EPA
United States
Environmental Protection
Agency

Overview and Update on EPA's ToxCast Chemical Prioritization Project

NICEATM-ICCVAM 5YPISC and RDWG, June 22, 2009

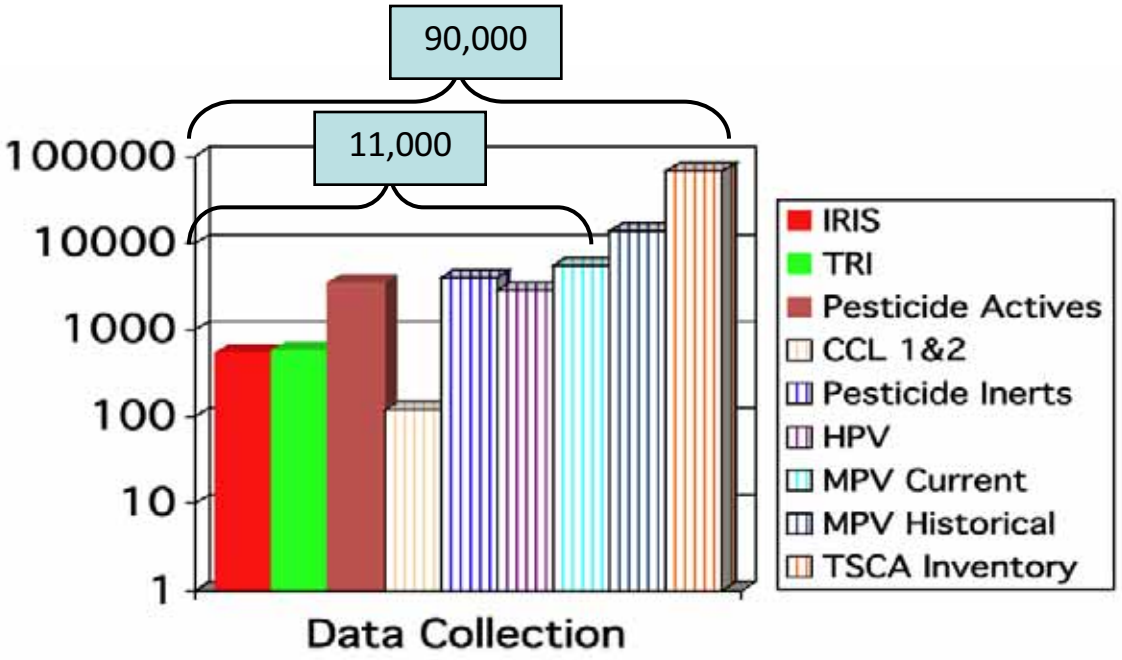
Keith Houck
houck.keith@epa.gov

Office of Research and Development
National Center for Computational Toxicology

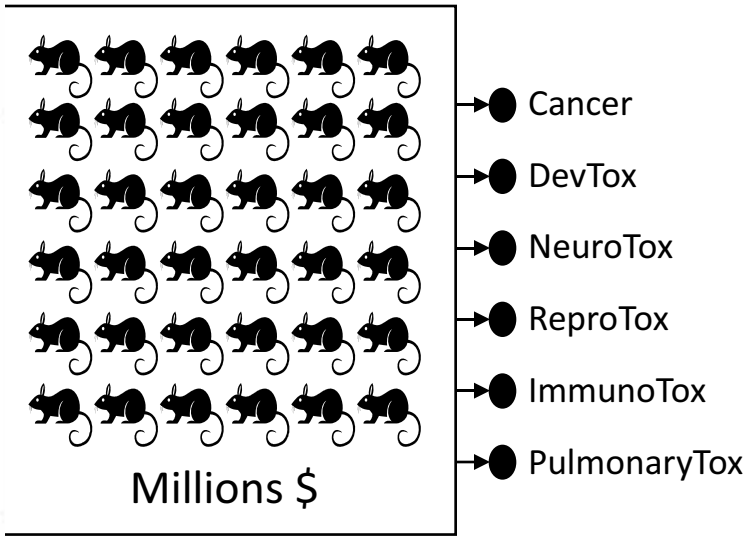
EPA United States Environmental Protection Agency

Change Needed Because …..

Too Many Chemicals



Too High a Cost



…and not enough data.

Judson, et al *EHP, 2008*

National Academy of Sciences Report (2007)
Toxicity Testing in the Twenty-first Century: A Vision and a Strategy

NAS PANEL SEEKS MAJOR SHIFT IN HOW EPA ASSESSES CHEMICALS' TOXICITY

Date: June 22, 2007 -

A National Academy of Sciences (NAS) panel is calling for a major shift in how EPA assesses chemicals' toxicity, recommending that the agency base its toxicological research and regulatory processes on how substances affect biological pathways -- which send information within and between cells -- rather than so-called health endpoints, such as cancer.

POLICYFORUM

Science: Feb 15, 2008

TOXICOLOGY

Transforming Environmental Health Protection

Francis S. Collins,[1*†] George M. Gray,[2*] John R. Bucher[3*]

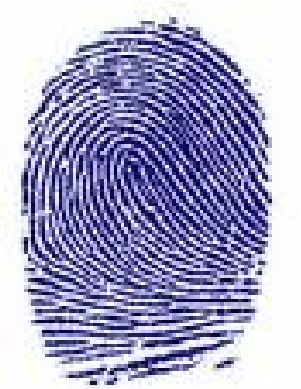
We propose a shift from primarily in vivo animal studies to in vitro assays, in vivo assays with lower organisms, and computational modeling for toxicity assessments.

In 2005, the U.S. Environmental Protection Agency (EPA), with support from the U.S. throughput screening (HTS) and other automated screening assays into its testing tion, usually between 2 and 10 μM, and tolerate high false-negative rates. In contrast, in

EPA United States Environmental Protection Agency

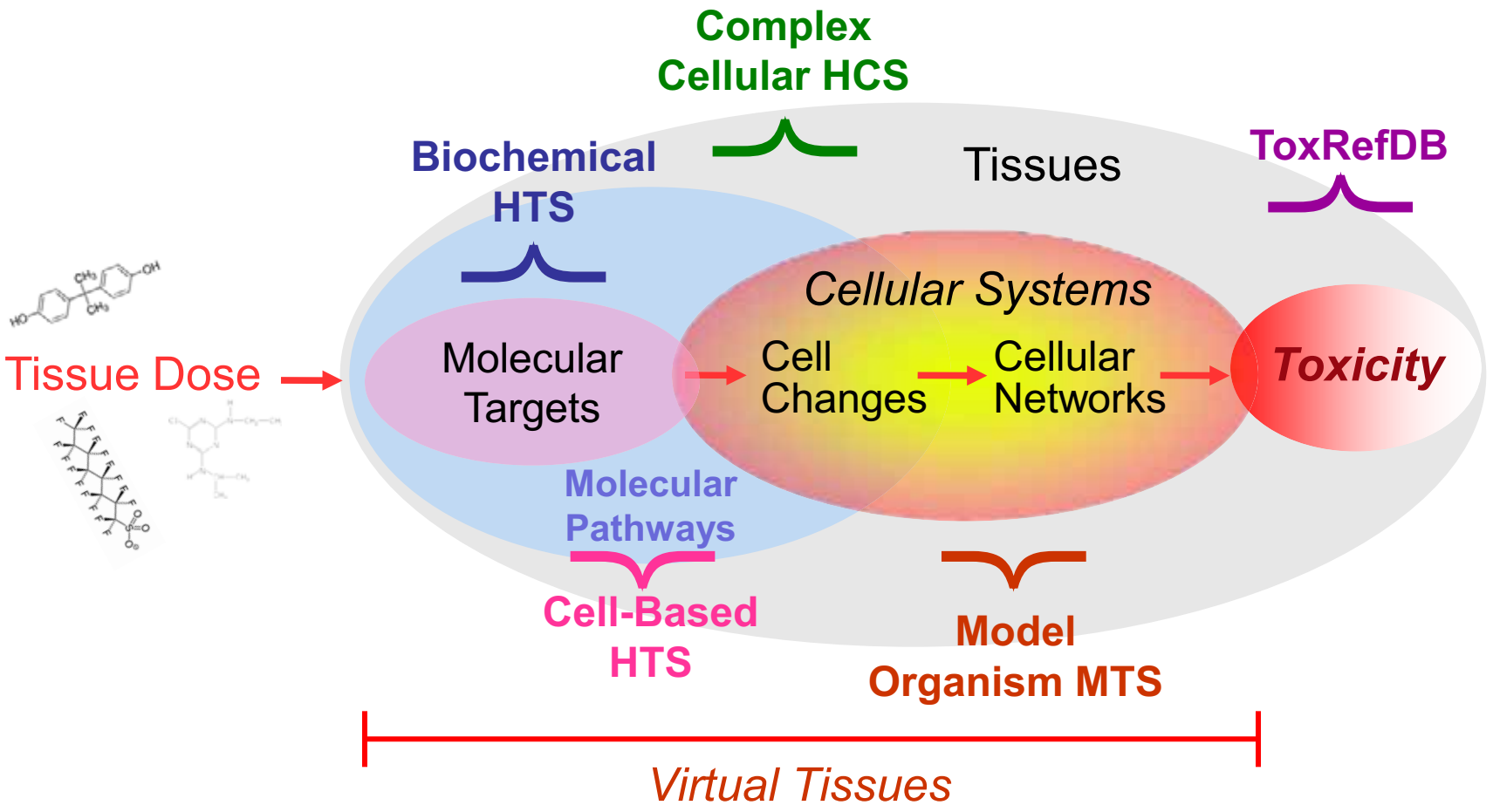
ToxCast Background

- Research program of EPA's National Center for Computational Toxicology (NCCT)
- Addresses chemical **screening** and **prioritization** needs for pesticidal inerts, anti-microbials, CCLs, HPVs and MPVs
- Comprehensive use of HTS technologies to generate biological fingerprints and predictive signatures
- Coordinated with NTP and NHGRI/NCGC via Tox21
- Committed to stakeholder involvement and public release of data
  - Communities of Practice- Chemical Prioritization; Exposure
  - NCCT website http://www.epa.gov/ncct/toxcast
    - *ACToR* *http://www.epa.gov/actor/*
    - *ToxRef DB* *http://www.epa.gov/ncct/toxrefdb/*
    - *DSSTox (PubChem)* *http://www.epa.gov/ncct/dsstox/*

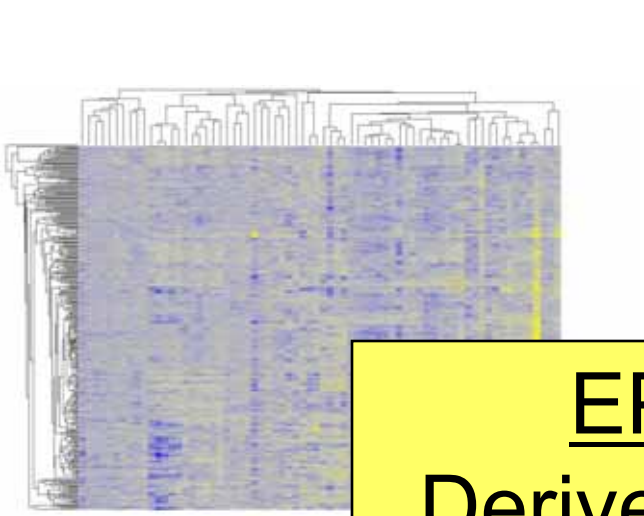


EPA United States Environmental Protection Agency

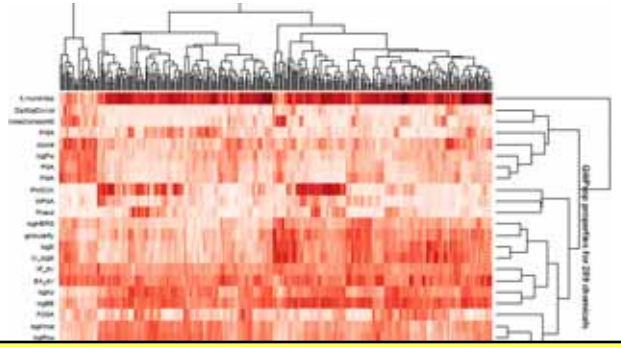
Ultimate Goal of ToxCast: Predicting Human Toxicity

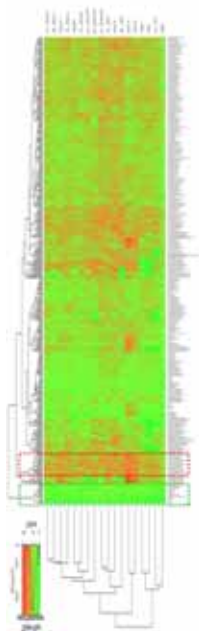


Correlating Domain Outputs



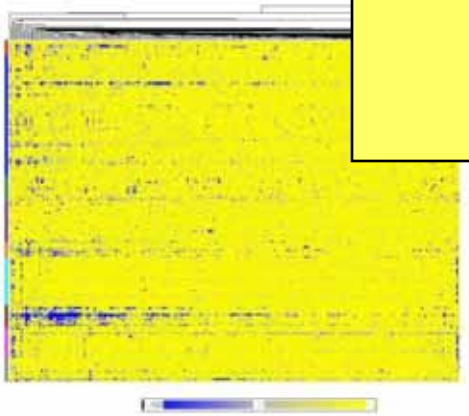
Cellular Assays



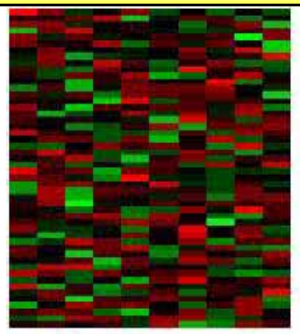


silico Predictions

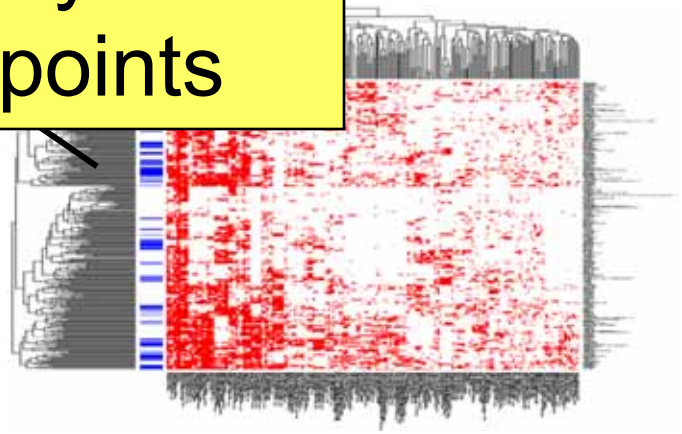
EPA ToxCast Goal:
Derive "Signatures" from *in vitro* & *in silico* assays to predict *in vivo* endpoints



Biochemical Assays



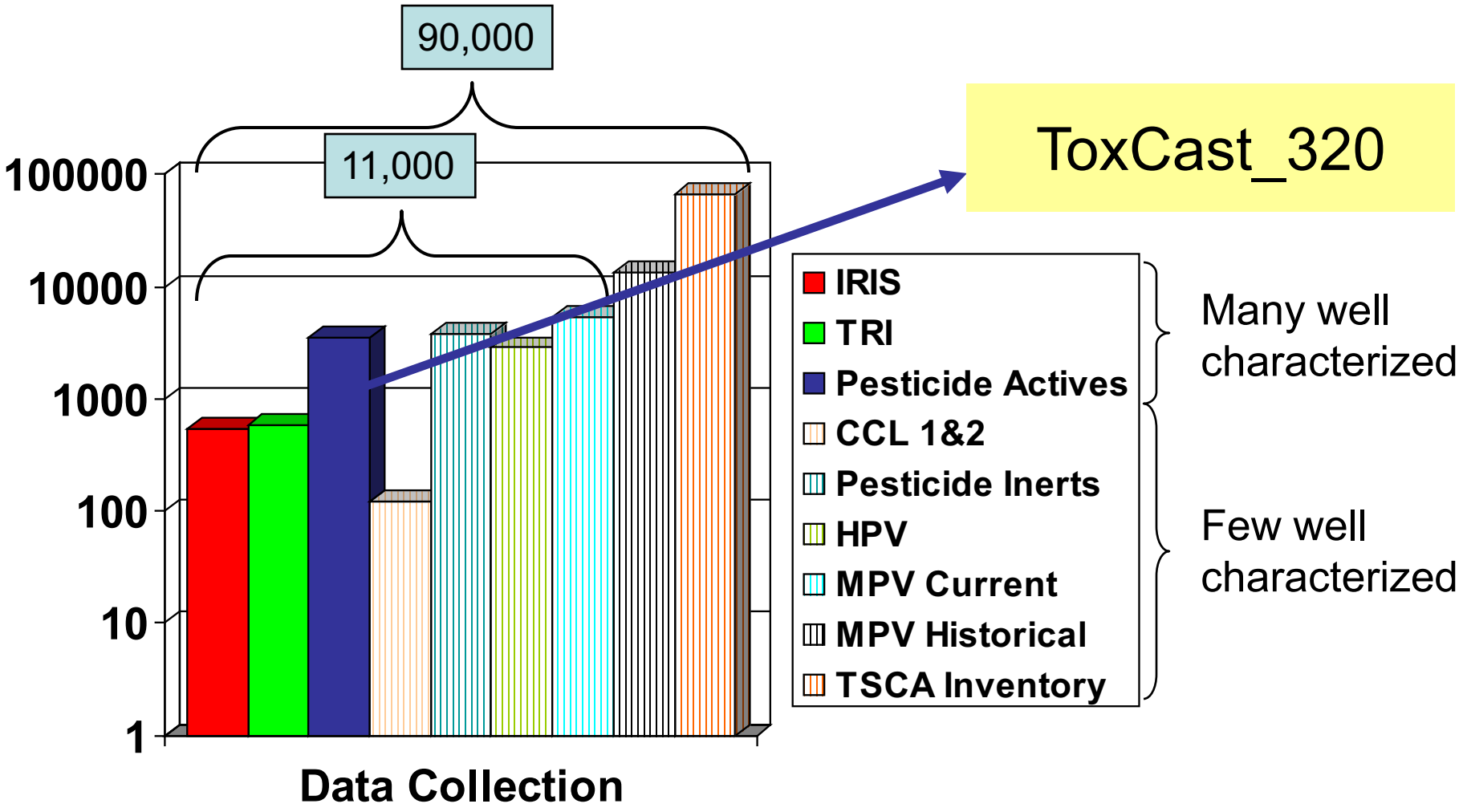
Genomic Signatures



Toxicology Endpoints

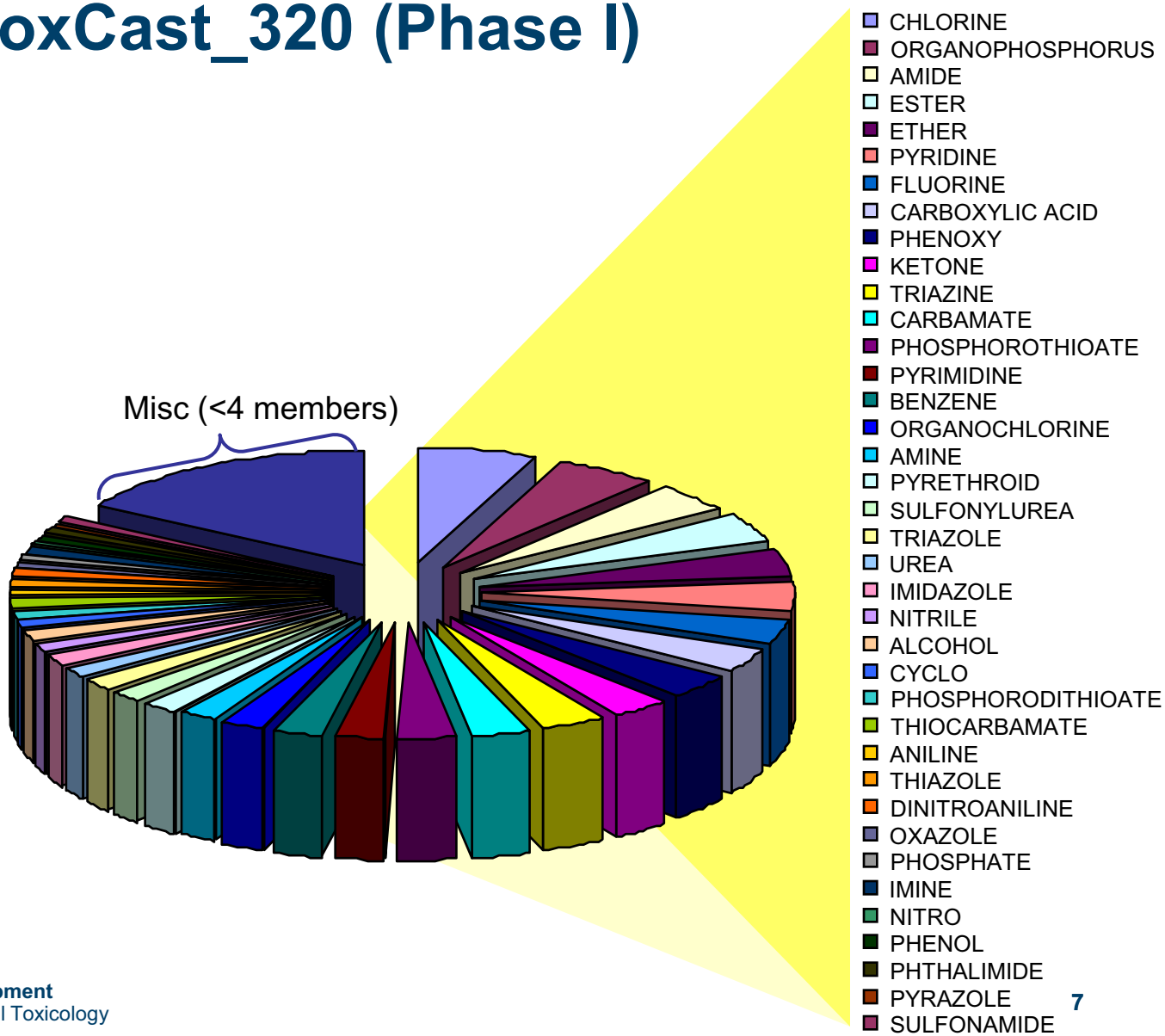
EPA United States Environmental Protection Agency

ToxCast Phase I Chemicals



Chemical Classes in ToxCast_320 (Phase I)

- 309 Unique Structures
- Replicates for QC
- 291 Pesticide Actives
- 9 Industrial Chemicals
- 13 Parent/Metablolite pairs
- 56/73 Proposed Tier 1 Endocrine Disruption Screening Program
- 14 High Production Volume Chemicals
- 11 HPV Challenge



EPA Pesticide Programs: Data Evaluation Records (DERs)

- Used for hazard identification and characterization
- Study Types
  - Chronic
  - Cancer
  - Subchronic
  - Multigeneration
  - Developmental
  - Others: DNT, Neurotox, Immu
- Derive Endpoints (NOAEL/LOAEL)
  - Systemic
  - Parental
  - Offspring
  - Reproductive
  - Maternal
  - Developmental
- Critical Effects for Endpoints

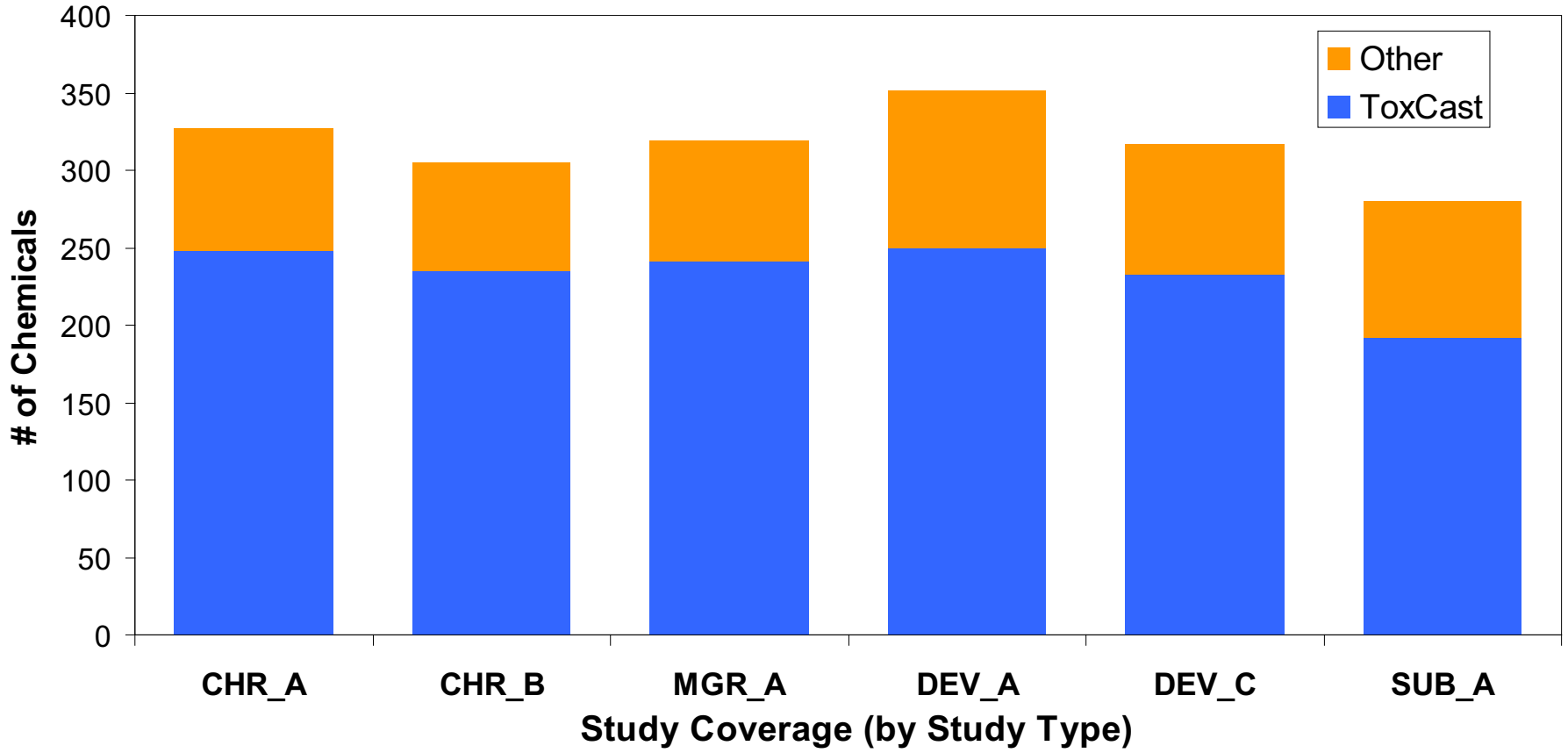
$10,000,000

DER Format

- Study Identifiers
  - Tested Chemical Information
    - IDs
    - Name
    - Purity
  - Study Type IDs
  - Reviewer Information
- Citation(s)
- Executive Summary
  - Summary Study Design
  - Summary Effects
  - Endpoints (NOAEL/LOAEL)
- Chemical Properties
- Animal Information
  - Species
  - Strain
  - Husbandry
- Results (full dose-response)
  - Clinical signs
  - Body weight
  - Clinical Chemistry/ Hematology
  - Gross Pathology
  - Non-neoplastic Pathology
  - Neoplastic Pathology
  - Parental vs. Offspring
  - Maternal vs. Fetal

United States Environmental Protection Agency

2073 Studies Entered For 480 Chemicals

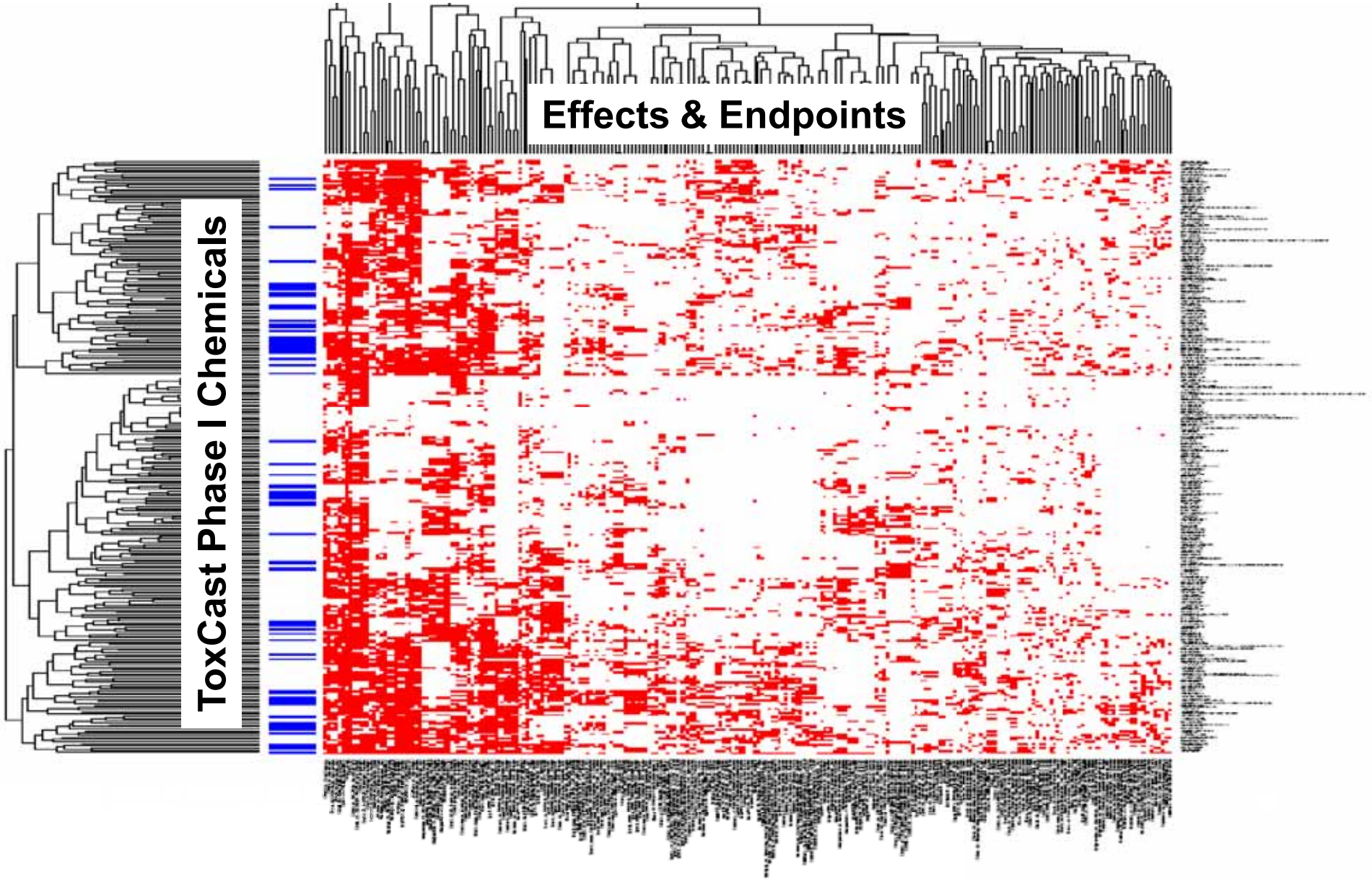


CHR = Chronic/Cancer
MGR = Multigeneration Reproductive
DEV = Prenatal Developmental

A = Rat
B = Mouse
C = Rabbit

EPA
United States
Environmental Protection
Agency

>$1Billion Million Dollars Worth of *In Vivo* Chronic/Cancer Bioassay Effects and Endpoints



EPA United States Environmental Protection Agency

ToxCast Data Sources

Compound Focus, Inc. a subsidiary of BioFocus DPI A Galápagos Company

CELLUMEN

attagene The Home of TFomics™

CellzDirect™ invitrogen corporation

NOVASCREEN a Caliper Life Sciences company

BioSeek

L'ORÉAL SOLIDUS

ACEA Biosciences, Inc.

gentronix

NCGC

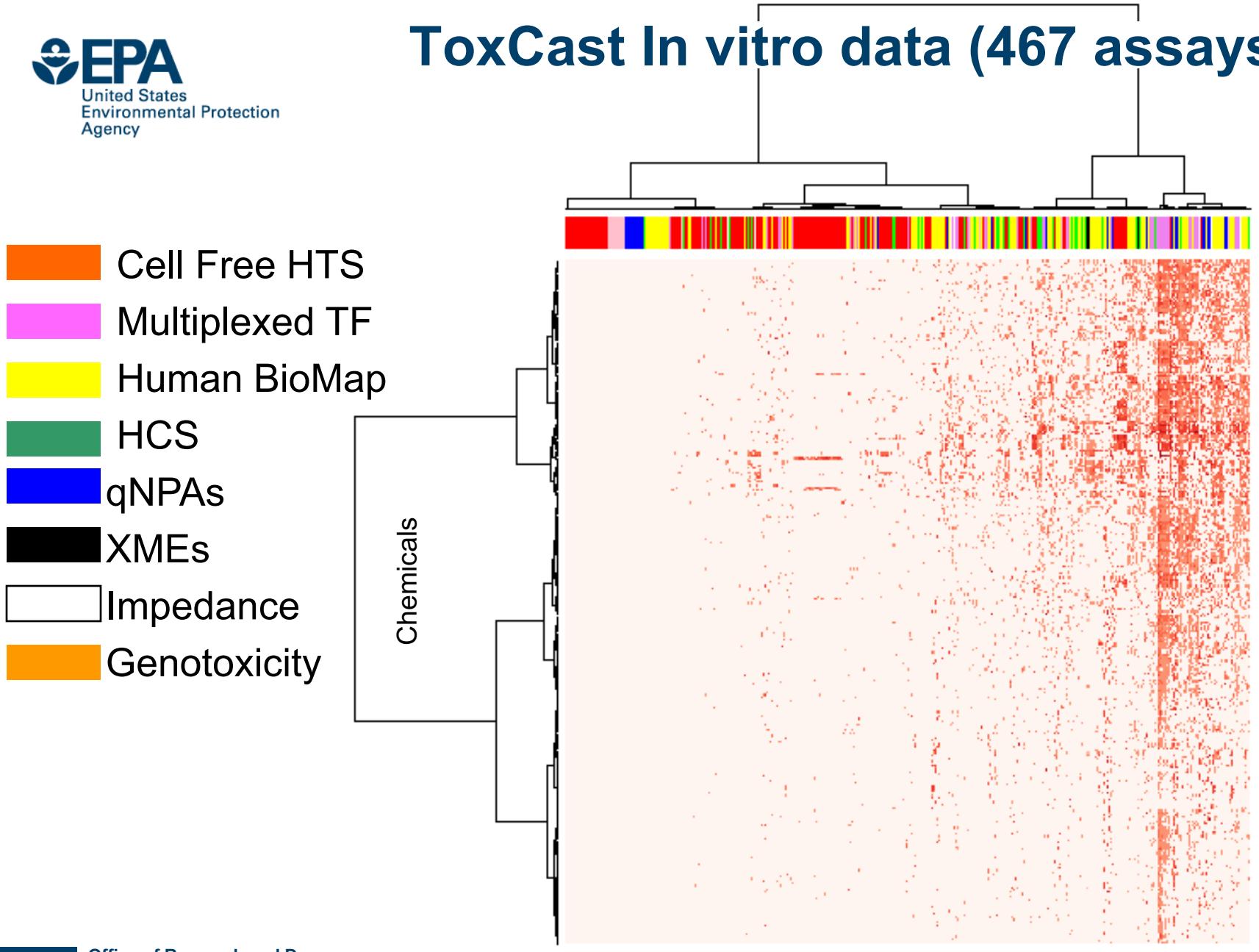
Tox21

National Toxicology Program U.S. Department of Health and Human Services

6 contracts, 4 collaborations
467 assays, 534 endpoints

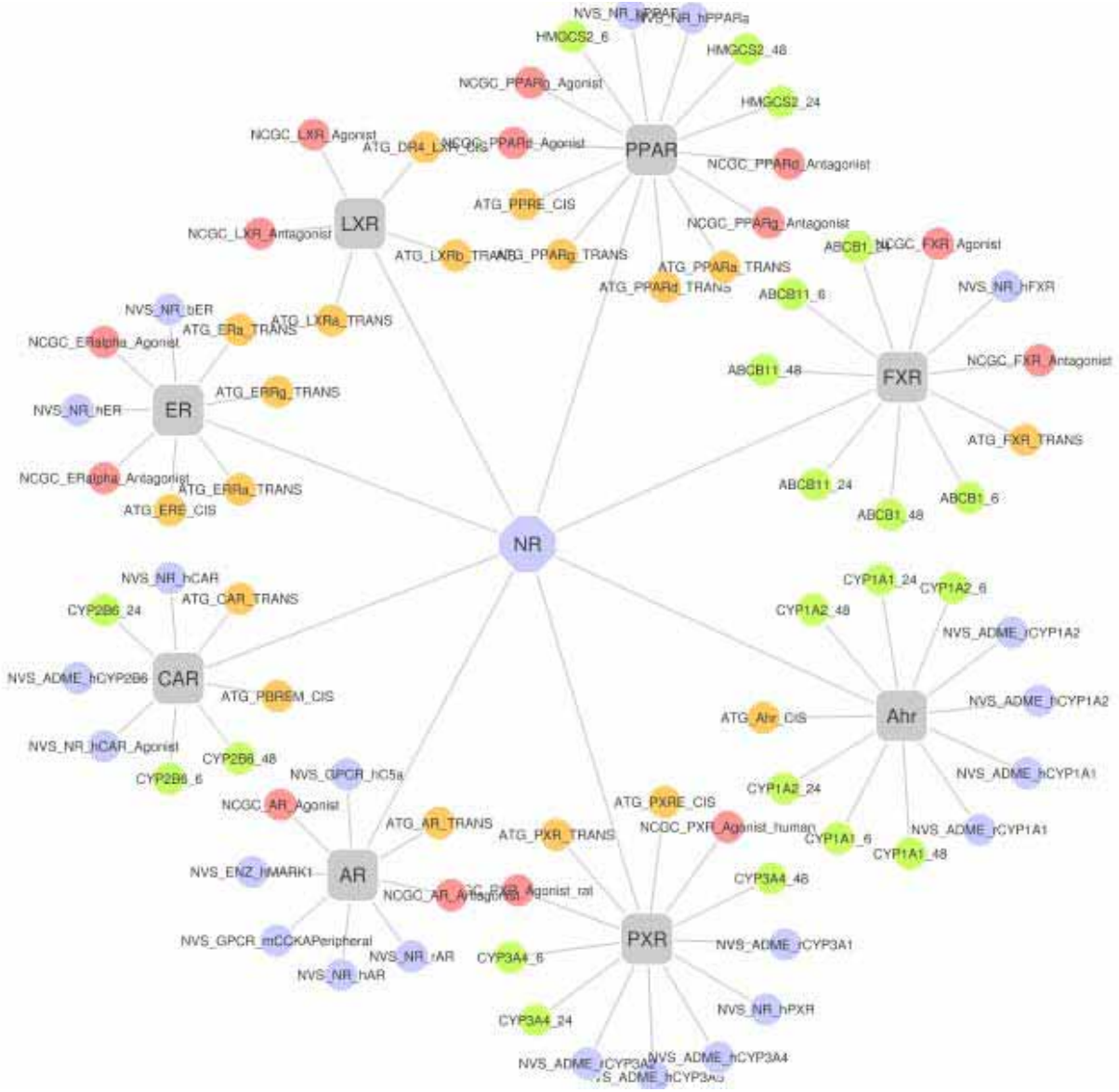
EPA United States Environmental Protection Agency

ToxCast In vitro data (467 assays)

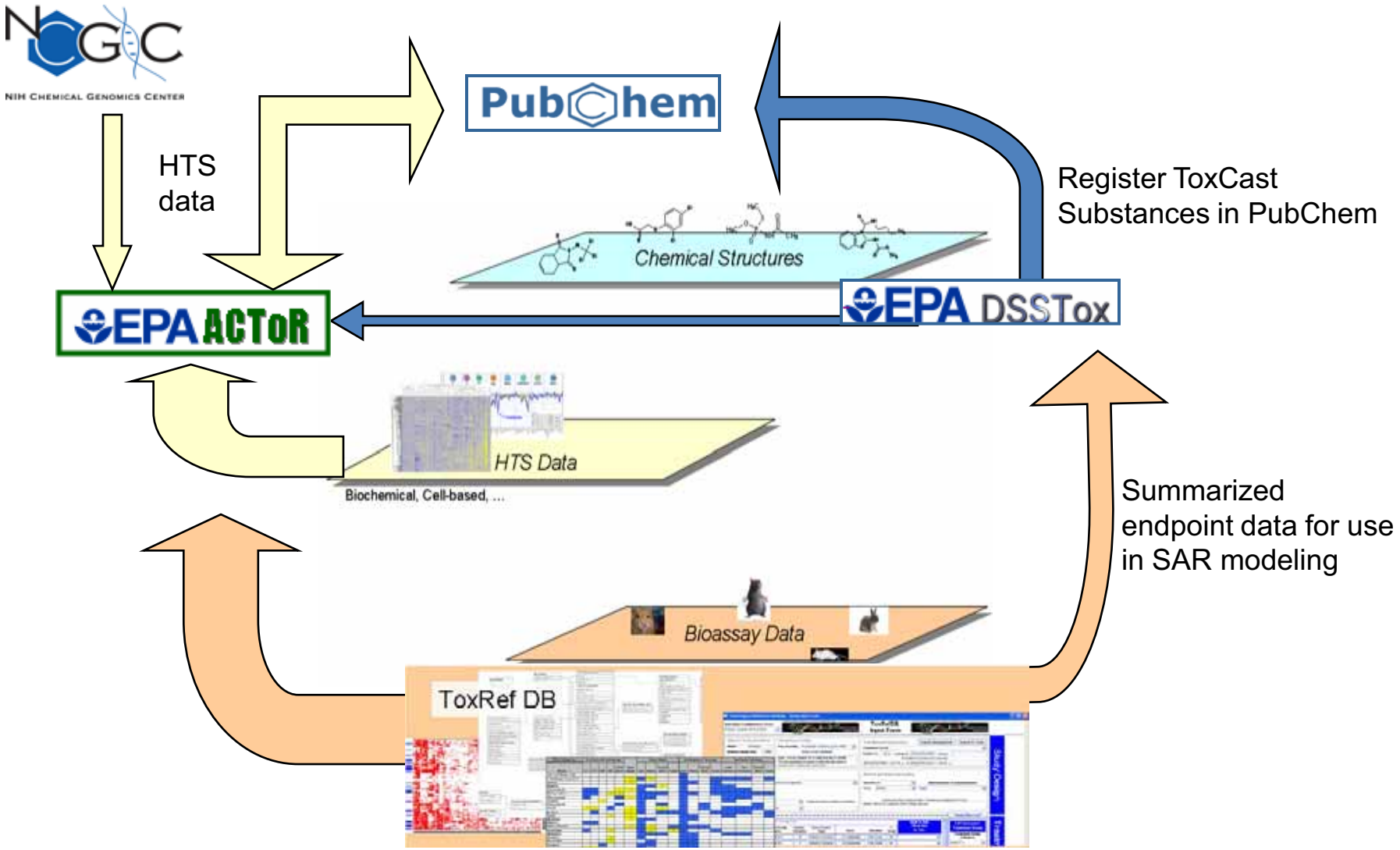


EPA United States Environmental Protection Agency

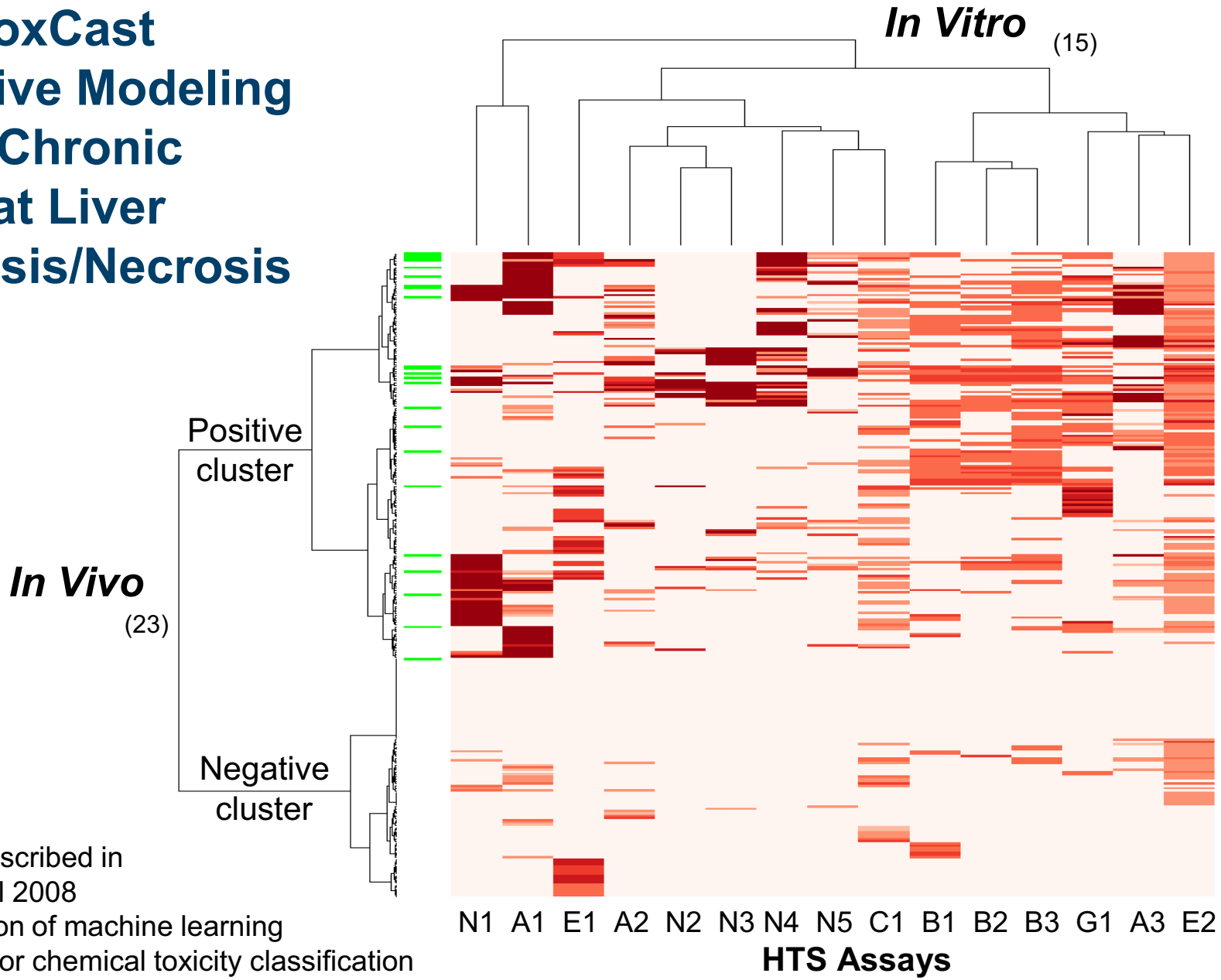
Multiple Assays per Endpoint



ToxCast: Data Publication & Exploration



ToxCast Predictive Modeling of Chronic Rat Liver Apoptosis/Necrosis



Methods described in
Judson et al 2008
A comparison of machine learning algorithms for chemical toxicity classification using a simulated multi-scale data model.
BMC Bioinformatics 9:241

ToxCast In Vitro/In Vivo Correlation Examples

Calculate Univariate Associations with Rat Liver Proliferative Lesions

- Significance Tests:
  - T-test (treat *in vitro* as continuous)
  - Chi-squared (treat *in vitro* as dichotomous, using 100μM as the cutoff)
- Significant associations are:
  - PPARA
  - PPARG
  - HMGCS2 (regulated by PPAR)
  - RXRA (dimerizes with PPAR)
  - CCL2
  - CCL26

PPAR signaling and Rodent Liver Tumors

- PPAR is involved with lipid and fatty acid metabolism
- Xenobiotics can activate PPAR
  - Leads to peroxisome proliferation and hepatocyte hypertrophy

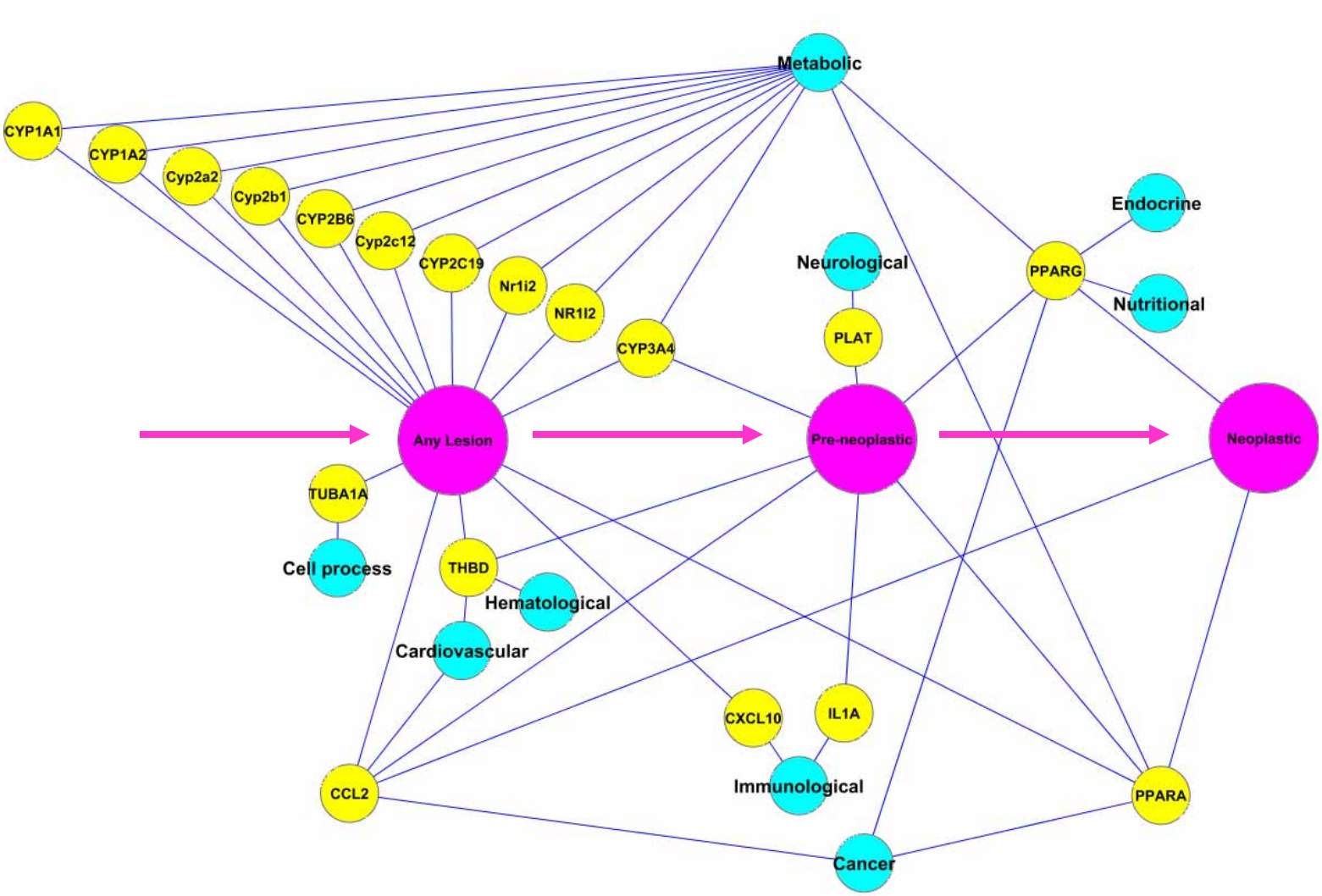
- PPAR-driven liver tumorigenesis does not seem to act in humans
  - But PPAR-driven hepatotoxicity is of concern (FDA)
  - PPAR is a target for human drugs to treat metabolic syndrome / diabetes

- 3 isoforms
  - PPARA / PPARα
  - PPARG / PPARγ
  - PPARD / PPARδ

CCL2 Associations with Environmental Chemicals and Liver Toxicity are Novel

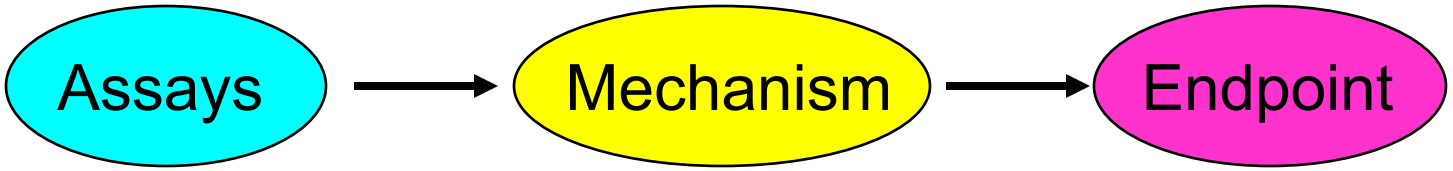
- Chemokine (C-C motif) ligand 2
- Drives angiogenesis and tumor cell invasion
- Seen in both humans and rodents
- Increased CCl2 levels associated with
  - Human Prostate cancer severity and progression
  - Human Gastric carcinomas
  - Human Oral carcinomas
  - Human Breast cancer
  - Human Thyroid cancer
  - Rat cholestatic liver injury
- May be related to PPAR signaling

Rat Liver Disease Progression Links



Links Drawn for Univariate Associations with $p<0.01$

Toxicity Signature Definition

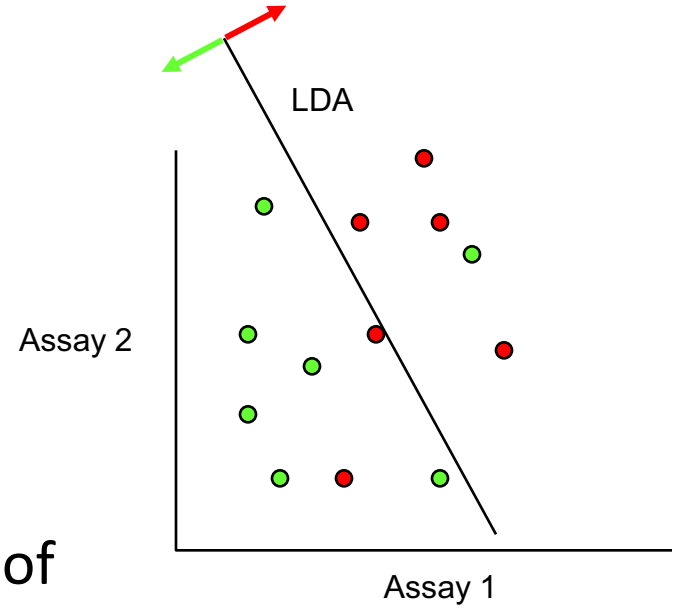


- An algorithm that takes as its input
  - A chemical
  - One or more *in vitro* assay measurement or *in silico* parameters
- And returns
  - A classification for that chemical for a toxicity endpoint

- Other terms
  - Model
  - Classifier

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*(assay values)
  - If
    - *Predictor* for a chemical meets criteria
  - Then
    - Predict endpoint to be positive for the chemical



	+ Truth	-
+ Test	TP	FP
-	FN	TN

Machine Learning Process

- ML Methods used
  - SVM – Support Vector Machines
  - NNET – Neural Networks
  - LDA – Linear Discriminant Analysis
  - SLR – Stepwise Logistic Regression

} Seemed to consistently overfit
Consistent with unbalanced data set

- Use AC50/LEC Data and log transform
- T-test Feature Selection
  - $p<0.1$ for cutoff
  - Accept maximum of n(chemical)/10 feature
- Use 5-fold cross validation
- Evaluate performance using balanced accuracy (BA)
  - BA=average of sensitivity and specificity

SLR Signature:
Rat Liver Proliferative Lesions

Assay	Coefficient	Gene	Gene Name
Intercept	-2.86		
ATG_PPARg_TRANS	0.298	PPARG	peroxisome proliferator-activated receptor gamma
NVS_ADME_hCYP3A4	0.614	CYP3A4	cytochrome P450, family 3, subfamily A, polypeptide 4
CLM_OxidativeStress_24hr	0.403	H2AFX	H2A histone family, member X (oxidative stress)
BSK_SM3C_MCP1_up	0.331	CCL2	chemokine (C-C motif) ligand 2
BSK_BE3C_IL1a_down	0.389	IL1A	interleukin 1, alpha
ATG_RORg_TRANS	0.51	RORC	RAR-related orphan receptor C
BSK_BE3C_tPA_up	0.386	PLAT	plasminogen activator, tissue
CLM_Hepat_Steatosis_24hr	0.181		
ATG_PPARa_TRANS	0.254	PPARA	peroxisome proliferator-activated receptor alpha
CLM_MitoticArrest_24hr	-0.322		
CLM_p53Act_72hr	0.28	TP53	tumor protein p53
ATG_Sp1_CIS	0.195	SP1	Sp1 transcription factor
ATG_NRF2_ARE_CIS	-0.171	NFE2L2	nuclear factor (erythroid-derived 2)-like 2 (oxidative stress)

Start with 624 Assay measurements, 3 p-chem, 103 chemical structure class variables
Genes associated with tumors or liver disease in red

Signature Performance – Proliferative Lesions

In vivo data

Signature	+	-
+	31	11
-	30	176

Sensitivity=51%
Specificity=94%

- 248/309 chemicals had rat data in ToxRefDB (used for model building)
- 8 other chemicals were predicted to be positive
  - PFOA: Causes rat liver adenomas
  - PFOS: Causes rat liver adenomas
  - Diniconazole: rat liver hypertrophy
  - Chlorothalonil: rat liver enlargement, kidney tumors
  - TCMTB: testicular and thyroid adenomas
  - No data for Niclosamide, Methylene bis(thiocyanate), Phenoxyethanol

Examine False Positives

- Look for data outside of ToxRefDB for highest scoring false positives

- Fenpyroximate
  - Liver hypertrophy in a rat 90-day subchronic study
- Bromoxynil
  - Non-proliferative lesions (2 year rat study)
  - Liver adenomas (2 year mouse study)
- Cyproconazole
  - Hepatocellular adenomas and carcinomas in mice
- Tribufos
  - Liver hemangiosarcomas in male mice

ToxCast™ Data Analysis Summit, May 14-15, 2009

- Phase 1 ToxCast data made available to analysis partners prior to full public release
  - >500 HTS assays – categorical (1/0)
  - 76 “bioassay” endpoints from ToxRefDB for modeling
  - Chemical structure SD file (DSSTox), chemical information files (descriptors)
- Over 200 registered attendees, 60 presenters
- Wide variety of prediction schemes
  - *In vitro* → *In vivo*
  - Chemical descriptors → *In vivo* (SAR)
  - Chemical descriptors + *In vitro* → *In vivo*
- Wide variety of approaches
  - Statistics, clustering, machine learning, particle swarm, etc.

ToxCast™ Data Analysis Summit, May 14-15, 2009

Impressions, Conclusions, Lessons…

- ToxCast Phase I data set poses highly challenging problems for prediction methods
- Global associations (*in vitro* to *in vivo*) trends not readily apparent → must go local to see meaningful associations
- Statistical means for dealing with highly dimensional, sparse, unbalanced data needed → new methods proposed
- Use of chemical descriptors and features improve model performance when combined with HTS (Is this accounting for ADME??)
- Public data availability and transparency successful in engaging wide range of researchers and capabilities in early analysis
- PASS, LAZAR, ToxTree indicate limited applicability of prior SAR carcinogenicity prediction models (based on public data) to ToxCast Phase I chemical space → reinforces need to enrich public data space, improve models

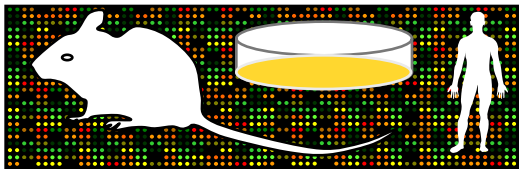
EPA United States Environmental Protection Agency

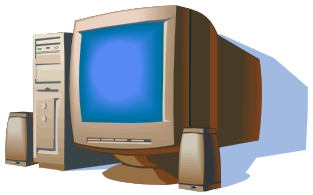
ToxCast Development

Phase	Number of Chemicals	Chemical Criteria	Purpose	Number of Assays	Cost per Chemical	Target Date
Ia	320	Data Rich (pesticides)	Signature Development	>500	$20k	FY07-08
Ib	15	Nanomaterials	Pilot	166	$10K	FY09
IIa	>300	Data Rich Chemicals	Validation	>400	~$20-25k	FY09
IIb	>100	Known Human Toxicants	Extrapolation	>400	~$20-25k	FY09
IIc	>300	Expanded Structure and Use Diversity	Extension	>400	~$20-25k	FY10
IId	>12	Nanomaterials	PMN	>200	~$15-20K	FY09-10
III	Thousands	Data poor	Prediction and Prioritization	>300	~$15-20k	FY11-12

January 2009

Tox21 Collaboration





National Health and Environmental Effects

National Center for Computational Toxicology

- Combined HTS plates (6x1408) high interest chemicals
- Joint assay development
- Use of NCGC HTS testing capabilities
- EPA informatics (ACToR/DSSTox)

National Toxicology Program
Department of Health and Human Services

NIH Chemical Genomics Center

Biomolecular Screening Branch

Toxicology Project Team