Application of matrix factorization method for kinase data to support safety profiling

Linlin Zhao            06/29/2022
Kinase inhibition and toxicity

Protein kinases are enzymes that mediate the phosphorylation reaction during the signaling process in human. Kinases inhibitors are well known to be involved in organ toxicities such as cardiotoxicity, ocular toxicity, and skin toxicity.

Screening compounds against human kinases is not only crucial for exploring primary therapeutic opportunities, but also important for risk assessment of drug candidates.

Kinase assay dataset

Overall, only ~2% of possible measurements were present in the data set.
Macau for data imputation

Modeling workflow

Kinase assay IC50 data/Inhibition percentage data

- Side information for kinases (TAPE embedding descriptors)

GNE compounds

- Side information for compounds (ECFP Fingerprints)

Training Set
- (80% of rows, 80% of columns)

Test Set 2
- (80% of rows, 20% of columns)

Test Set 1
- (20% of rows, 80% of columns)

Test Set 3
- (20% of rows, 20% of columns)

Internal Test Set
- (20% masked data)

Model A: Training Set

Model B: Training Set + Side information for compounds (ECFP Fingerprints)

Model C: Training Set + Side information for kinases (TAPE embedding descriptors)

Model D: Training Set + Side information for compounds (ECFP Fingerprints) + Side information for kinases (TAPE embedding descriptors)
## Results

### IC50 dataset

<table>
<thead>
<tr>
<th>Kinases</th>
<th>RMSE</th>
<th>R2</th>
<th>Compounds (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kinase_1</td>
<td>0.47</td>
<td>0.36</td>
<td>20</td>
</tr>
<tr>
<td>Kinase_2</td>
<td>0.74</td>
<td>0.33</td>
<td>22</td>
</tr>
<tr>
<td>Kinase_3</td>
<td>0.77</td>
<td>0.28</td>
<td>2</td>
</tr>
<tr>
<td>Kinase_4</td>
<td>0.4</td>
<td>0.25</td>
<td>11</td>
</tr>
<tr>
<td>Kinase_5</td>
<td>0.87</td>
<td>0.2</td>
<td>3</td>
</tr>
<tr>
<td>Kinase_6</td>
<td>0.68</td>
<td>0.15</td>
<td>6</td>
</tr>
<tr>
<td>Kinase_7</td>
<td>0.83</td>
<td>0.1</td>
<td>130</td>
</tr>
</tbody>
</table>

### Inhibition Percentage dataset

<table>
<thead>
<tr>
<th>Kinases</th>
<th>RMSE</th>
<th>R2</th>
<th>Compounds (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kinase_8</td>
<td>17.99</td>
<td>0.57</td>
<td>2</td>
</tr>
<tr>
<td>Kinase_9</td>
<td>17.87</td>
<td>0.34</td>
<td>13</td>
</tr>
<tr>
<td>Kinase_10</td>
<td>18.8</td>
<td>0.24</td>
<td>17</td>
</tr>
<tr>
<td>Kinase_11</td>
<td>20.3</td>
<td>0.23</td>
<td>15</td>
</tr>
<tr>
<td>Kinase_12</td>
<td>19.77</td>
<td>0.18</td>
<td>7</td>
</tr>
<tr>
<td>Kinase_13</td>
<td>18.54</td>
<td>0.12</td>
<td>13</td>
</tr>
<tr>
<td>Kinase_14</td>
<td>26.09</td>
<td>0.12</td>
<td>10</td>
</tr>
</tbody>
</table>

Inhibition Percentage model performance on selected kinases that are relative to Training Set.

IC50 model performance on selected kinases that are relative to Training Set.
Doing now what patients need next
Thank you!
Revealing Adverse Outcome Pathways from Public High-Throughput Screening Data to Evaluate New Toxicants by a Knowledge-Based Deep Neural Network Approach

Heather L. Ciallella, Daniel P. Russo, Lauren M. Aleksunes, Fabian A. Grimm, and Hao Zhu

Heather L. Ciallella, Ph.D.
Zhu Research Lab
Rutgers Center for Computational & Integrative Biology

SOT CTSS Award Winner Webinar
June 29, 2022
The uterotrophic assay

molecular interaction
chemical

immature or OVX female mouse or rat
(\geq 6\) control and treated)

\sim \$30,000, 9 days

organ
cellular response

ER agonism
↑ ER activation
↑ uterine wt
↑ ovarian wt

organ system
altered estrous cyclicity

organism
decreased age/wt at VO

adverse outcome pathway

population
altered development

bottom figure adapted from Browne P et al. Environ Health Perspect. 2017;125(9):096001.
Regulatory acceptance of alternative ER methods

- **<1996**: Pharma uses high-throughput screening to rapidly screen chemicals
- **1996**: EPA kicks off EDSP
- **1998**: Scientists propose endocrine effects of chemicals
- **2007**: ToxCast screening initiated
- **2009**: EPA requests the National Research Council to review strategy for toxicity testing
- **2015**: 18 ToxCast/Tox21 assays and computational model accepted as alternative to ER EDSP Tier 1 screening

Conclusions

-Estrogen receptor QSAR
-Estrogen receptor k-DNN
-Dev tox read-across

adapted from Daniel Russo
Estrogen receptor signaling

Figure created with Biorender
Pathway modeling for toxicity predictions

Embedding ER pathway assays into a neural network
Opening the black box of neural networks

Opening the black box of neural networks

Improving predictions with pathway analysis

Acknowledgements

**Zhu research group**

Dr. Hao Zhu
Elena Chung
Nada Daoood
Xuelian Jia
Dr. Daniel Russo
Swati Sharma
Saeideh Nasiri

**Yitao Shen**
Dr. Tong Wang
Dr. Wenyi Wang
Dr. Min Wu
Dr. Xiliang Yan
Dr. Linlin Zhao

**Doctoral committee members**

Dr. Sunil Shende, Rutgers-Camden Computer Science
Dr. Nir Yakoby, Rutgers-Camden Biology
Dr. Lauren Aleksunes, Rutgers-New Brunswick Pharmacology & Toxicology

**Industry collaborators**

Dr. Fabian Grimm, EMBSI
Dr. Colin North, EMBSI
Leveraging high-throughput screening data, deep neural networks, and conditional generative adversarial networks to advance predictive toxicology

Dr. Adrian Green
Postdoctoral Research Scholar
NC State University
The data gap

- About 89,000 chemicals registered with the US Environmental Protection Agency (EPA)
- The EPA started the ToxCast program in 2007 and currently have rigorous toxicological data ~4,600
- To address this gap, high-throughput screening (HTS) and computational methods are vital.
What options are available to screen environmental chemicals?

Humans  Mammals  Vertebrates  Invertebrates  Cells  Biochemical

Each offers pros/cons in terms of: throughput, cost, human relevance, specificity (targets), complexity (development, systemic interactions).

Toxicological endpoints such as abnormal behavior or development are difficult to measure using purely *in vitro* systems.

High-throughput studies using embryonic zebrafish complement targeted approaches and provide systematic data that can be used for integrated analysis across *in silico, in vitro, and multi-scale in vivo* endpoints.
Computational tools

Computational approaches to bridge this data gap have been developed utilizing chemical structure.

- Quantitative Structure-Activity Relationship (QSAR)
- Read-Across

Recent machine learning advances such as deep neural networks (DNNs) and conditional generative adversarial networks (cGANs) have not been thoroughly explored.

Study design

Data collected on individual zebrafish for each of 1000+ unique chemicals:
- x 6 concentrations
- x 32 biological (well) replicates
- x time series
- x 22 morphological endpoints

Chemical space: How similar are the datasets

![Chemical space diagram]

<table>
<thead>
<tr>
<th>Data</th>
<th>Active Chemical</th>
<th>Inactive Chemical</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training data</td>
<td>159</td>
<td>844</td>
<td>1003</td>
</tr>
<tr>
<td>External testing data</td>
<td>7</td>
<td>49</td>
<td>56</td>
</tr>
</tbody>
</table>

Table 1. Summary of training and testing data used in this study.
# Toxicity data

## Morphology Assessment

### 24 hpf End Points

<table>
<thead>
<tr>
<th>End Point</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortality</td>
<td>MO24</td>
</tr>
<tr>
<td>Developmental Delay</td>
<td>DP24</td>
</tr>
<tr>
<td>Spontaneous Movement</td>
<td>SM24</td>
</tr>
<tr>
<td>Notochord</td>
<td>NC24</td>
</tr>
</tbody>
</table>

### 120 hpf End Points

<table>
<thead>
<tr>
<th>End Point</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortality</td>
<td>MORT</td>
</tr>
<tr>
<td>Yolk Sac Edema</td>
<td>YSE</td>
</tr>
<tr>
<td>Body Axis</td>
<td>AXIS</td>
</tr>
<tr>
<td>Eye Defect</td>
<td>Eye</td>
</tr>
<tr>
<td>Snout</td>
<td>SNOU</td>
</tr>
<tr>
<td>Jaw</td>
<td>JAW</td>
</tr>
<tr>
<td>Otic Vesicle</td>
<td>OTIC</td>
</tr>
<tr>
<td>Pericardial Edema</td>
<td>PE</td>
</tr>
<tr>
<td>Brain</td>
<td>BRAI</td>
</tr>
<tr>
<td>Somite</td>
<td>SOMI</td>
</tr>
<tr>
<td>Pectoral Fin</td>
<td>PFIN</td>
</tr>
<tr>
<td>Caudal Fin</td>
<td>CFIN</td>
</tr>
<tr>
<td>Pigment</td>
<td>PIG</td>
</tr>
<tr>
<td>Circulation</td>
<td>CIRC</td>
</tr>
<tr>
<td>Truncated Body</td>
<td>TRUN</td>
</tr>
<tr>
<td>Swim Bladder</td>
<td>SWIM</td>
</tr>
<tr>
<td>Notochord &amp; Bent Tail</td>
<td>NC</td>
</tr>
<tr>
<td>Touch Response</td>
<td>TR</td>
</tr>
</tbody>
</table>

---

3D Chemical Structure

2D chemical compounds (EPA’s Chemistry Dashboard) → Open Babel Ligand Preparation algorithm → 3D chemical structures

Example Chemical Structure

Green,A.J. et al. (2021) PLOS Computational Biology.
The DNN (Go-ZT) architecture

\[ [v_i] \rightarrow \text{Generator chemical feature layers} \rightarrow \text{Calculate weighted sum of output} \rightarrow f(v_i) \cdot w_i \rightarrow \text{Generator toxicity layer} \rightarrow \text{Generated toxicity matrix} \rightarrow \text{Empirical toxicity matrix} \rightarrow \text{Compare and fit using regression} \]
The cGAN (GAN-ZT) architecture
Training results

A

Go-ZT Training Loss

B

GAN-ZT Training Loss

Green, A.J. et al. (2021) PLOS Computational Biology.
## Training results

Performance of different methods in activity classification with 10-fold cross-validation.

<table>
<thead>
<tr>
<th>Model</th>
<th>SE</th>
<th>SP</th>
<th>PPV</th>
<th>Kappa</th>
<th>AUROC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>SVM</td>
<td>17.7</td>
<td>10.9</td>
<td>92.1</td>
<td>3.6</td>
<td>0.30</td>
</tr>
<tr>
<td>MLP</td>
<td>12.2</td>
<td>7.60</td>
<td>94.5</td>
<td>1.6</td>
<td>28.2</td>
</tr>
<tr>
<td>RF</td>
<td>6.5</td>
<td>6.10</td>
<td>98.2</td>
<td>3.1</td>
<td>56.7</td>
</tr>
<tr>
<td>GAN-ZT</td>
<td>58.4</td>
<td>20.7</td>
<td>64.1</td>
<td>19.4</td>
<td>28.4</td>
</tr>
<tr>
<td>Go-ZT</td>
<td>44.6</td>
<td>7.25</td>
<td>97.1</td>
<td>1.65</td>
<td>76.1</td>
</tr>
</tbody>
</table>

- **SE** – Sensitivity, how well the model identifies active chemicals
- **SP** – Specificity, how well the model identifies inactive chemicals
- **PPV** – Positive Predictive Value, proportion of chemicals identified as active that are truly active
- **Cohen's Kappa** – Evaluates the performance of a classification model
- **AUROC** – Evaluates the performance of a classification model

Green, A.J. et al. (2021) PLOS Computational Biology.
**Training results**

Performances of different methods in activity prediction of test set chemicals.

<table>
<thead>
<tr>
<th>Model</th>
<th>SE</th>
<th>SP</th>
<th>PPV</th>
<th>Kappa</th>
<th>AUROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>28.6</td>
<td>95.9</td>
<td>50.0</td>
<td>0.300</td>
<td>0.649</td>
</tr>
<tr>
<td>MLP</td>
<td>28.6</td>
<td>89.8</td>
<td>28.6</td>
<td>0.184</td>
<td>0.660</td>
</tr>
<tr>
<td>RF</td>
<td>28.6</td>
<td>98.0</td>
<td>66.7</td>
<td>0.351</td>
<td>0.459</td>
</tr>
<tr>
<td>GAN-ZT</td>
<td>71.4</td>
<td>59.2</td>
<td>20.0</td>
<td>0.146</td>
<td>0.653</td>
</tr>
<tr>
<td>Go-ZT</td>
<td>71.4</td>
<td>91.8</td>
<td>55.6</td>
<td><strong>0.564</strong></td>
<td><strong>0.816</strong></td>
</tr>
</tbody>
</table>

SE – Sensitivity, how well the model identifies active chemicals  
SP – Specificity, how well the model identifies inactive chemicals  
PPV – Positive Predictive Value, proportion of chemicals identified as active that are truly active  
Cohen's Kappa – Evaluates the performance of a classification model  
AUROC – Evaluates the performance of a classification model
Ensemble model

Performances of combined model performance.

<table>
<thead>
<tr>
<th>Model</th>
<th>PPV</th>
<th>Kappa</th>
<th>AUROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAN-ZT</td>
<td>20.0</td>
<td>0.146</td>
<td>0.653</td>
</tr>
<tr>
<td>Go-ZT</td>
<td>55.6</td>
<td>0.564</td>
<td>0.816</td>
</tr>
<tr>
<td>Comb-ZT</td>
<td>71.4</td>
<td>0.673</td>
<td>0.837</td>
</tr>
</tbody>
</table>
Conclusions and next steps

- My combined model is a good predictor of empirical toxicity.
- A DNN utilizing 3D chemical structural information is a useful prescreening tool for predicting the toxic outcomes of the approximately 80,000 untested chemicals registered with the USEPA.

Next steps:
- Create a priority list of chemicals to screen.
- Empirically validate predictions.
- Develop deep learning methodologies to zebrafish behavior.
Acknowledgements

Postdoc Mentor
David Reif

Collaborators
Robyn Tanguay
Lisa Truong

Lab Members
Skylar Marvel
Preethi Thunga
Dylan Wallis
Nnamdi Osakwe
Jonathan Fleming
Melody Hancock

Funding
National Institute of Environmental and Health Sciences (P30 ES030287, R56 ES030007, P30 ES025128)
and the National Cancer Institute (R01 CA161608)
Mechanism-driven Modeling of DILI Using Structural Alerts and an In Vitro Screening Assay

Xuelian Jia
Dr. Hao Zhu Cheminformatics lab
Rutgers - Center for Computational and Integrative Biology
Drug-induced liver injury

- Liver is an organ for transforming/eliminating chemicals, vulnerable to the toxicity.

- Drug-induced liver injury (DILI) is an adverse reaction to drugs or other xenobiotics and can be potentially fatal.

- DILI is a leading cause of drugs failure during clinical trials and being withdrawn from the market.

**Intrinsic**

(direct or predictable) type is dose-related and occurs shortly after exposure in most individuals exposed to the drug, which is toxic at a given threshold level (e.g., acetaminophen).

**Idiosyncratic**

(indirect or unpredictable) type occurs less frequently, has a longer latency period, determined by the interaction of host factors with the drug.

http://chembl.blogspot.com/2018/06/withdrawn-drugs.html
Traditional toxicology testing

Expensive  
Time consuming  
Ethics issue

Fail to identify toxicants that cause liver injury in clinical stage

Alternative non-animal approaches
An adverse outcome pathway (AOP) is a structured representation of biological events leading to toxicological effects. It involves:

- **molecular initiating event (MIE)**,
- a series of intermediate steps and **key events (KEs)**,
- an **adverse outcome** (hepatotoxicity).

The pathway includes:

- **Toxicants**
- **Chemical features**
- **Molecular interaction**
- **Gene upregulation**
- **Stress response**
- **Apoptosis/cytotoxicity**
- **Liver injury**

**Virtual AOP modeling** and **Computational approaches** are used to study these pathways.

**In vitro assays** are also utilized to test the outcomes.
Structure-activity relationships

- A method to identify a drug’s inherent liability leading to toxicity; chemical mechanism
- Search for structural motifs or structural alerts that commonly occur in compounds known to cause a particular effect.

MULTICASE Algorithm

- hierarchical algorithm that breaks the learning set into logical subsets
- commonality between the molecules is based on a rational evaluation of their structures rather than on an arbitrary choice of common structural features.

Workflow of study

1. Dataset collection
2. Profiling PubChem assays
3. QSAR modeling for ARE activation
4. Identification of structural alerts
5. Experimental validation and DILI prediction
Individual assay’s correlation to DILI

- Number of test results > 150
- Specificity > 0.75
- PPV > 0.75
- CCR > 0.53

543 assays

24 assays

Target information of individual assays

<table>
<thead>
<tr>
<th>ID</th>
<th>AID a</th>
<th>Target</th>
<th>Class</th>
<th>ID</th>
<th>AID a</th>
<th>Target</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1347083</td>
<td>Arenaviruses</td>
<td>A</td>
<td>13</td>
<td>743219</td>
<td>ARE</td>
<td>D</td>
</tr>
<tr>
<td>2</td>
<td>1347085</td>
<td>Arenaviruses</td>
<td>A</td>
<td>14</td>
<td>1346795</td>
<td>PR</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>1347090</td>
<td>cancer</td>
<td>A</td>
<td>15</td>
<td>1345199</td>
<td>tumor</td>
<td>A</td>
</tr>
<tr>
<td>4</td>
<td>743122</td>
<td>AhR</td>
<td>E</td>
<td>16</td>
<td>1296009</td>
<td>carcinoma</td>
<td>A</td>
</tr>
<tr>
<td>5</td>
<td>1347100</td>
<td>cancer</td>
<td>A</td>
<td>17</td>
<td>1259378</td>
<td>ER-beta</td>
<td>B</td>
</tr>
<tr>
<td>6</td>
<td>743140</td>
<td>PPARg</td>
<td>B</td>
<td>18</td>
<td>1159523</td>
<td>ROR-gamma</td>
<td>B</td>
</tr>
<tr>
<td>7</td>
<td>1347093</td>
<td>cancer</td>
<td>A</td>
<td>19</td>
<td>743202</td>
<td>ARE</td>
<td>D</td>
</tr>
<tr>
<td>8</td>
<td>1347031</td>
<td>PR</td>
<td>B</td>
<td>20</td>
<td>686971</td>
<td>tumor</td>
<td>A</td>
</tr>
<tr>
<td>9</td>
<td>743069</td>
<td>ER-alpha</td>
<td>B</td>
<td>21</td>
<td>662</td>
<td>CRE-BLA</td>
<td>A</td>
</tr>
<tr>
<td>10</td>
<td>743078</td>
<td>ER-alpha</td>
<td>B</td>
<td>22</td>
<td>884</td>
<td>CYP 3A4</td>
<td>C</td>
</tr>
<tr>
<td>11</td>
<td>743199</td>
<td>PPARg</td>
<td>B</td>
<td>23</td>
<td>1347050</td>
<td>hNpr2</td>
<td>A</td>
</tr>
<tr>
<td>12</td>
<td>1259403</td>
<td>ERR</td>
<td>B</td>
<td>24</td>
<td>1825</td>
<td>KLF5</td>
<td>A</td>
</tr>
</tbody>
</table>

- AID: PubChem Assay Identifier
- Class: Categories of target and number of assays belong to this class
  A: Cancer/diseases  11
  B: Nuclear receptors  9
  C: CYP450 enzymes  1
  D: oxidative stress  2
  E: transcription factor  1
Structural alerts identified using DILI data

Potential structural alerts for hepatotoxicity and their correlations to DILI.

<table>
<thead>
<tr>
<th>Structures</th>
<th>Occurrences</th>
<th>PPV</th>
<th>Structures</th>
<th>Occurrences</th>
<th>PPV</th>
<th>Structures</th>
<th>Occurrences</th>
<th>PPV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>48</td>
<td>0.73</td>
<td></td>
<td>43</td>
<td>0.72</td>
<td></td>
<td>40</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>0.81</td>
<td></td>
<td>35</td>
<td>0.69</td>
<td></td>
<td>22</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.92</td>
<td></td>
<td>11</td>
<td>0.82</td>
<td></td>
<td>9</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>1.00</td>
<td></td>
<td>5</td>
<td>1.00</td>
<td></td>
<td>5</td>
<td>1.00</td>
</tr>
</tbody>
</table>

exist in \( \geq 5 \) compounds \( \text{ppv} > 0.68 \) 27 alerts

Some of these chemical structures have been known or suspected to be metabolized into reactive intermediates and induce oxidative stress in liver.
oxidative stress mechanistic DILI model

Drug compounds
Alerts check
ARE testing
DILI prediction

MIE
Key events
Adverse Outcome

ARE

Toxic
Inconclusive
Inconclusive
Non-Toxic

oxidative stress mechanistic DILI model

Drug compounds
Alerts check
ARE testing
DILI prediction

MIE
Key events
Adverse Outcome

ARE

Toxic
Inconclusive
Inconclusive
Non-Toxic

oxidative stress mechanistic DILI model
Many other toxicity mechanisms can also lead to DILI.
<table>
<thead>
<tr>
<th>No.</th>
<th>Drug Name</th>
<th>Structural alerts</th>
<th>ARE result</th>
<th>DILI Prediction</th>
<th>DILI Exp.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>QSAR</td>
<td>Exp. of ARE+Alerts</td>
<td>Exp. of ARE+Alerts</td>
</tr>
<tr>
<td>1</td>
<td>Cinchophen</td>
<td>Yes</td>
<td>Active</td>
<td>Active</td>
<td>Tox</td>
</tr>
<tr>
<td>2</td>
<td>Exifone</td>
<td>Yes</td>
<td>Active</td>
<td>Active</td>
<td>Tox</td>
</tr>
<tr>
<td>3</td>
<td>Thiabendazole</td>
<td>Yes</td>
<td>Active</td>
<td>Active</td>
<td>Tox</td>
</tr>
<tr>
<td>4</td>
<td>Diclofenac</td>
<td>Yes</td>
<td>Active</td>
<td>Active</td>
<td>Tox</td>
</tr>
<tr>
<td>5</td>
<td>Benzbromarone</td>
<td>Yes</td>
<td>Active</td>
<td>Active</td>
<td>Tox</td>
</tr>
<tr>
<td>6</td>
<td>Clozapine</td>
<td>Yes</td>
<td>Active</td>
<td>Active</td>
<td>Tox</td>
</tr>
<tr>
<td>7</td>
<td>Ritonavir</td>
<td>Yes</td>
<td>Active</td>
<td>Active</td>
<td>Tox</td>
</tr>
<tr>
<td>8</td>
<td>Sorafenib</td>
<td>Yes</td>
<td>Active</td>
<td>Active</td>
<td>Tox</td>
</tr>
<tr>
<td>9</td>
<td>Nortriptyline</td>
<td>Yes</td>
<td>Active</td>
<td>Active</td>
<td>Tox</td>
</tr>
<tr>
<td>10</td>
<td>Dexpanthenol</td>
<td>No</td>
<td>Inactive</td>
<td>Inactive</td>
<td>Non-Tox</td>
</tr>
<tr>
<td>11</td>
<td>L-carnitine</td>
<td>No</td>
<td>Inactive</td>
<td>Inactive</td>
<td>Non-Tox</td>
</tr>
<tr>
<td>12</td>
<td>Mannitol</td>
<td>No</td>
<td>Inactive</td>
<td>Inactive</td>
<td>Non-Tox</td>
</tr>
<tr>
<td>13</td>
<td>Benztropine</td>
<td>No</td>
<td>Inactive</td>
<td>Inactive</td>
<td>Non-Tox</td>
</tr>
<tr>
<td>14</td>
<td>Edrophonium</td>
<td>No</td>
<td>Inactive</td>
<td>Inactive</td>
<td>Non-Tox</td>
</tr>
</tbody>
</table>

Structural alerts are highlighted in red.
Experimental validation - external compounds

<table>
<thead>
<tr>
<th>No.</th>
<th>Drug Name</th>
<th>Structural alerts</th>
<th>ARE result</th>
<th>DILI Prediction</th>
<th>DILI Exp.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>QSAR</td>
<td>Exp.</td>
<td>QSAR of ARE+Alerts</td>
</tr>
<tr>
<td>1</td>
<td>Boceprevir</td>
<td>Yes</td>
<td>Active</td>
<td>Active</td>
<td>Tox</td>
</tr>
<tr>
<td>2</td>
<td>Ixabepilone</td>
<td>Yes</td>
<td>Active</td>
<td>Active</td>
<td>Tox</td>
</tr>
<tr>
<td>3</td>
<td>Crizotinib</td>
<td>Yes</td>
<td>Active</td>
<td>Active</td>
<td>Tox</td>
</tr>
<tr>
<td>4</td>
<td>Cabazitaxel</td>
<td>Yes</td>
<td>Active</td>
<td>Active</td>
<td>Tox</td>
</tr>
<tr>
<td>5</td>
<td>Hypericin</td>
<td>No</td>
<td>Active</td>
<td>Inactive</td>
<td>Inconclu</td>
</tr>
<tr>
<td>6</td>
<td>Cediranib</td>
<td>No</td>
<td>Active</td>
<td>Active</td>
<td>Inconclu</td>
</tr>
<tr>
<td>7</td>
<td>Etifoxine</td>
<td>No</td>
<td>Active</td>
<td>Active</td>
<td>Inconclu</td>
</tr>
<tr>
<td>8</td>
<td>Panobinostat</td>
<td>No</td>
<td>Inactive</td>
<td>Active</td>
<td>Non-Tox</td>
</tr>
<tr>
<td>9</td>
<td>Mitoguazone</td>
<td>No</td>
<td>Inactive</td>
<td>Inactive</td>
<td>Non-Tox</td>
</tr>
<tr>
<td>10</td>
<td>Tremorine</td>
<td>No</td>
<td>Inactive</td>
<td>Inactive</td>
<td>Non-Tox</td>
</tr>
<tr>
<td>11</td>
<td>Potassium hydroxocitrate</td>
<td>No</td>
<td>Inactive</td>
<td>Inactive</td>
<td>Non-Tox</td>
</tr>
<tr>
<td>12</td>
<td>fotemustine</td>
<td>Yes</td>
<td>Inactive</td>
<td>Inactive</td>
<td>Inconclu</td>
</tr>
</tbody>
</table>

Structural alerts are highlighted in red.
Incorporating multiple assays in hepatotoxicity predictions

Oxidative stress mechanistic model:
high number of FNs indicates these compounds lead to hepatotoxicity through other toxicity mechanisms

1) agonists of PXR (AID 1347033);
2) CYP3A4 induction through PXR (AID 1346984);
3) disruptors of the mitochondrial membrane potential (MMP) (AID 720637);
4) agonists of the constitutive androstane receptor (CAR) pathway (AID 1224839);
5) cytotoxicity in HepG2 cells (AID 1224878).

<table>
<thead>
<tr>
<th>NAME</th>
<th>CID</th>
<th>Hepatotoxicity</th>
<th>Model prediction</th>
<th>PXR</th>
<th>CYP3A4_PXR</th>
<th>MMP</th>
<th>CAR</th>
<th>Cytotoxicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>albendazole</td>
<td>2082</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fenofibrate</td>
<td>3339</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>riluzole</td>
<td>5070</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ticlopidine</td>
<td>5472</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

most modeling set compounds have missing data in these assays
Summary

- This mechanistic model explored the potential of integrating the AOP concept and HTS screening into a new computational approach.
- Using experimental ARE results could correct possible false prediction.
- Extra data obtained from assays measuring other key events of hepatotoxicity AOPs can aid in reducing FNs in our model.
- This strategy can be applied to develop predictive models for other hepatotoxicity mechanisms and other complex toxicity endpoints.
Acknowledgements

**Zhu Research Lab**
- PI: Dr. Hao Zhu
- Dr. Dan Russo
- Dr. Heather Ciallella
- Elena Chung
- Tong Wang
- Nada Daood
- Yitao Shen
- Dr. Linlin Zhao
- Swati Sharma

**Collaborators**
- Dr. Lauren Aleksunes
- Dr. Xia Wen