Outcomes of the ToxCast data analysis summit: Challenges in transforming toxicity testing from *in vitro* to *in vivo*
EPA’s Need for Prioritization

Too Many Chemicals

Too Little Data (%)

The Future of Toxicity Testing

Bioinformatics/Machine Learning in silico analysis

Cancer
ReproTox
DevTox
NeuroTox
PulmonaryTox
ImmunoTox

Future of Toxicity Testing

TOKOCYLOGY
Transforming Environmental Health Protection

Office of Research and Development
National Center for Computational Toxicology

www.epa.gov/ncct/toxcast
ToxCast™ Background

- 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 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- Aggregated Computational Toxicology Resource
    http://www.epa.gov/actor/
Key Challenges Of Pathway Profiling

- Find the Toxicity Pathways
  - Hepato vs developmental neurotoxicity

- Obtain HTS Assays for Them
  - Including metabolic capability

- Screen Chemical Libraries
  - Coverage of p-chem properties

- Link Results to in vivo Effects
  - Gold standard and dosimetry
309 Unique Structures
Replicates for QC
291 Pesticide Actives
9 Industrial Chemicals
8 Metabolites
56/73 Proposed Tier 1 EDSP
53 of 80 with DNTs
122 in IRIS
14 HPV
11 HPV Challenge

Classification based on OPPIN

56/73 Proposed Tier 1 EDSP
53 of 80 with DNTs
122 in IRIS
14 HPV
11 HPV Challenge
ToxRefDB

- Relational phenotypic/toxicity database
- Provides in vivo anchor for ToxCast predictions

Three study types
- Chronic/Cancer rat and mouse (Martin, et al, EHP 2008)
- Rat & Rabbit developmental (Knudsen, et al, Repro Tox, in press)

Two types of synthesis
- Supervised (common individual phenotypes)
- Unsupervised (machine based clustering of phenotype patterns)
ToxCast In Vivo Data from ToxRefDB

Chronic/Cancer
Multigenation
Developmental
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)
  – Gentronix - GreenScreen GeneTox assay (1 assay)
  – NCGC - nuclear receptor assays (11 assays)
  – Novascreen / Caliper - receptor binding and enzyme inhibition assays (239 assays)
  – Solidus - P450 vs. cytotoxicity assays (4 assays)

467 Endpoints

• 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 (Hepregen)
  – 3D Cellular Zebrafish vascular/cardiotoxicity (Zygogen)
  – HTS stress response (NHEERL+NCGC)
  – Embryonic Stem Cells (NHEERL)
  – Metabolic Phenotyping (Biolog)

+ New contract proposals under review
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

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
Multiple Assays per Endpoint
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
  - \( \text{Predictor} = F(\text{assay values}) \)
  - If
    - \( \text{Predictor} \) for a chemical meets criteria
  - Then
    - Predict endpoint to be positive for the chemical
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
Rat Liver Disease Progression Links

Links Drawn for Univariate Associations with p<0.01
Focus on Rodent Liver Toxicity

- Several liver lesion classes for rat, similar for mouse
  - All chemicals: 248
  - No Liver lesions: 122
  - Any Lesion: 126
  - Pre-neoplastic or neoplastic: 58
  - Neoplastic: 21
  - Liver Proliferative Lesions: 61
  - Liver Tumors (=neoplastic): 21
Calculate Univariate Associations with Rat Liver Proliferative Lesions

• Significance Tests:
  – T-test (treat \textit{in vitro} as continuous)
  – Chi-squared (treat \textit{in vitro} as dichotomous, using 100\textmu M as the cutoff)

• Significant associations are:
  – PPARA
  – PPARG
  – HMGCS2 (regulated by PPAR)
  – RXRA (dimerizes with PPAR)
  – CCL2
  – CCL26
Signature Performance – Proliferative Lesions

- 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

<table>
<thead>
<tr>
<th>Signature</th>
<th>+</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>31</td>
<td>11</td>
</tr>
<tr>
<td>-</td>
<td>30</td>
<td>176</td>
</tr>
</tbody>
</table>

Sensitivity=51%
Specificity=94%
Presentations explored multiple aspects of the ToxCast Phase-I data

- Jim Jones  Using ToxCast™ for Chemical Screening and Prioritization in the Real World (EPA/OPPTS)
- David Dix  Overview of the ToxCast™ Research Program: Applications to Predictive Toxicology and Chemical Prioritization
- Keith Houck  Characteristics of the ToxCast™ In Vitro Datasets from Biochemical and Cellular Assays
- Matt Martin  Characteristics of the ToxRefDB In Vivo Datasets from Chronic, Reproductive and Developmental Assays
- Richard Judson  NCCT Predictive Signatures from ToxCast™
- Barry Hardy  Initial OpenTox Evaluation of ToxCast™ Phase 1 Datasets
- Alex Tropsha  Prediction of animal toxicity endpoints of ToxCast™ Phase I compounds using a combination of chemical and biological in vitro descriptors
- Lyle Burgoon  Biomarker Identification using Graph Theoretic and Particle Swarm Optimization-based Support Vector Machine Analysis of the Phase I ToxCast™ Dataset
- William Welsh  Biological profile analysis of the ToxCast™ chemicals: Linking biological activity profiles to molecular structure
- Rusty Thomas  An Integrated In Vitro and Computational Approach to Define the Exposure-Dose-Toxicity Relationships In High-Throughput Screens
- Weida Tong  Prediction of liver toxicity in the animal study using the mechanistically relevant in vitro screening assay data – Lessons learned from the EPA ToxCast™ project
- Fred Wright  Prediction of in vivo toxicity endpoints from ToxCast™ Phase I data using a variety of machine learning approaches
- Alison Motsinger  Predictive Modeling of Toxicity Outcomes with Grammatical Evolution Neural Networks
- James Rathman  Navigating through chemical and biological data domains of ToxCast™
- Christodoulos Floudas  Predicting in vivo toxicities using optimal methods for re-ordering and machine learning
- Nina Jeliazkova  Hierarchical Multi-label Classification of ToxCast™ Datasets
- Jun Huan  Learning from Multiple Data Sources and Its Applications in Chemical Toxicity Prediction
- Jessie Xia  Hierarchical Predictive Analysis of Chemical Toxicity By Recursive Partitioning Using the Phase-I ToxCast™ Dataset
- Johann Gasteiger  Chemical Reactivity in Metabolism and Degradation Reactions in the Risk Assessment Workflow
- Matt Martin  Biotransformation and ToxCast™
- Holly Mortensen  Mapping Human Toxicity and Disease Pathways in ToxCast™

(Plus over 40 poster presentations!)
Opportunities and challenges in the ToxCast Phase-I data

Signature utility:

Prediction versus/and/or prioritization?
Is the burden of proof greater for validation of extant hypotheses versus the generation of novel hypotheses?

Technical issues:

• What are the best measures of performance?
• How robust are the signatures?
• Is data and mechanism heterogeneity an issue?
• How can data be compared across assays?
• Which data are most useful?
• Where is more data needed?
Recurrent themes & insights from this meeting

- Data “meaning” (in vitro assays, in vivo endpoints, chemical descriptors)
- Heterogeneity (chemicals, endpoint pathways, . . . .)
- Non-linearity
- Pathways and putative mechanisms
- Class imbalance
- Method Implementation is not analysis
- Interpretation across sets (and types) of data
- Structure (formal interrelationships) in the data (in vitro assays, chemical descriptors, . . . .)
- Global models versus cluster-specific models
- Incorporation of external knowledge
- Endpoint aggregation and/or progression
- Endpoint prioritization (prevalence, exposure, species extrapolation, severity, . . . .)
- Look to other scientific disciplines for valuable methods/lessons/strategies
- Metrics for success
- Interpret results in context (chemistry ←→ toxicology ←→ biology ←→ . . . .)
- Important data gaps
- Expanding scale → exploding combinatorics
Toxicity is Multi-factorial
ToxCast™ Program
Predicting Hazard, Characterizing Toxicity Pathways, and Prioritizing the Toxicity Testing of Environmental Chemicals

ToxCast™ Data Analysis Summit
Transforming Toxicity Testing From In Vivo to In Vitro: A Computational Toxicology Challenge

The First ToxCast™ Data Analysis Summit
Hosted by U.S. EPA’s National Center for Computational Toxicology
EPA Campus, Research Triangle Park NC
May 14-15, 2009

The EPA ToxCast™ Program is developing novel approaches to predict chemical toxicity using high-throughput and high-content in vitro assays. ToxCast™ has produced data on 320 chemicals from over 500 in vitro assays, and about 75 in vivo endpoints, providing a powerful dataset for evaluating analysis approaches.

The EPA made the ToxCast™ data available to analysis partners who signed a Materials Transfer Agreement (MTA). We invited our partners to apply their best analysis strategies, and present their findings at the first ToxCast™ Data Analysis Summit (TDAS). The goal is to find robust analysis methods for predicting whole animal or human chemical toxicity from the in vitro data.

The meeting will be held in the Building C auditorium, room C111, on EPA’s campus in Research Triangle Park, NC.

AGENDA
(All times are Eastern Daylight Savings Time (EDT))

Below are presentations from the conference that can be downloaded. Please note many of the files below are PDF documents.

<table>
<thead>
<tr>
<th>Day 1: Thursday, May 14</th>
<th>EPA RTP campus building C auditorium (Room C111)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:30am-8:40</td>
<td>Robert Kavlock</td>
</tr>
<tr>
<td></td>
<td>Welcoming Remarks and Goals of the Summit (EPA/ORD/NCCT)</td>
</tr>
<tr>
<td>8:40am-9:00</td>
<td>Jim Jones</td>
</tr>
<tr>
<td></td>
<td>Using ToxCast™ for Chemical Screening and Prioritization in the Real World (EPA/OPPTS)</td>
</tr>
<tr>
<td>9:00am-10:10am</td>
<td></td>
</tr>
<tr>
<td>9:00-9:15</td>
<td>David Dix</td>
</tr>
<tr>
<td></td>
<td>Overview of the ToxCast™ Research Program: Applications to Predictive Toxicology and Chemical Prioritization (EPA/ORD/NCCT)</td>
</tr>
<tr>
<td>9:15-9:45</td>
<td>Keith Houdk</td>
</tr>
<tr>
<td></td>
<td>Characteristics of the ToxCast™ In Vitro Datasets from Biochemical and Cellular Assays (EPA/ORD/NCCT)</td>
</tr>
<tr>
<td>9:45-10:00</td>
<td>Matt Martin</td>
</tr>
<tr>
<td></td>
<td>Characteristics of the ToxReIDB In Vitro Datasets from Chronic, Reproductive and Developmental Assays (EPA/ORD/NCCT)</td>
</tr>
<tr>
<td>10:00-10:30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BREAK--LIGHT REFRESHMENTS PROVIDED</td>
</tr>
</tbody>
</table>