Cheminformatics and Toxicogenomics for Toxicity Prediction and Mechanistic Insight

George Daston
Overview

• Tiered approach to predicting toxicity of new chemicals
  – Cheminformatics- supported SAR
  – High-content methods to assess SAR solutions

• Identifying MOA using cheminformatics and toxicogenomics
  – MOA ontology
  – Connectivity mapping
Toxicology: From an Empirical to a Predictive Science

Traditional Approach (Black box): Use a model that we have (some) confidence in, but incomplete understanding of how it works

Desired Approach: Predictions based on deep, fundamental understanding
Taking Advantage of the Existing Literature

• Considerable outcome data in DART (almost 12,000 entries in publicly available databases)
• Pressing need is to identify initial molecular events
• Effort needed to connect initial events with tissue/organ level effects
Initial Screening for Human Hazards

• Substructure searching
  – Genotoxicity (19,300)
  – Carcinogenicity (15,800)
  – Skin Sensitization (9,400)
  – Skin Irritation (10,400)
  – Reproductive/Developmental Toxicity (11,300)
  – Subchronic/Chronic Toxicity (15,100)
  – Acute Toxicity (68,500)

• All assessment captured in CHS

Flow chart of new analog identification & evaluation process

Wu et al., RTP, 2010
Searching GRASP- Substructure
Searching

Search Structure

\[
\begin{align*}
\text{Search Structure} & \quad \rightarrow & \quad \text{Product} \\
\text{Search Structure} & \quad \rightarrow & \quad \text{Product}
\end{align*}
\]
Output – Substructure Searching

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>SEARCH STRUCTURE</td>
<td></td>
</tr>
<tr>
<td><strong>Substances</strong></td>
<td><strong>A. Alias</strong></td>
</tr>
<tr>
<td>4</td>
<td>103652</td>
</tr>
<tr>
<td>5</td>
<td>103852</td>
</tr>
<tr>
<td>6</td>
<td>103652</td>
</tr>
<tr>
<td>7</td>
<td>103852</td>
</tr>
<tr>
<td>8</td>
<td>103652</td>
</tr>
<tr>
<td>9</td>
<td>103652</td>
</tr>
<tr>
<td>10</td>
<td>103652</td>
</tr>
<tr>
<td>11</td>
<td>104455</td>
</tr>
<tr>
<td>12</td>
<td>104455</td>
</tr>
</tbody>
</table>

Ready
Suitable Analogs

CAS# 80-27-3

CAS# 1338-43-8

CAS# 8007-35-0

CAS# 8007-43-0
Possibly Suitable

CAS# 1338-43-8

CAS# 1338-92-3
Unsuitable Analogs

CAS# 177406-46-1 VS CAS# 131-17-9

Chemical structures showing chemical transformations and their implications on hepatotoxicity.
Nrf2 qHTS screen for inhibitors: counterscreen for cytotoxicity

qHTS Assay for Inhibitors of RanGTP induced Rango (Ran-regulated importin-beta cargo) – Importin beta complex dissociation

qHTS Assay for Inhibitors of JMJD2A-Tudor Domain

Chemical Probe  Active  Inactive  Inconclusive  Unspecified
Cheminformatics: Ontology

- Use large database to organize chemicals into mode of action groupings
- Start to estimate the extent of “the universe of toxicity mechanisms”
- This will allow us to design a suite of model systems that is comprehensive
Initial Concept

• An initial list of ~260 chemicals with DART data was originally developed as part of an evaluation of Threshold of Toxicologic Concern (TTC) (Laufersweiler et al., 2012)
• These chemicals were grouped based on their chemical characteristics and this tree was published in concept in Blackburn et al. (2011)
P&G DART tree + CAESAR for test set (106 active, 73 non-active)

Accuracy: ~86%, Sensitivity: 93% and Specificity: 77%
Putative MOA Grouping by Chemical Structure

• 25 major categories, multiple sub-categories
• Highest level of confidence has
  – Similar structures
  – Identified molecular target
  – Similar DART outcome (e.g., common syndrome or highly specific effect)
• Along with toxicogenomics, has the potential to accelerate assigning MOA to DART compounds
Hierarchy Examples

- Nuclear hormone receptor ligands
- Prostaglandin receptor ligands
- Nicotinic ACh receptor ligands and AChesterase inhibitors
- Shh signaling interference/ cholesterol synthesis inhibitors
- Nucleotide derivatives
Nuclear Hormone Receptor Ligands

- Estrogen and androgen receptor ligands
- Glucocorticoid receptor ligands
- Retinoic acid receptor ligands
- Thyroid hormone receptor ligands
- Ah receptor ligands
Nuclear hormone receptor ligands

- Estrogen and androgen receptor ligands
  - steroid nucleus-derived compounds
    - Estradiol-like
    - Progesterone, androgens, steroidal anti-androgens
  - Non-steroidal compounds
    - Flavones and mycoestrogens
    - Alkylphenols
    - N-aryl-substituted ureas, carbamides, amides
    - other
Nuclear hormone receptor ligands

• Estrogen and androgen receptor ligands
  – steroid nucleus-derived compounds
    • Estradiol-like
    • Progesterone, androgens, steroidal anti-androgens
  – Non-steroidal compounds
    • Flavones and mycoestrogens
    • Alkylphenols
    • N-aryl-substituted ureas, carbamides, amides
    • other
Key functional groups:
C=O at C-3; -OH and alkyl (C1-C3 carbons), ethyn at C-17
Nuclear hormone receptor ligands

• Estrogen and androgen receptor ligands
  – steroid nucleus-derived compounds
    • Estradiol-like
    • Progesterone, androgens, steroidal anti-androgens
  – Non-steroidal compounds
    • Flavones and mycoestrogens
    • Alkylphenols
    • N-aryl-substituted ureas, carbamides, amides
    • other
R=OH, H; R₁=OH
R₂=mono-, di-, tri-, OH-Ph, MeO-Ph
R₃=mono-, di-, tri-, OH-Ph
R₂ and R₃ can not be present at C-2 and C-3 simultaneously

R=OH, =O
X-X=C-C, C=C
Nuclear hormone receptor ligands

• Estrogen and androgen receptor ligands
  – steroid nucleus-derived compounds
    • Estradiol-like
    • Progesterone, androgens, steroidal anti-androgens
  – Non-steroidal compounds
    • Flavones and mycoestrogens
    • Alkylphenols
    • N-aryl-substituted ureas, carbamides, amides
      • other
R=OH, NH₂
n=1, X=C, R₁=Alkyl (C₁-C₄)
R₂=Me
R₁, R₂=isobenzofuranone
n=2, R₁ and R₂ are on different C's
n=2, X=C-C, R₁, R₂=H, Me, Et
n=2, X=C=C, R₁, R₂=H, Me, Et
n=1, X=O, S, SO₂, R₁=R₂=none

X-Y= C-C
R=OH, Cl, OMe
R₁=H, Cl
X-Y= C=C
R=OH, Cl, OMe
R₁=none
Ah Receptor Ligands

- TCDD-like chemicals
  - cleft palate, hydronephrosis and reproductive system defects
- Indole-related compounds: repro system
- Polycyclic aromatics
- Halogenated aromatics (e.g., PCBs)
  - Liver cyp induction leads to DART effects?

\[ \text{structures with } R = \text{Cl, Br}, \ X = O, \text{none} \]
Problems with the chemical approach

• Promiscuous chemicals that have more than one molecular target

• Seemingly similar compounds that have different developmental outcome
  – PK differences?
  – More than one target?
  – Insufficient potency against target?
Outcome of chemistry assessment is hypothesis generation

• Chemical is metabolized to a tested chemical, or to a known active metabolite
  – Currently, assessment is done by wet lab metabolism

• Chemical is sufficiently similar in structure to analogs of known toxicity that similar biological activity is inferred
  – Currently, assessment is done by MOA-specific evaluation
  – Add ToxCast and other PubChem data to our databases and our expert considerations about mechanism
  – Global analysis of gene expression
Using Gene Expression Analysis to Inform MOA and AOP

• Gene expression is specific for MOA
• In vitro models may have great potential to identify MOA via gene expression
Two Close Structural Analogs

DEHP

DINP

DEHP-24
DEHP-48
DINP-24
DINP-48
Connectivity Mapping: High-throughput toxicogenomics

• Concept developed by Lamb in 2006
  – A relatively small number of carefully selected cell types contained all of the pathways necessary to define gene expression profiles for all therapeutic agents in current use

• Can we do the same for toxicants?
  – Cell types: rich in either small molecule receptors or metabolizing enzymes
MOAs to Interrogate with CMAP

- Estrogens, environmental estrogens
- Anti-estrogens
- PPAR agonists
- Anti Androgen agonists
- CAR/PXR agonists
- RAR agonists
- TR agonists
- AhR agonists
- Vitamin D agonist
- Glucocorticoid receptor agonists

- EGFR receptor agonists
- FXR receptor agonists
- Progesterone receptor agonists
- EGFR antagonist
- Steroid synthesis inhibitors
- HDAC inhibitors
- Folate/one-carbon metabolism inhibitor
- Glycolytic inhibitors
- Oxidative phos/mitochondrial inhibitors
- Iron chelators
- Microtubule inhibitors
- Liver cholestasis inducers
# Genes Significantly Changed

<table>
<thead>
<tr>
<th>Chemical</th>
<th>MCF7</th>
<th>Ishikawa</th>
<th>HepG2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bisphenol A</td>
<td>76</td>
<td>5262</td>
<td>9247</td>
</tr>
<tr>
<td>Trenbolone</td>
<td>188</td>
<td>18</td>
<td>3</td>
</tr>
<tr>
<td>methotrexate</td>
<td>3296</td>
<td>16</td>
<td>5376</td>
</tr>
<tr>
<td>vorinostat</td>
<td>17342</td>
<td>19432</td>
<td>21798</td>
</tr>
<tr>
<td>RU486</td>
<td>106</td>
<td>4</td>
<td>22</td>
</tr>
<tr>
<td>Vitamin D3</td>
<td>519</td>
<td>93</td>
<td>2</td>
</tr>
<tr>
<td>Amoxicillin</td>
<td>6</td>
<td>29</td>
<td>810</td>
</tr>
</tbody>
</table>
Connectivity Mapping: Example

Bisphenol A comparisons

- DES
- resveratrol
- epitiostanol
- equilin
- genistein
- genistein
- estrone
- genistein
- estradiol
- levonorgestrel
- resveratrol
- equilin
landmark genes

• expression of 978 *landmark* genes measured
  – selected from large, diverse, high-quality microarray dataset
  – orthogonal expression and validated predictive power

• inputs for genome-wide inference model
  – compute expression of transcripts not explicitly measured
  – flagged as *LM* (rather than *INF*) in output data file

>100,000 Affy U133 scans

gene  gene correlation  landmarks
AFFX versus L1000

• create *signatures* for each treatment from AFFX data
  – treatment (n=1) *versus* corresponding vehicle control (n=1)
  – 50 up- and 50 down- regulated genes by signal-to-noise

• create *instances* for each treatment from L1000 data
  – rank all features by extent of differential expression
  – treatment compared with matched control sample

• compute enrichment of each signature in each instance
  – rank instances based on these connectivity scores
  – AFFX signature finds expected L1000 instances in 9 of 10 tests

signature: bisphenol A
AFFX versus L1000

ranks of L1000 instances of each treatment with specified AFFX signature
Modeling PK to Ensure the Right In Vitro Concentration

- Dose matters: data obtained in vitro at irrelevant concentrations is also irrelevant in predicting risk
- Active concentrations at the target tissue in vivo are predictable
PK workflow

Chemical structure, vehicle or formulation information

Phys-chem property prediction

Kasting skin macro penetration model

PBPK ADME workbench

Physiological descriptors and their variability

Partition and diffusion coefficients for different layers of the skin

Metabolism, protein binding data

Partition coefficients for different organs

Concentration

Time
Modeling AUC for a Range of Absorption Values
Conclusions

• Chemical ontology can aid in assigning chemicals to groups with same putative MOA
• It is possible to estimate the size of the MOA universe
• Linking initial molecular event with outcome will require considerable hypothesis testing, aided by gene expression data, modeling and simulation
• It is already possible to estimate tissue dose using computation, phys chem parameter estimation, and judicious data generation
Acknowledgements

• Cheminformatics
  – Shengde Wu
  – Karen Blackburn
  – Jorge Naciff
  – Joan Fisher

• Toxicogenomics
  – Jorge Naciff
  – Yuching Shan
  – Xiaohong Wang
  – Jay Tiesman
  – Greg Carr
  – Nadira deAbrew

• PK Modeling
  – Joanna Jaworska
  – Russell Devane
  – John Troutman

• CMAP (external)
  – Justin Lamb
  – Rusty Thomas
  – Ed Carney