Tox21 & ToxCast Programs for using *in vitro* HTS data to predict *in vivo* outcomes: *A Chemical Perspective*

*Ann M. Richard*
*U.S. EPA, National Center for Computational Toxicology*

*Cosmetics Inventory Review Meeting, Wash DC*
*June 11, 2012*
Overview

- EPA’s ToxCast Program
- Navigating data domains
- MOA QSAR
- Expanding ToxCast’s chemical landscape
- Tox21 Project
- New cheminformatics approaches
- Tox21-COSMOS structure overlaps
ToxCast Project

2007: EPA NCCT launches ToxCast
2008: Phase I testing begins
2009: Phase I data results
2010: Full Phase I data publication/release
2012: Phase II testing begins
2013: Phase II & e1k data publication/release
2012: ToxRefDB web access w/in ACToR

Phase I (309)
Phase II (1060)
+ Phase II & e1k (1860)
Problem Statement

Too many chemicals to test with standard animal-based methods
- Cost, time, animal welfare

Need for better mechanistic data
- What is human relevance
- What is the Mode of Action (MOA)?
- What is the Adverse Outcome Pathway (AOP)?
ToxCast Goals

• Identify targets or pathways linked to toxicity
  – Chemicals perturbing these can lead to adverse events

• Develop assays for these targets or pathways
  – Assays probe “Molecular Initiating Events” or “Key Events” [MIE / KE]

• Develop predictive models: *in vitro* → *in vivo*
  – “Toxicity Signature”

• Use signatures:
  – Prioritize chemicals for targeted testing (“Too Many Chemicals” problem)
  – Suggest / distinguish possible AOP / MOA for chemicals
Toxicity Signature Generation

In Vitro Data

Predictive Models – “Signatures”

In Vivo Data

R. Judson, SOT Presentation, March 2012
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
  - Rat hepatocytes
  - Mouse embryonic stem cells (Sid Hunter)

- Biotransformation competent cells
  - Primary rat hepatocytes
  - Primary human hepatocytes

- Assay formats
  - Cytotoxicity
  - Reporter gene
  - Gene expression
  - Biomarker production
  - High-content imaging for cellular phenotype

Primarily Human / Rodent
Exception: Zebrafish development (S. Padilla)
ToxCast Phase I: Published Models

- Predictive models: endpoints
  - liver tumors: Judson et al. 2010, Env Hlth Persp 118: 485-492
  - zebrafish vs ToxRefDB: Sipes et al. 2011, Birth Defects Res C 93: 256-267

- Predictive models: pathways
  - endocrine disruption: Reif et al. 2010, Env Hlth Persp 118: 1714-1720
  - angiogenesis: Kleinstreuer et al. 2011, Env Hlth Persp 119: 1596-1603
Developmental Rat Toxicity Model Features

22 Assays Across 12 Features
Balanced Accuracy:
Training: 71%
Test: 70%

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 issues: don’t have all the right assays (FN)
  - Tissue issues: may need multi-tissue signaling networks (FN)
  - Statistical power issues: need enough chemicals acting through a given MOA to be able to build and test model (FN)
  - Compensation: system may adapt to initial insult (FP)
  - *In vitro* assays are not perfect! (FP, FN)
  - *In vivo* rodent data are not perfect! (FP, FN)
ToxCast Data Landscapes

Structures
(Q) SAR

In Vitro/HTS

Biological Modeling

In Vivo
Rat Liver Tumorigens: diverse chemical structures and in vitro “signatures”
Structure Class vs Bioactivity Class

Chemical structure class:
- Cluster according to structure similarity
- Range of activities

Bioactivity profile class:
- Implies mechanistic similarity
- Diverse chemical structures
Toxicity Prediction Challenge

How do we make best use of all data domains to extract meaningful relationships from these mechanistically diverse data?
Navigating a data-rich landscape

P. Volarath et al., Fall ACS Mtg, Boston 2010

Developmental Toxicity:
- Rat
- Morphology
- External Malformation
- Cleft Palate

Structures

ToxRefDB

In Vitro/HTS

In Vivo

Existing knowledge
Target of Modeling: *In Vivo* Endpoint

**Developmental Toxicity:**
- *Rat*
  - *Morphology*
  - *External Malformation*
  - *Cleft Palate*

**Literature support for biological pathways implicated in Cleft Palate:**

- Transforming growth Factor-β (TGF-β) family
- Other growth factors:
  - *fibroblast growth factor (FGF)*,
  - *epidermal growth factor (EGF)* or *TGFα*,
  - *insulin-like growth factor (IGF)*,
  - *platelet-derived growth factor (PDGF)*
- Hedgehog signaling pathway (SHH) & Wnt signaling pathways

**Biological targets**

*In vitro* assays, e.g. TGF-β

---

ToxRefDB: Rat Cleft Palate Actives

<table>
<thead>
<tr>
<th>Chemical Structure 1</th>
<th>Chemical Structure 2</th>
<th>Chemical Structure 3</th>
<th>Chemical Structure 4</th>
<th>Chemical Structure 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Chemical Structure 1" /></td>
<td><img src="image2.png" alt="Chemical Structure 2" /></td>
<td><img src="image3.png" alt="Chemical Structure 3" /></td>
<td><img src="image4.png" alt="Chemical Structure 4" /></td>
<td><img src="image5.png" alt="Chemical Structure 5" /></td>
</tr>
</tbody>
</table>

**10 Actives**

**240 Inactives**

*A total of 250 ToxCast PhaseI_309 chemicals have ToxRefDB Rat DevelopmentalTox studies*
Rat Cleft Palate Associated Assays: Univariate Analysis of ToxCast_309

Univariate Analysis: Which ToxCast assays are statistically associated with cleft palate endpoint?

Assays appear to relate to known mechanisms of cleft palate activity?

Fisher’s Exact Test: p value < 0.05
Rat Cleft Palate Associated Assays: Literature Search

Literature Search: *Which ToxCast assays relate to known pathways for cleft palate?*

- TGF pathways
  - BE3C_TGFβ1_down
  - BE3C_TGFβ1_up
  - KF3CT_TGFβ1_down
  - KF3CT_TGFβ1_down

Are these assays statistically associated with cleft palate in the ToxCast data set?
# ToxCast_309: HTS Assay Associations with Rat Cleft Palate Activity

## Literature-based association
- **TGFP_CIS**: Odds Ratio 3.7, p values > 0.05
- **BE3C_TGFβ1_up**: Odds Ratio 0.0
- **KF3CT_TGFβ1_down**: Odds Ratio 4.8
- **KF3CT_TGFβ1_up**: Odds Ratio 0.0
- **BE3C_TGFβ1_down**: Odds Ratio 4.0

## In vivo endpoint
- **DEV_Rat_Pregnancy_EmbryoFetalLoss**: Odds Ratio 5.5

## Statistical-based associations - Pathways
- **Ceramide_signaling**: Odds Ratio 10.1, p values < 0.05
- **KEGG_T_cell_receptor_signaling**: Odds Ratio 10.1
- **Gene_CYP2d2_Rattus_norvegicus**: Odds Ratio 11.0
- **KEGG_mTOR_signaling**: Odds Ratio 7.3
- **IGF_1_signaling**: Odds Ratio 6.5

## Statistical-based association - Assays
- **ADME_rCYP2D2**: Odds Ratio 11.0
- **IC_rNaCh_site2**: Odds Ratio 20.1
- **ADME_rCYP2A2**: Odds Ratio 8.2
- **HIF1a_CIS**: Odds Ratio 5.3
- **ABCG2_48**: Odds Ratio 8.2
Structure-Feature Approach

Developmental Toxicity:
- Rat
- Morphology
- External Malformation
  - Cleft Palate

In Vitro/HTS
In Vivo
Existing knowledge
ToxRefDB: Rat Cleft Palate Actives

<table>
<thead>
<tr>
<th>10 Actives</th>
<th>240 Inactives</th>
</tr>
</thead>
</table>

**Chemical Structures:**

1. [Structure 1]
2. [Structure 2]
3. [Structure 3]
4. [Structure 4]
5. [Structure 5]
6. [Structure 6]
7. [Structure 7]
8. [Structure 8]
9. [Structure 9]
10. [Structure 10]

**Chemical Structures with Green Circles:**

1. [Structure 11]
2. [Structure 12]
3. [Structure 13]
4. [Structure 14]
5. [Structure 15]
6. [Structure 16]
7. [Structure 17]
8. [Structure 18]
9. [Structure 19]
10. [Structure 20]

**Chemical Structure with Red Circle:**

- **Formula:** $f_1$
- **Explanation:** $A = \text{any atom}$, $X = \text{any halide}$
Sterol Biosynthesis Inhibitors (SBI Fungicides)

Chemicals containing the target feature, f1, belong to this triazoles MOA group

G1: SBI class I: DMI fungicides
►C14 demethylase (erg11/cyp51)
# 3 DeMethylation Inhibitors (=DMI fungicides)

http://www.frac.info
Chemicals containing the target feature in the 309 dataset

Cleft Palate Actives
4 out of 10

Cleft Palate Inactives:
5 out of 240

Odds Ratio: 30:1
### Assays Hits within f1 Feature Class

<table>
<thead>
<tr>
<th>Assay Name</th>
<th>Cyproconazole</th>
<th>Flusilazole</th>
<th>Propiconazole</th>
<th>Triadimefon</th>
<th>Hexaconazole</th>
<th>Myclobutanil</th>
<th>Paclitaxel</th>
<th>Tetraconazole</th>
<th>Triadimenoil</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEV_rat_CleftLipPalate_endpoint</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Target Feature</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>DEV_rat_PregnancyRelated_Embryo_FetalLoss_endpoint</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TGFβ_CIS</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BE3C_TGFβ1_up</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>KF3CT_TGFβ1_down</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>KF3CT_TGFβ1_up</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BE3C_TGFβ1_down</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ceramide_signaling</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>KEGG_T_cell_receptor_signaling</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Gene_CYP2d2_Rattus_norvegicus</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>KEGG_mTOR_signaling</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>IGF_1_signaling</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ADME_rCYP2D2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>IC_rNaCh_site2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ADME_rCYP2A2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>HIF1a_CIS</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ABCG2_48</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

**Evaluate new chemical for Cleft Palate Activity**

- **Does chemical contain f1 feature?**
  - Yes (9) No (241)

- **Is TGFβ1_down assay positive?**
  - Yes (2) No (7) 100%

- **Assay + f1 feature together can strengthen association**
- **TGFβ1 assay (literature) has greater significance within f1 feature class**
Remove feature (f1) class: How do assay statistics compare?

Significance of TGFβ assays increases when f1 feature class removed

With the f1 class (9 chemicals) removed, most assay associations are significantly weakened in the remaining 241 set (6 Actives, 235 Inactives)

Significance of CYP2D2, CYP2A2 assays could point to particular metabolic requirements for Rat Cleft Palate actives containing f1 feature
Developmental training set

- Tox data (1154 chemicals) merged from public sources*
  - EPA ToxRefDb
  - FDA CDER (drugs@fda)
  - FDA CFSAN (food contact substance)
  - ILSI DevTox
  - NTP

<table>
<thead>
<tr>
<th></th>
<th>FDA</th>
<th>NTP</th>
<th>ILSI</th>
<th>ToxRef</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDA</td>
<td>399</td>
<td>10</td>
<td>24</td>
<td>12</td>
</tr>
<tr>
<td>NTP</td>
<td>86</td>
<td>21</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>ILSI</td>
<td>195</td>
<td></td>
<td></td>
<td>18</td>
</tr>
<tr>
<td>ToxRef</td>
<td></td>
<td></td>
<td></td>
<td>384</td>
</tr>
</tbody>
</table>

- Dose-level effects
  - Maternal toxicity
  - Fetal survival and growth
  - Fetal morphology
    - e.g., cleft palate, microthalmia, kidney/renal/ureter anomalies


Slide courtesy of Chihae Yang, Altamira
Possible Mode-of-Actions for cleft palate

**Chemical classes**
- Glucocorticoid
- Steroid
- SHh antagonist
  - Cyclopamin
- Sterol biosynthesis inhibitors
  - triazine
  - imidazole
  - pyrimidine ...
- Retinoids
- Phthalates...

**Biological key events during embryogenesis**
- Wnt signaling
- SHh signaling
- Glucocorticoid receptors
- Retinoic acid receptors
- P-glycoprotein
- Glycogen synthase kinase
- Nuclear receptor
- Androgen receptor

**Phenotype effects**
- disruptions in EMT
  - cell proliferation
  - palate growth
  - Fusion

**EMT**: epithelial – mesenchymal transformation

*Slide courtesy of Chihae Yang, Altamira*
## Cleft palate chemotypes

<table>
<thead>
<tr>
<th>Steroids, glucocorticoids, cyclopamine</th>
<th>![Steroid Molecule] ![Cyclopamine Molecule]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triazole, imidazole, pyridines, pyrimidine, piperazole, dioxolane…</td>
<td>![Triazole Molecule]</td>
</tr>
<tr>
<td>Retinoids, conjugated dienes</td>
<td>![Retinoid Molecule] ![Conjugated Dienes Molecule]</td>
</tr>
<tr>
<td>Phthalic ester</td>
<td>![Phthalic Ester Molecule]</td>
</tr>
<tr>
<td>Organofluorides</td>
<td>![Organofluorides Molecule]</td>
</tr>
<tr>
<td>Thiocarbamates</td>
<td>![Thiocarbamates Molecule]</td>
</tr>
</tbody>
</table>

*Slide courtesy of Chihae Yang, Altamira*
MoAs: features for biology and chemistry

Biological features

Chemical features

In vitro Assays

In vivo Toxicity

MoA

QSAR21

Slide courtesy of Chihae Yang, Altamira
• QSAR 21
  ➢ Bring chemistry to Tox21 and biology to QSAR

➢ Successfully applied to:
  o developmental effects
  o cleft palate (rat); eye, kidney, ureter dysmorphology
  o genotoxicity
  o skin sensitization, skin irritation

*Slide courtesy of Chihae Yang, Altamira*
Chemical library serves to probe in vitro biology

- need sufficient diversity of structures to sample wide range of target interactions, pathways, and MOAs
- need sufficient representation of structures & features to enable QSAR inferences
ToxCast_1060 (Phases I & II)

**ToxCast Phase I (293 unique cmpds)**
- *EPA pesticidal actives w/ rich in vivo data*
- *PFOAs, BPA, metabolite/parent pairs*

**ToxCast Phase II (767 unique cmpds)**
- *EPA pesticides, high interest EPA and stakeholder inventories, data rich chemicals (EDSP, OPPT, Antimicrobials, Inerts, …)*
- *FDA CFSAN data rich, NCTR LTKB Priority 1 drugs*
- *Toxicity reference chemicals, data-rich chemicals, NTP immunotox compounds*

- *135 Donated pharma cmpds -- failed drugs w/ pre-clinical or clinical tox data*
- *Donated green chemicals; L’Oreal sponsored compounds*
Tox21: EPA Chemical Inventories

To be tested in endocrine-related subset of ToxCast assays (approx 60): ER, AR agonist & antagonist assays

Pesticidal actives, failed drugs, antimicrobials, CFSAN cmpds, green alternatives, HPV, MPV, endocrine ref cmpds, other tox ref cmpds, NTP in vivo, EPA high interest compounds
e1k Assays and Protocols

- **ACEA**
  - Cell growth assay in T47D cells; 2
  - 4-28; 8 half log dilutions; in duplicate

- **Attagene**
  - Transcription factor profiling (cis and trans), in HEPG2
  - MTC; 8 concentrations; singlet

- **Novascreen (Caliper)**
  - Biochemical assays for NR binding; single conc (25 uM); in duplicate
  - Eight point concentration follow-up on hits
  - hAR, rAR, hER, mERα, bER, hTRα, rMR, hGR, hPR, bPR, hPPARα, hPPARγ, hPBR, rPBR, (+/-)hRAR, hPXR, hCYP19A1

- **Odyssey Thera**
  - Protein Complementation or High Content Imaging in HEK293, PC3 or HeLa cells
  - 6 concentrations, times vary (1.5-24 hr); in triplicate
  - ERα & β, ERRγ, AR, GPR30, GPR78, NURR1, PPARγ
  - ERα & β and AR will be run with +/- rat S9

*Slide courtesy of R. Judson, NCCT*
### Tox21 Chemicals: EPA Selection Strategy

- Nominations & tracking of inventory overlaps by CAS
- Prioritize for procurement
- Cost & availability
- Filter by MW, physchem, volatility, MSDS cautions
- DMSO solubility

<table>
<thead>
<tr>
<th>CAS Number</th>
<th>Chemical Name</th>
<th>MW</th>
<th>Volatility</th>
<th>MSDS Cautions</th>
<th>DMSO Solubility</th>
</tr>
</thead>
<tbody>
<tr>
<td>12345</td>
<td>Acetone</td>
<td>57</td>
<td>High</td>
<td>None</td>
<td>Yes</td>
</tr>
<tr>
<td>67890</td>
<td>Ethanol</td>
<td>46</td>
<td>Medium</td>
<td>None</td>
<td>No</td>
</tr>
<tr>
<td>23456</td>
<td>Water</td>
<td>18</td>
<td>Low</td>
<td>None</td>
<td>Yes</td>
</tr>
<tr>
<td>78901</td>
<td>Isopropanol</td>
<td>88</td>
<td>Low</td>
<td>None</td>
<td>No</td>
</tr>
</tbody>
</table>
ToxCast PhI&PhII 1060: # Compounds per Inventory

- **AIR**: 90
- **GRAS**: 26
- **EDSP**: 130
- **IRIS**: 240
- **TRI**: 216
- **MPV**: 83
- **HPV**: 232
- **Green Chemistry**: 85
- **Antimicrobials**: 91
- **Consumer**: 210
- **Water**: 217
- **Pesticide Inerts**: 243
- **Pesticide Actives**: 329
- **Donated Pharmaceuticals**: 135
- **NTP In Vivo**: 202
- **FDA CFSAN**: 94
- **Total In vivo**: 580

- **1060 Total chemicals → 2806 total overlaps across 16 diverse inventories**
- **Excellent coverage of multiple high-interest inventories**
- **Broad diversity of chemical-use and types**
- **Large overlap with data-rich inventories**
- **Majority of cmpds on 2 or more lists**

**Office of Research and Development**
**National Center for Computational Toxicology**
ToxCast_1060 Pharmaceuticals: Multipurposing

- 255 of 1060 cmpds classified as drugs based on presence on FDAMDD, NCGC, NCTR LKB, Donated pharma lists
- 150 of these drugs appear on other high interest lists, i.e. multipurpose
- 137 drugs in ToxCast_1060 appear on 2 or more lists
- Caffeine appears on 18 lists
Assay Similarity Analysis: NVS Biochemical Assays

Steroid Receptors

<table>
<thead>
<tr>
<th>Assay</th>
<th>Count</th>
<th>Low Similarity</th>
<th>High Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>NR_hER</td>
<td>8</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>NR_bER</td>
<td>9</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>NR_mERa</td>
<td>10</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>NR_hAR</td>
<td>8</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>NR_hPR</td>
<td>3</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>NR_bPR</td>
<td>3</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>NR_rAR</td>
<td>3</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>NR_hGR</td>
<td>6</td>
<td>105</td>
<td></td>
</tr>
</tbody>
</table>
# Assays Affected

Chemical Similarity Analysis: NVS Biochemical Assays

- **High Similarity**
  - 90
  - Pharmaceuticals
  - GPCR inhibitors
  - Steroid receptor binding chemicals
  - Pesticides
    - PPO inhibitors
    - Organochlorine pesticides

- **Low Similarity**

*Slide courtesy of Nisha Sipes*
Tox21 Chemical x Assay Landscape

- **Tox21**: ~10,000 Chem x 50-100 assays
- **ToxCast Phase I & II**: 1060 Chem x >500 assays
- **NTP HTS Plate A**: 1408 Chem x >100 assays

**# Assays**
- >500
- ~1000
- 100

**# Chemicals**
- ~10,000
- 50-100
- 1408

**1536 well microplate format (1408 cmpds/plate) x 9 plates**

**Agencies**
- NCCT/EPA: Environmental, Industrial Pesticides, Food Use, Drugs
- NIEHS/NTP: Toxicology
- NIH/NCGC: FDA, CFSAN/CDER

**Tox21**

**Office of Research and Development**
National Center for Computational Toxicology
Tox21 Chemicals: NTP Selections

- NTP-studied compounds of all types, including:
  - *Chronic, Rodent carcinogenicity*
  - *Genotoxicity*
  - *Reproductive & developmental toxicity*
  - *Immunotoxicity*

- NTP nominations and related compounds

- ICCVAM & NICEATM validation and reference compounds

- Nominations from outside collaborators
  - *e.g., U.S. Army Public Health Command*
### Drug Source

<table>
<thead>
<tr>
<th>Drug Source</th>
<th>In house</th>
<th>Procurement in process</th>
</tr>
</thead>
<tbody>
<tr>
<td>US FDA</td>
<td>1635</td>
<td>182</td>
</tr>
<tr>
<td>UK/EU/Canada/Japan</td>
<td>756</td>
<td>177</td>
</tr>
<tr>
<td>Investigational</td>
<td>928</td>
<td>3953</td>
</tr>
<tr>
<td>Total Approved</td>
<td>2391</td>
<td>359</td>
</tr>
<tr>
<td>Total</td>
<td>3319</td>
<td>4312</td>
</tr>
</tbody>
</table>

#### Informatics sources
- US FDA: Orange Book, OTC, NDC, Green Book, Drugs at FDA
- Britain NHS
- EMEA
- Health Canada
- Japan NHI

#### Physical sources
- Procurement from >70 suppliers worldwide
- In-house purification of APIs from marketed forms
- Synthesis

*In house drug plate composition*
Tox21/ToxCast: Analytical QC

A copy of each parent Tox21 assay plate (352 cmpds/plate) will be subjected to analytical QC for assessing purity, identity, stability.

PASS = Confirm parent ion peak and >90% purity

Fail, inconclusive or analytical method inappropriate

Retest at later time point under assay conditions for stability

Publish QC summary results in association with assay data
Analytical QC Summary PDF results will be available for each analyzed sample.

NCGC Analytical Chemist (B. Leister) will review all preliminary LC results and QC “FAIL” compounds, and supervise all follow-up testing.

Prime objective of QC results is to inform analysis and interpretation of assay results:

- high confidence → low confidence → fail
Tox21/ToxCast Test Sample Registry

- Procured chemicals
  - Source, Lot/Batch, Bottle code
- Test sample info from CoA/MSDS
- Stock solution used in plating
- Substance & structure-annotation and registration in DSSTox

DSSTox
Chemical structure - CID
Substance details - SID
Project inventory record - RID

Tox21 Sample Tracking Database

DSSTox RID
Bottle ID (→ COA ID)
Solution ID (→ QC ID)
Tox21 Cheminformatics

Chemical Substance: SID

Structure: CID

Parent: PID

Features
Classes
Descriptors
Properties

2D, 3D Conformers
Tautomers

Salts
Complexes
(Metabolites?)

Tox21/ToxCast

Assay Results

Bottle IDs

Solution IDs

Plate IDs, Plate Addresses

CAS/Name

Mixture IDs

Analytical QC

Office of Research and Development
National Center for Computational Toxicology
### Tox21 Structural Library

<table>
<thead>
<tr>
<th>Feature</th>
<th>NCGC Tox21</th>
<th>EPA Tox21</th>
<th>NTP Tox21</th>
</tr>
</thead>
<tbody>
<tr>
<td>urea</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>thiocarboxylates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>thiocarboxamide</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sulfoxide</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sulfonic acid</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sulfone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sulfonate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sulfonamide</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sulfide</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>silicon groups</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>quinones</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>phosphorous groups</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>organometal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nitroso</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nitro</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nitrile</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mercaptan</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ketone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>isocyanate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>iminomethyl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>imine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hydroxylamine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hydrazine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>carbonyl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>carbamate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>boron groups</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>azide</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>amines</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>amidine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alkynes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alkenes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aldehydes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alcohols</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>acid halides</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>acid anhydrides</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Feature enrichment

#### Feature diversity
Tox21 Reaction Features: Commonalities

Metabogen reaction features generated using “MOSES” software by Molecular Networks

Drugs
Food additives
Antimicrobials
Water contaminants
HPVs
Toxicants
…
ToxCast & Tox21 Property Space

- **LOG P** = Octanol/Water partition coefficient
- **TPSA** = \(\log (\text{Total Polar Surface Area})\)
- **Complexity** = \(\log (\text{complexity based on paths, branching, atoms})\)

*Chemical properties computed using “Adrianna” software by Molecular Networks (P. Volarath)*
ToxCast & Tox21 Property Space

- LOG P = Octanol/Water partition coefficient
- TPSA = log (Total Polar Surface Area)
- Complexity = log (complexity based on paths, branching, atoms)

ToxCast Phase II (767)
ToxCast Phase I (293)

Chemical properties computed using “Adrianna” software by Molecular Networks (P. Volarath)
ToxCast & Tox21 Property Space

ToxCast PhaseI (293)
ToxCast PhaseII (767)
Donated Pharma (135)

LOG P = Octanol/Water partition coefficient
TPSA = log (Total Polar Surface Area)
Complexity = log (complexity based on paths, branching, atoms)

Chemical properties computed using “Adrianna” software by Molecular Networks (P. Volarath)
LOG P = Octanol/Water partition coefficient
TPSA = log (Total Polar Surface Area)
Complexity = log (complexity based on paths, branching, atoms)

Chemical properties computed using “Adrianna” software by Molecular Networks (P. Volarath)
ToxCast & Tox21 Property Space

ToxCast PhaseII (767)
ToxCast PhaseI (293)
Tox21 (7324 unique)
ToxCast e1k (+800)
Donated Pharma (135)

- LOG P = Octanol/Water partition coefficient
- TPSA = log (Total Polar Surface Area)
- Complexity = log (complexity based on paths, branching, atoms)

Chemical properties computed using “Adrianna” software by Molecular Networks (P. Volarath)
Chemical/Substance representations for hazard & risk evaluation

- Nanomaterial Characterizations
  - Nano dispersion media
  - Fate & Bioavailability
  - Exposure route

- Toxicity Data
  - Toxicity structural alerts
  - HPV categories
  - TTC categories

- Biological & Chemical Reactivity
  - EPA MoA classes
  - Organic chemistry classes
  - Chemical & metabolic reactivity

- Intrinsic Properties of Chemical or Substance
  - PhysChem properties
  - Calculated descriptors
  - Structural features

- Exposure Scenario-Informed
  - Usage
  - Biology
  - Interactions
  - Chemistry
- MOA QSAR can inform assay aggregation & pathway hypothesis
- Reactivity & toxicity-informed features & chemical clusters serve to guide & facilitate such efforts
Meteor and Derek (Lhasa, Ltd.) prediction systems rely on “structure alerts” extracted from available data & expert knowledge to predict potential for:

- metabolic biotransformation (Meteor)
- in vitro and in vivo toxicities (Derek Nexus)

Use for cheminformatics analysis:

1. Employ Meteor and Derek alerts for clustering, and building chemical-reactivity and toxicity-informed similarity metrics for use in grouping chemicals for modeling;
2. Develop & apply a Meteor-Derek Workflow (MDW) to identify chemicals predicted to be toxic (by Derek) only after metabolic activation (by Meteor).

Volarath et al., SOT poster, Mar 2012
### Asses Meteor Content & Coverage

<table>
<thead>
<tr>
<th>Meteor (Rat) – Totals</th>
<th># ToxCast Chemicals (All Enzymes)</th>
<th># Tox21 Chemicals (CYP450 only)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processed with bio-transformations</td>
<td>636 (66%)</td>
<td>5432 (76%)</td>
</tr>
<tr>
<td>Processed but no bio-transformations found</td>
<td>324</td>
<td>1569</td>
</tr>
<tr>
<td>Chemicals not processed</td>
<td>0</td>
<td>142</td>
</tr>
<tr>
<td><strong>Total # chemicals in the dataset</strong></td>
<td><strong>960</strong></td>
<td><strong>7143</strong></td>
</tr>
</tbody>
</table>

- Majority ToxCast & Tox21 chemicals trigger alerts in Meteor (Rat)
- Approx 50% of Meteor enzymes “hit” by ToxCast
- Tox21 chemicals hit large number (all?) of CYP450 alerts

**Meteor has 64 metabolic enzymes modules for 28 mammalian species, with 484 total biotransformations (structure alerts).**
### Assess Derek Content & Coverage

<table>
<thead>
<tr>
<th>Derek (Rat) – Totals</th>
<th># ToxCast Chemicals</th>
<th># Tox21 Chemicals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processed by Derek – alerts found</td>
<td>637 (66%)</td>
<td>4387 (61%)</td>
</tr>
<tr>
<td>Processed by Derek but no alerts found</td>
<td>323</td>
<td>2642</td>
</tr>
<tr>
<td>Chemicals not processed</td>
<td>0</td>
<td>114</td>
</tr>
<tr>
<td><strong>Total # chemicals in the dataset</strong></td>
<td><strong>960</strong></td>
<td><strong>7143</strong></td>
</tr>
</tbody>
</table>

- 60% of all Derek alerts found in ToxCast
- 83% of all Derek alerts found in Tox21
- 86% of Derek (Rat) endpoints predicted by ToxCast cmpds
- 98% of Derek (Rat) endpoints predicted by Tox21 compounds

*Derek has 45 Rat endpoints with 478 alerts*
Meteor Biotransformation Alert Fingerprint Clustering (Rat, CYP450)

Chemicals cluster based on common set of Meteor (biotransformation) alert features
Chemicals cluster based on common set of Derek alert features

Chemicals cluster differently based on Meteor (biotransformation) vs. Derek (toxicity) alerts

Derek Alert Fingerprint Clustering
Chemicals are grouped by endpoint or endpoint cluster (sharing common alert features).
METEOR predicts possible metabolic transformations based on recognition of structural features.

DEREK predicts in vivo toxicity outcomes based on recognition of structural features (e.g., rat carcinogenicity).

• Workflow identifies ToxCast & Tox21 compounds more likely to require metabolic activation to produce *in vivo* (Derek endpoint) response [Parent-/Metabolite+]

• HTS *in vitro* to *in vivo* models are more likely to poorly predict such compounds (if assays are not metabolically competent)
Toxicity Prediction Challenge

MOA QSAR & Cheminformatics

- MOA chemotypes; Reactivity & toxicity-informed features & classes

Aggregation

- Mechanistically well-defined toxicity endpoint; Sufficient data

Data-mining

- MOA:
  > Pathways
  > Genes
  > Assays
  + Statistical associations

In Vitro/HTS

In Vivo

Existing knowledge
COSMOS Substance Inventory

- Over 7000 unique substances
- Structurally diverse
- Many well populated clusters of similar compounds
- Structure curation & QC by Altimira & DSSTox
- Will be made publicly available thru DSSTox & COSMOS

Office of Research and Development
National Center for Computational Toxicology
Tox21 & COSMOS Property Space

- Tox21 unique (7324)*
- + COSMOS-Tox21 overlap (1063)**

- Significant overlap of COSMOS with Tox21 structure inventory
- 481 ToxCast overlaps (incl. 234 PhI&II and 247 e1k)
- Overlaps span large portion of the property space of Tox21

LOG P = Octanol/Water partition coefficient
TPSA = log (Total Polar Surface Area)
Complexity = log (complexity based on paths, branching, atoms)

* desalted, duplicates eliminated, computable structures
** overlap includes structurable & computable unique entries

Chemical properties computed using “Adrianna” software by Molecular Networks (P. Volarath)
Tox21 & COSMOS Property Space

- Tox21 unique (7324)*
- + COSMOS-Tox21 overlap (1063)**
- x COSMOS-Not in Tox21 (1746)**

- LOG P = Octanol/Water partition coefficient
- TPSA = log (Total Polar Surface Area)
- Complexity = log (complexity based on paths, branching, atoms)

* desalted, duplicates eliminated, computable structures
** overlap includes structurable & computable unique entries

Chemical properties computed using “Adrianna” software by Molecular Networks (P. Volarath)
ToxCast & Tox21 Property Space

- Tox21 unique (7324)*
- + COSMOS-Tox21 overlap (1063)**
- x COSMOS-Not in Tox21 (1746)**

- COSMOS structures not in Tox21 concentrate in same regions of property space as those overlapping w/Tox21
- Tox21 cmpds provide excellent coverage of COSMOS structure property space
- Overlaps span large portion of the property space of Tox21

- LOG P = Octanol/Water partition coefficient
- TPSA = log (Total Polar Surface Area)
- Complexity = log (complexity based on paths, branching, atoms)

* desalted, duplicates eliminated, computable structures
** overlap includes structurable & computable unique entries

Chemical properties computed using “Adrianna” software by Molecular Networks (P. Volarath)
Conclusions

- Toxicity prediction challenge is challenging!
- HTS *in vitro* data provide intermediate biological input that can inform biological-based modeling of *in vivo* outcomes
- HTS chemical libraries serve as probes of this biology
- Statistical approaches alone (chemical or biological) insufficient
- MOA & chemistry concepts guide modeling into productive domains
- Knowledge-based chemical representations can facilitate MOA-based and statistical modeling
- *Challenge is to use ALL available data domains*
Acknowledgements:

- EPA NCCT ToxCast Team:
  Robert Kavlock
  David Dix
  Keith Houck
  Matt Martin (ToxRefDB)
  Richard Judson (ACToR)
  David Reif (ToxPi)
  Tom Knudsen (vEmbryo)
  Imran Shah (vLiver)
  Patra Volarath
  Nisha Sipes
  Nicole Kleinstrauer
  Alicia Frame

- EPA NCCT DSSTox:
  Maritja Wolf – Lockheed Martin, Contractor to the EPA
  Indira Thillainadarajah – EPA:SEE

- External Collaborators:
  MOA QSAR: Chihae Yang (Altimira); Andrew Worth, Kirk Arvidson (FDA)
  Tox21: Chris Austin & colleagues, NCGC/NIH
  Tox21: Ray Tice & colleagues, NTP/NIEHS

This work was reviewed by EPA and approved for publication but does not necessarily reflect official Agency policy.