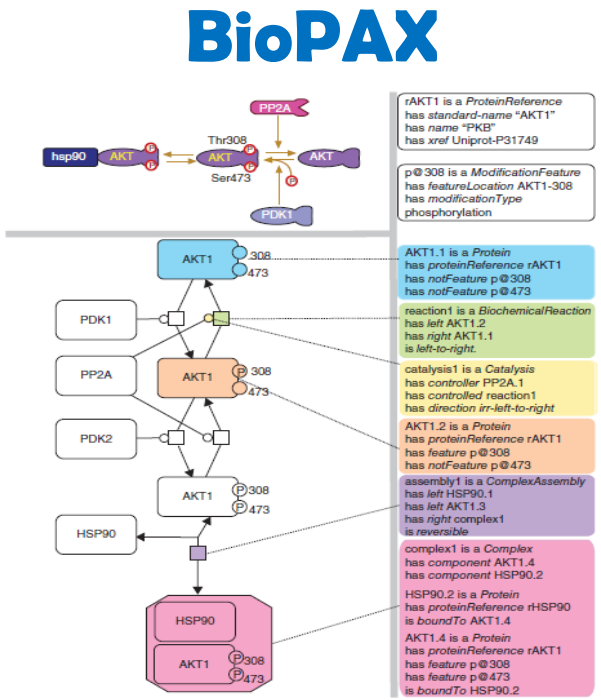
Use Semantic Tools to Link Disparate Data

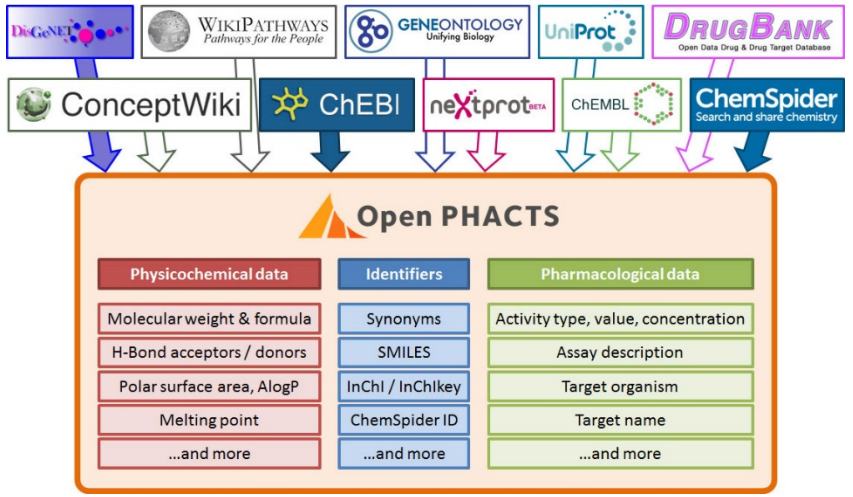
Chemical → Uses → Exposure → ADME → Internal Dose ↔ Molecular Changes ↔ Cellular Pathways ↔ Organ Effects ↔ Toxicity → Population Outcomes

Can we use semantic tools (OWL/RDF, existing ontologies to meaningfully integrated disparate resources ?



- Free and open source
- Uses Semantic Web standards
- Release 2 (Jan 2013): 1B+ interlinked statements from 19 conventional and high value datasets
- Provenance, statistics
- Partnerships with EBI, NCBI, DBCLS, NCBO, OpenPHACTS, and commercial tool providers





Data Challenges

- Transparently sharing complex data streams – adequately capturing chemical (treatment dose and time), biological (experimental modal, assay, etc.) context to ease re-use
- Systematically integrating disparate data streams – representing linkages across molecular, cellular, tissue, organs. This is vital for relating early molecular changes to adverse (e.g. histopathological) outcomes
- Effectively extracting evidence from unstructured textual data – the literature is one of the largest resource for information about apical outcomes
- Using linked data to better discriminate between adaptation vs adversity – predicting which molecular markers lead to apical outcomes
- Quantifying and incorporating uncertainty and variability in predictions
- Legal defensibility of new methods and assessment products

Acknowledgements and Questions

Tox21 Colleagues:
- NTP Crew
- FDA Collaborators
- NCATS Collaborators

EPA Colleagues:
- NERL
- NHEERL
- NCEA

EPA's National Center for Computational Toxicology