DEVELOPMENT OF MECHANISM-DRIVEN VIRTUAL ADVERSE OUTCOME PATHWAY (VAOP) FOR HEPATOTOXICITY

By

XUELIAN JIA

A thesis submitted to the

Graduate School-Camden

Rutgers, the State University of New Jersey

in partial fulfillment of the requirements

For the degree of

Master of Science

Graduate Program in Computational and Integrative Biology

Written under the direction of

Dr. Hao Zhu

and approved by

_____

Dr. Hao Zhu

_____

Dr. Jinglin Fu

_____

Dr. Sunil Shende

_____

Dr. Lauren M. Aleksunes

Camden, New Jersey

October, 2022

THESIS ABSTRACT

Development of Mechanism-driven Virtual Adverse Outcome Pathway (vAOP) for Hepatotoxicity

By XUELIAN JIA

Thesis Director:

Dr. Hao Zhu

The liver is an important organ for transforming and eliminating chemicals and thus is vulnerable to toxicity from the toxicants. A broad class of chemicals can be potential liver toxicants, including environmental and industrial chemicals, herbal and dietary supplements, traditional medicines, and medications. Drug-induced liver injury (DILI) represents the acute and chronic liver injuries that are caused by medications. Drug attrition during clinical trials and post-marketing because of DILI can cause extremely high expenses. As a result, there is great interest by regulators to develop *in vitro* and computational modeling to help identify which chemicals have the propensity to cause liver injury in the early stage of safety evaluation. For the prediction of complex toxicity endpoints like hepatotoxicity, using traditional computational strategies (e.g., Quantitative Structure-Activity Relationship, QSAR) and structural, chemical properties is not sufficient and often error-prone. As an advanced framework of risk assessment, the Adverse Outcome Pathway (AOP) was introduced to describe the mode and mechanism

of toxicant action. The mechanisms of DILI are complex and be explained by various AOPs. In this study, we specifically focus on oxidative stress-involved hepatotoxicity. Reactive metabolites formed during drug metabolism or inhibition of the bile salt export pump can cause oxidative stress, which triggers the transcription of antioxidative enzymes found in the antioxidant response element (ARE) signaling pathway. The quantitative HTS (qHTS) ARE activation assay screened more than 10,000 compounds of interest and is an indicator of chemical-induced oxidative stress and subsequent hepatotoxicity. This assay, along with other *in vitro* mechanism-related assays in public big data sources will be collected and combined with advanced machine learning and deep learning algorithms, for the development of the virtual AOP (vAOP). The resulting vAOP framework will reveal hepatotoxicity mechanisms within the available big data and resolve the limitations of traditional QSAR modeling by providing accurate mechanism-based predictions for new compounds.

INTRODUCTION

Because of its relationship to the gastrointestinal tract and unique role in metabolism, the liver is a vulnerable target organ subject to the toxicity of drugs, xenobiotics, and oxidative stress. Hepatotoxicity is the leading cause of drug failure in clinical trials and after withdrawal from the market. During 1975 and 2007, 32% of drug withdrawals were attributed to hepatotoxicity (Andrade et al., 2019; Stevens and Baker, 2009). As a result, pharmaceutical companies suffer major losses due to such late-stage attrition of clinical candidates, black box warnings, and post-marketing drug withdrawals (Gijbels and Vinken, 2019). Some of these adverse events can be serious in nature and as evidenced by DILI. The mechanism of DILI is complicated and can be classified into intrinsic and idiosyncratic types based on the chemical's presumed mechanism of action. The intrinsic DILI is relatively direct and predictable, 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). In contrast, idiosyncratic DILI is unpredictable, occurs less frequently, has a longer latency period, and is determined by the interaction of environmental and host factors with the drug (Andrade et al., 2019). Normal drug metabolism processes involve drug uptake, Phase I and Phase II metabolism, and drugs/metabolites efflux, which are controlled by a large family of proteins that collectively influence the accumulation of drugs or their metabolites and lead to the stress effects in the liver. Drugs are taken up into hepatocytes passively or by an array of transporters located in the basolateral membrane. After that, drugs are metabolized by Phase I and Phase II metabolism. After the phase I reactions (oxidation, reduction, and hydrolysis), the metabolites usually have only minor structural differences from the parent drug but can

have very different pharmacological actions. Phase II metabolism involves the conjugation of a drug or metabolite with endogenous molecules such as glucuronic acid, sulfate or glutathione resulting in a more polar product that usually does not have pharmacological activity. Drugs and metabolites efflux from hepatocytes into the bile or back into the sinusoidal blood for subsequent renal excretion, which is mediated mainly by ATP-binding cassette (ABC) transporters such as multidrug resistance protein 1 (MDR1), also called P-glycoprotein. Inhibition of drug efflux transporters, formation of reactive metabolites during phase I and II reactions, and inhibition of the bile salt export pump can be possible mechanisms of DILI.

Determination of the liver safety profile for a drug is time-consuming and expensive, usually necessitating the exposure of hundreds of thousands of animals to the drug compound. Furthermore, drugs that cause severe DILI in humans typically do not show clear hepatotoxicity in animals, do not show dose-related toxicity, and cause low rates of severe injury (Xu et al., 2015). This could partly be explained by significant gaps in the mechanistic understanding of DILI (Bale et al., 2014). People hope to predict DILI by using biomarkers and a series of *in vitro* assays at the molecular level or cellular level to predict the *in vivo* liver injury, the so-called adverse outcome pathway. The AOP is a conceptual construct that portrays existing knowledge concerning the linkage between a direct molecular initiating event and an adverse outcome at a biological level relevant to risk assessment (Vinken, 2018). Many AOPs are under development for predicting DILI. The AOP developed by Mathieu's team describes the mechanistic basis of drug-induced cholestatic hepatotoxicity, with inhibition of the bile salt export pump (BSEP) as the molecular initiating event (MIE). Inhibition of BSEP causes the accumulation of bile

acids in the hepatocyte cells. The latter key event (KE) induces two types of cellular responses: a deteriorative response and an adaptive response. The deteriorative response involves the occurrence of inflammation, change of mitochondrial permeability, oxidative stress, and cell death. The adaptive response involves the activation of several nuclear receptors (NRs), which induces an array of transcriptional changes to facilitate the removal of bile salts and their products (Gijbels et al., 2020). Oxidative stress is a KE is the AOP of DILI caused by BSEP inhibitor. For other drug compounds, the formation of reactive metabolites by cytochrome P450 enzymes could also induce oxidative stress (Park et al., 2011). To counterbalance excessive reactive oxidants, the human body developed complex antioxidant defense systems which are regulated by a web of pathways. As a regulator of cellular response to oxidants, the nuclear factor erythroid 2–related factor 2 (Nrf2) controls the basal and induced expression of an array of antioxidant response element–dependent genes to control oxidant homeostasis in addition to drug metabolism (Ma, 2013). When the antioxidant defenses are inadequate, mitochondrial dysfunction, hepatocyte cell death can occur and process to liver injury.

In recent years, many computational models have been developed for predicting DILI. For example, Kotsampasakou et al built classification models to predict drug-induced cholestasis, using physicochemical descriptors and predicted transporter inhibition profiles as features (Kotsampasakou and Ecker, 2017). Gadaleta et al developed QSAR models to predict the MIE leading to hepatic steatosis (Gadaleta et al., 2018). In another study, Li et al described a deep learning-powered DILI (DeepDILI) prediction model created by combining conventional machine learning algorithms with a deep learning framework (Li et al., 2021). However, these models were mainly developed using

chemical structures and properties, lacking the integration of enormous *in vitro* bioactivity data. Moreover, the predictions were often generated from "black box" models and cannot help revealing the mechanisms of toxicity. For several years, both individual initiatives and consortium organizations have produced large amounts of data, including *in vitro* assays and gene expression analysis. Benefiting from the high throughput screening (HTS) technique development and many associated data-sharing projects, modern drug discovery has stepped into a big data era (Zhu, 2020). For example, PubChem is a publicly available big data resource with over 96 million compounds, including many drugs and drug-like compounds, tested against over 1 million bioassays (Kim, Thiessen, et al., 2016a; Wang et al., 2017a). Significant efforts in HTS toxicology have been made by the United States Environmental Protection Agency (US EPA) research program Toxicity Forecaster (ToxCast) and Toxicology in the 21st Century (Tox21). These HTS assays test a large number of chemicals against various human cells and have quantitative results that allow for mechanistic interpretation (Gibb, 2008). This data landscape enables researchers to create predictive computational models that incorporate the concept of the AOP with publicly available big data, resulting in mechanism-driven virtual AOP (vAOP) models. This project aims to build vAOP models that can not only predict the DILI risk of new compounds but also illustrate toxicity mechanisms of importance in humans.

RESEARCH DESIGN AND METHODS

AIMS

The liver plays an important role in detoxification and metabolism, which makes it highly vulnerable to injury by environmental chemicals, commercial products, and drugs. Late-stage drug attrition and post-marketing withdrawal because of hepatotoxicity cause drug failure and major loss of drug companies. The animal models used in preclinical hepatotoxicity evaluation are expensive and often fail to identify toxicants that cause liver injury in the clinical stage, emphasizing the demand for developing new approaches that predict drug-induced liver injury (DILI) in the early stages of drug development (Vorrink et al., 2018). In 2016, the Frank R. Lautenberg Chemical Safety for the 21st Century Act (LCSA) was signed into law to progress chemical risk assessment. The LCSA calls for novel computational approaches and associated predictive models for safety evaluation purposes. Adverse outcome pathway (AOP) is an important tool that maps the mechanisms underlying toxic events relevant for chemical risk assessment. The central goal of this project is to develop a virtual Adverse Outcome Pathway (vAOP) model that could accurately identify chemicals with a high propensity to induce DILI.

Section I. Collect *in vitro* and *in vivo* data related to oxidative stress and hepatotoxicity

In this project, we specifically focus on oxidative stress-driven hepatotoxicity. Activation of the Antioxidant Response Element (ARE) is a sensitive indicator of chemical-induced oxidative stress and subsequent hepatotoxicity (Shukla et al., 2012). The quantitative High Throughput Screening (qHTS) ARE beta-lactamase (bla) reporter gene assay was one of the assays included in the Tox21 program, which screened >10,000 compounds of interest (Betts, 2013; Shukla et al., 2012). Hepatotoxicity databases which are reported by U.S. Food & Drug Administration (FDA) (Chen et al., 2016), as well as Toxicity Reference Database (ToxRefDB) (Judson et al., 2012) and in the literature, will be collected as major sources of human hepatotoxicity data. Bioactivity repositories such as PubChem (Kim, Thiessen, et al., 2016a; Wang et al., 2017a) will also be used for short-term bioassay data. Bioactivity of qHTS assays in toxicology programs like Toxicity Forecaster (ToxCast) and Tox 21 will be explored to unveil potential toxicological mechanisms. These datasets will be used for developing, validating, and sharing toxicity models needed for Section II and III.

Results - datasets collection

To study DILI and develop content-rich resources to improve basic understanding of liver toxicity, FDA's National Center for Toxicological Research (NCTR) started the project Liver Toxicity Knowledge Base (LTKB) (Chen et al., 2011). The initial benchmark dataset (LTKB-BD) contains 287 drugs whose potential to cause DILI in humans has been established using the FDA-approved prescription drug labels. The DILIrank dataset is an updated version of the LTKB-BD and is the largest reference drug list ranked by the

risk for developing DILI in human (Chen et al., 2016). DILIrank consists of 1,036 FDA-approved drugs that are divided into four classes according to their potential for causing DILI: vMost-, vLess-, vNo-DILI concern and Ambiguous-DILI-concern. The classification is derived from analyzing the hepatotoxicity descriptions presented in the FDA-approved drug labeling documents and assessing causality evidence in the literature. Specifically, this largest publicly available annotated DILI dataset contains three groups (vMost-, vLess- and vNo-DILI concern) with confirmed causal evidence linking a drug to liver injury and one additional group (Ambiguous-DILI-concern) with causality undetermined.

Known data resources that can be utilized for data collection and future modeling is provided in **Table 1**. LiverTox is a publicly available website that provides information about DILI caused by prescription and non-prescription medications, herbal products, and dietary supplements (ncbi.nlm.nih.gov/books/NBK547852/) (Hoofnagle et al., 2013). LiverTox is produced by the Liver Disease Research Branch of the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), with the aim to help physicians, patients as well as researchers understanding the idiosyncratic DILI. Comparable Toxicogenomics Database is a premier public big data resource that provides data describe relationships between chemicals, gene expressions, phenotypes, diseases, and environmental exposures (Mattingly et al., 2003; Mattingly et al., 2004; Mattingly et al., 2006). The AOPwiki website (aopwiki.org) is the repository of qualitative information for the international AOP development effort coordinated by the Organisation for Economic Co-operation and Development (OECD). There are 15 liver-related AOPs on the AOPwiki. Most of these AOPs are still under development and may undergo

modifications before being accepted by the OECD. Although much of the data stored on these websites may not be suitable for modeling, we could use the information for understanding the mechanism of action for drug compounds of interest.

The qHTS assays from ToxCast and Tox21 project tested a large amount of chemicals against various human cells for the toxicity-related molecular and pathway perturbations, which allow for mechanistic interpretation (Gibb, 2008). Phase I of ToxCast project evaluated 300 pesticides using about 500 HTS assays (Judson et al., 2010). Phase II evaluated an additional 767 compounds, including some failed pharmaceutical compounds, using about 700 HTS assays (Kavlock et al., 2012). Phase I of Tox21 used 75 HTS assays, which were selected and refined from ToxCast assays, to screen an initial set of about 2800 compounds (Attene-Ramos et al., 2013). Phase II began in 2010 to screen a more extensive set of approximately 10,000 environmental compounds (Tice et al., 2013). The Connectivity Map project (Lamb et al., 2006) created gene-expression profiles from cultured human cells treated with bioactive small molecules, providing functional connections among diseases, genetic perturbation, and drug action. As the next generation Connectivity Map, the L1000 project has developed a low-cost high throughput transcriptomic assay using 978 “landmark” genes from human cells (Subramanian et al., 2017). Measurement of these “landmark” genes were able to infer the expression levels of 81% of non-measured transcripts. The resulted transcriptomic profiles for multiple cell lines were generated in response to around 20,000 small molecule perturbagens. The L1000 provides us an excellent resource for studying chemical toxicity at the transcriptome level, and some relevant data could be used in our mechanism driven model development.

In this project, we specifically focus on oxidative stress-driven DILI. The quantitative HTS (qHTS) Antioxidant Response Element (ARE) beta-lactamase (bla) reporter gene assay was one of the assays included in the Tox21 program, a federal collaboration between the US FDA, EPA, and NIH to improve chemical toxicity prediction (Betts, 2013; Shukla et al., 2012). Activation of the ARE is a sensitive indicator of chemical-induced oxidative stress and subsequent hepatotoxicity (Shukla et al., 2012). The qHTS ARE-bla data sets can also be downloaded from PubChem using Bioassay Accession Identifiers (AIDs) 743202 and 651741.

Table 1. Summary of public data resources

Name	Data type	Description
DILIrank[1]	In vivo hepatotoxicity	1,036 FDA-approved drugs: 192 vMost, 278 vLess, 312 vNo, 254 Ambiguous-DILI-concern
LiverTox [2]	In vivo hepatotoxicity	up-to-date information on the diagnosis, cause, frequency, clinical patterns, and management of liver injury attributable to ~1000 medications
PubChem[3]	In vitro assays	over 96 million compounds, over 1 million bioassays, over 13 billion data points related to toxicity, genomics and literature data
ToxCast/Tox 21[4]	Toxicity-related in vitro assays	test ~10,000 chemicals against a panel of nuclear receptor and stress response pathway assays
CTD*[5]	chemical-gene interaction	includes more than 30.5 million toxicogenomic connections relating chemicals/drugs, genes/proteins, diseases, taxa, Gene Ontology (GO) annotations, pathways, and gene interaction modules.
L1000[6]	chemical-gene interaction	almost two million gene expression profiles for ~20,000 small molecules and drugs against 978 “landmark” genes from human cells

* Comparable Toxicogenomics Database, [1] (Chen et al., 2016), [2] (Hoofnagle et al., 2013), [3] (Kim, Thiessen, et al., 2016b; Wang et al., 2017b), [4] (Attene-Ramos et al., 2013; Gibb, 2008; Judson et al., 2012; Kavlock et al., 2012; Tice et al., 2013), [5] (Mattingly et al., 2003; Mattingly et al., 2004; Mattingly et al., 2006), [6] (Subramanian et al., 2017).

Results **- data curation**

The DILIrank dataset was used as the primary source of high-quality compounds for model development. First, 254 drug compounds belong to Ambiguous-DILI-concern class were excluded. Compounds belonging to the Most and Less-DILI-concern were classified as hepatotoxic, and compounds belong to the No-DILI-concern were classified as non-toxic. Then, structures in the DILIrank dataset were curated and standardized using the CASE-Ultra DataKurator 1.8.0.0 software (MulitCASE Inc., Beachwood, OH). This includes the removal of duplicates, mixtures, inorganics, and correction of structural errors. The curated DILIrank dataset consisted of 680 unique compounds, including 432 hepatotoxic and 248 non-toxic compounds. There are more toxic compounds than non-toxic compounds, which provides bias chemical space and is not suitable for model development. To balance the toxic/non-toxic ratio, we need to add more non-toxic compounds. As shown in **Table 2**, we collected and curated hepatotoxicity datasets from multiple research papers, whose *in vivo* hepatotoxicity defined using different standards. We harmonized various hepatotoxicity classifications into binary classifications of 1 (hepatotoxic) and 0 (nontoxic). Overlap compounds were identified among different datasets. If the overlap compounds yield conflict hepatotoxicity classifications, these compounds will be excluded. The high-quality non-toxic compounds were collected by selecting compounds that showed consistent hepatotoxicity classifications among the 7 literature datasets (at least two datasets). Those compounds that already exist in the DILIrank dataset (including Ambiguous-DILI-concern drugs) were excluded. As a result, the balanced hepatotoxicity dataset consists of 432 hepatotoxic and 450 non-toxic compounds.

Table 2. Hepatotoxicity Datasets from literature

Dataset	Initial size	After curation	Curation	Reference	Original source
1	534	345	salt not included	(Ekins et al., 2010)	Self - compilation
2	951	909	Only human data were used	(Fourches et al., 2010)	Self - compilation
3	605	596	Excluding inconclusive	(Liu et al., 2015)	LiverTox 2014
4	287	263	Most, less-concern as 1; no concern as 0	(Chen et al., 2011)	LTKB-DB
5	1314	1309	Remove no CID structures	(Kim, Huang, et al., 2016)	FDA liver damage data, 2014
6	3712	2167	Only use Level 0 human data	(Mulliner et al., 2016)	Self - compilation
7	1274	1248	Remove no CID structures	(Liew et al., 2011)	Self - compilation

Section II. Develop mechanism-driven vAOP models for hepatotoxicity

This section will identify the novel chemical *in vitro*-*in vivo* relationships to generate vAOP models for hepatotoxicity. Initial profiles for the model include important chemical features and bioprofiles across multiple assays. Relevant portions of the chemical structure related to toxic properties (known as toxicophores) will be identified using our custom fragment generation tool along with commercial software (i.e., Chemotyper®(Yang et al., 2015)). The toxicophore information will serve as the chemical and structural profiles for target compounds. Using our automatic data mining tool, compounds will then be used as the probe to search for relevant *in vitro* biological and toxicological data within public sources (e.g., PubChem, ToxCast/Tox 21, L1000). Compilation of these data using novel knowledgebase deep neural networks (k-DNN) will result in a bioprofile containing billions of data points generated from thousands of different assays with established toxicity relationships for the target compounds.

Results **- QSAR modeling of ARE activation assay**

Understandably, not all the *in vivo* hepatotoxicity compounds have the *in vitro* ARE activation activity. In our balanced hepatotoxicity dataset, 221 out of the 882 drug compounds have ARE activation testing results (**Figure 1**). Missing data severely limits the identification of the *in vitro* bioactivity and *in vivo* hepatotoxicity relationships in this study. Thus, we built QSAR models to predict the drug compounds that do not have the testing results of ARE activation assay. The ARE assay with PubChem AID 743202 was selected for QSAR modeling. The downloaded data table contains 9305 items, which were curated and standardized as following: 1) items have inconclusive responses or do not have PubChem Compound Identifiers (CIDs) were removed 2) items have the same

CID but showed different activity outcomes were also removed, 3) for the remaining compounds, those share the same structure (identified by CASE-Ultra) but have different activities will be removed. These efforts resulted in 3,394 unique compounds with unambiguous activity results (277 actives, 3,117 inactives). To get the antive/inactive balanced training set, we include all the 277 active compounds and randomly selected 300 compounds from the 3,117 inactive compounds.

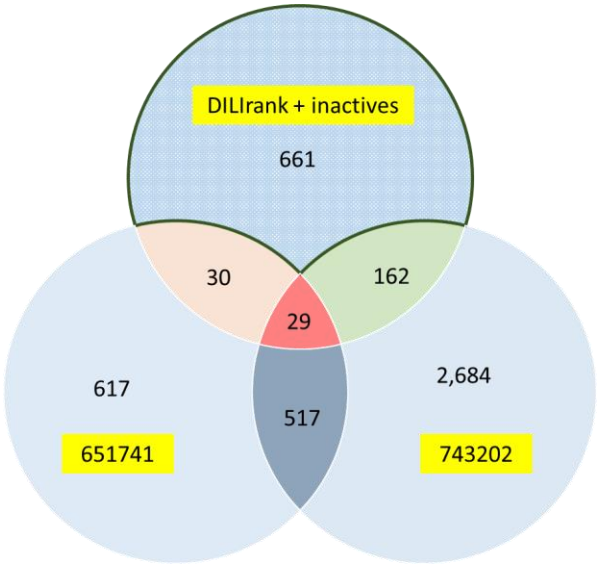

Figure 1. The number of compounds and overlap in three datasets.

Five types of chemical descriptors, including chemical fingerprints and molecular descriptors, were used in QSAR model development. Chemical fingerprints include Extended-connectivity fingerprints (ECFPs), functional-class fingerprints (FCFPs) (Rogers and Hahn, 2010), Molecular ACCess system (MACCS) keys (Durant et al., 2002). These three types of fingerprints were calculated using the RDKit (www.rdkit.org) package. Molecular descriptors include 200 RDKit molecular descriptors calculated using the RDKit package and 334 Dragon descriptors calculated using the commercial Dragon software v.6.0 (Talete s.r.l., Milano, Italy). All the molecular descriptor values were

normalized to the range from zero to one for the training set compounds before model development. Three types of machine learning approaches implemented by scikit-learn were used to develop QSAR models for each bioassay endpoint: k-Nearest Neighbors (kNN), Random Forest (RF), and Support Vector Machines (SVM). Individual regression models for each bioassay endpoint were developed using the combination of one type of descriptors (ECFP6, FCFP6, MACCS, rdkit) and one of the modeling approaches (kNN, RF, SVM), resulting in 12 individual models. The consensus QSAR model which was generated by averaging predictions of various individual models were also used in this study.(Ciallella et al., 2020; Golbraikh et al., 2017; Wang et al., 2015) All models were evaluated using a standard five-fold cross-validation procedure, with 20% of the training set compounds left out for testing purposes during each iteration, as described in previous studies.(Ciallella et al., 2020; Russo et al., 2019; Wang et al., 2015) Each bioassay training set was randomly split into five equal subsets, four subsets (80% of the total compounds) were used for model training, and the remaining 20% was used to test the resulted model. This procedure was repeated five times so that every compound was used for prediction once.

Another ARE assay with PubChem AID 651741 was used as the external validation set. After the above data curation and standardization, the dataset has 1,193 unique compounds (394 active and 799 inactive). Furthermore, compounds already in the ARE assay 743202 were excluded (**Figure 1**). The remaining 647 compounds were used for external validation. The results of sensitivity (Equation 1), specificity (Equation 2), and average correct classification ratio (CCR, Equation 3), when predicting the external validation set using individual and consensus models are shown in **Figure 2**. The two

model kNN-ECFP and kNN-FCFP6 showed poor model performance (sensitivity less than 0.3) and was not shown in **Figure 2**. After examining the probability prediction values, which ranged from [0, 1], we defined two consensus prediction thresholds (CPTs) to classify compounds as active or inactive (Kim, Huang, et al., 2016). CPT-1 (≥ 0.5 as active and < 0.5 as inactive), which is same as the default mode method, and CPT-2 (≥ 0.7 as active and ≤ 0.3 as inactive), where predictions between < 0.7 and > 0.3 were inconclusive. The consensus model generated under CPT-1 threshold showed better performance than individual models. Under stricter thresholds CPT-2, the performance of consensus model was further improved. However, this would decrease the coverage of the prediction compounds. Same with the strictest thresholds that only evaluate compounds showed consistent predictions among all the 13 models.

$$\text{sensitivity} = \frac{TP}{(TP+FN)} \qquad (1)$$

$$\text{specificity} = \frac{TN}{(TN+FN)} \qquad (2)$$

$$CCR = \frac{(Sensitivity+Specificity)}{2} \qquad (3)$$

Where TP represents the number of true positives (active compounds correctly predicted as active), FP represents the number of false positives (inactive compounds incorrectly predicted as active), TN represents the number of true negatives (inactive compounds correctly predicted as inactive), and FN represents the number of false negatives (active compounds incorrectly predicted as inactive).

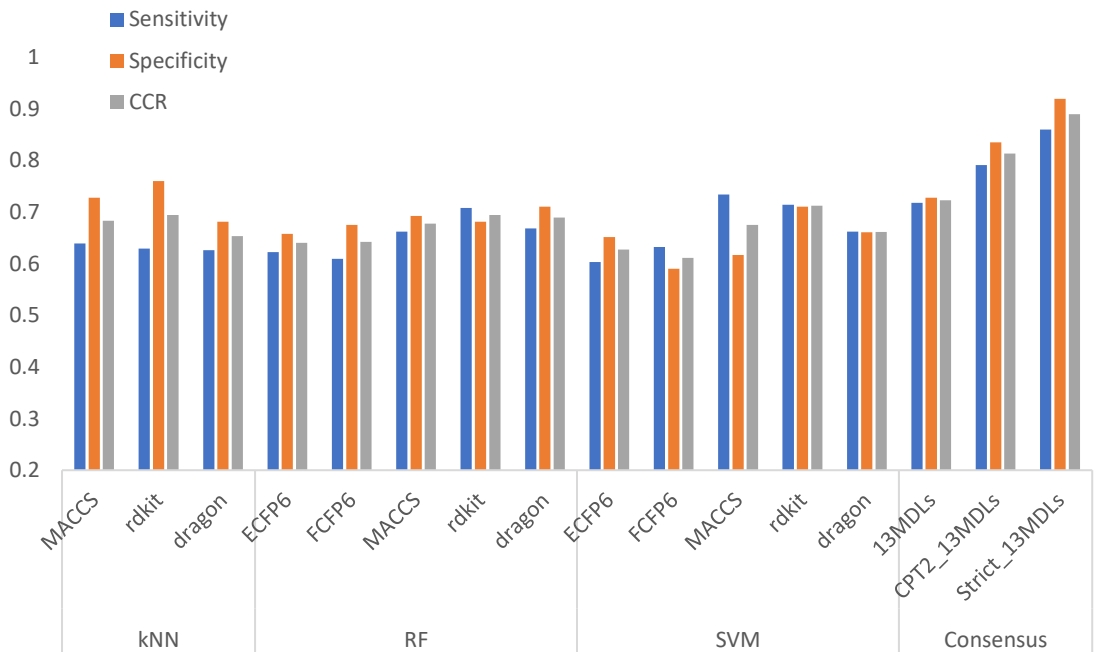

Figure 2. Sensitivity, specificity, and CCR results of predicting external validation set using individual and consensus models. 13MDLs: consensus model built using 13 combinatorial models under CPT-1 thresholds. CPT2_13MDLs: consensus model built using 13 combinatorial models under CPT-2 thresholds. Strict_13MDLs: only compounds showed consistent predictions among all the 13 models were used for evaluation.

***Results* - Profiling Target Compounds using Public Big Data Sources**

The balanced hepatotoxicity dataset consists of 882 compounds: 432 hepatotoxic and 450 non-toxic. These 882 compounds were used to search the PubChem portal for all the bioactivity responses. The result is the bioactivity response profile, so-called bioprofile. The initial bioprofile for hepatotoxicity compounds consisted of more than 100,000 PubChem bioassays, most of which were sparse, consisted of little data, and needed further curations. It is important to select critical bioassays based on their relationships to the drug compounds. The initial bioprofile is optimized by selecting assays that have at least five active responses and one inactive response across the drug compounds. The

optimized bioprofile has 544 assays and 715 compounds (358 hepatotoxic, 357 non-toxic) involved (**Figure 3**).

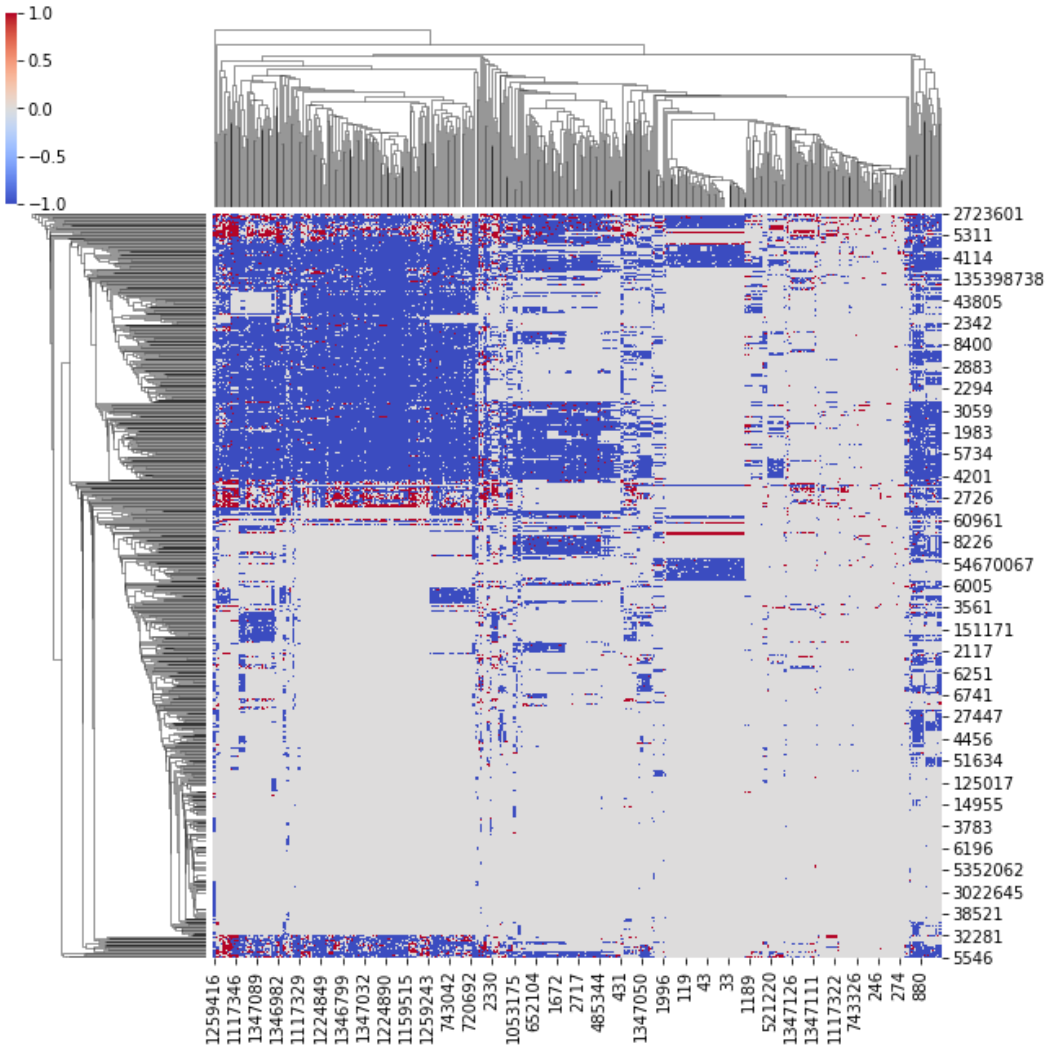

Figure 3. Optimized PubChem bioprofile. Only assays showed at least five active responses and at least one inactive response across the compounds of interest were selected. The resulting optimized bioprofile has 715 compounds and 544 related. Active results (1) were represented by red; inactive results (−1) were represented by blue, and inconclusive or untested results (0) were represented by gray.

***Results* – subspace clustering of PubChem *in vitro* assays.**

To identify potential toxicity mechanisms and further optimize the initial bioprofile, the 544 PubChem assays were clustered based on shared chemical fragments relevant to bioassay responses. To achieve this, we used the established ToxPrint fingerprints, a set

of 729 chemical fragments relevant to toxicity reported in a previous study (Yang et al., 2015). ToxPrint fingerprints were generated using ChemoTyper software version 1.0 (Molecular Networks GmbH, Erlangen, Germany). We also used saagar descriptors, another set of 834 extensible chemistry-aware substructures, as described in the literature (Sedykh et al., 2021). By constructing the contingency table of active/inactive response with the presence/absence of a fingerprint, and conduct Fisher's exact test, we determine whether an assay is correlated with a fingerprint. If the p-value of Fisher's exact test < 0.05, we think the assay is correlated with the fingerprint, and assign a 1 at the assay-fingerprint matrix. Details of this method are described in our previous study (Russo et al., 2019).

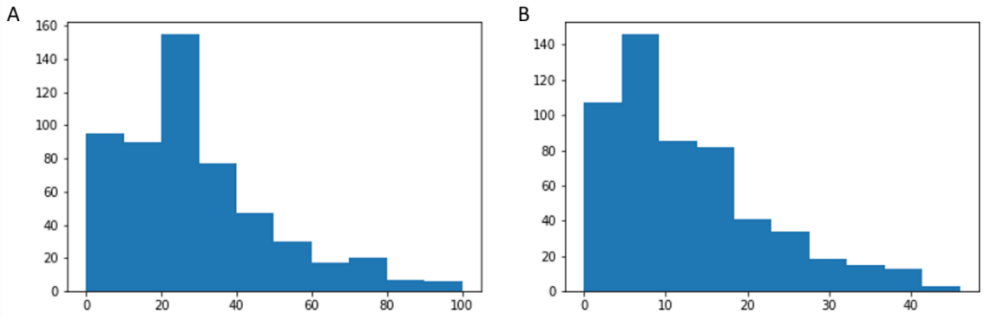

Figure 4. distributions of how many correlated fingerprints an assay has using A) saagar and B) ToxPrint fingerprints.

The distributions of how many correlated fingerprints an assay has are shown in **Figure 4**. For saagar fingerprint, there are 604 fragments that had at least one correlated assay, and 525 assays had at least one correlated fragment. For ToxPrint fingerprint, 301 fragments had at least one correlated assay, and 504 assays had at least one correlated fragment. With the assay-fingerprint matrix, we can calculate the Jaccard distances for every assay pairs. Use the assays as nodes and Jaccard distances as edges, we can construct a similarity map of assays and further cluster the assays into different

communities using the software Gephi version 0.9.1 (https://gephi.org/), as shown in **Figure 5**. Each assay cluster represents relevant biological data for compounds of interest that can be integrated into specific toxicity pathways. We got 36 clusters using saagar fingerprints, and 11 clusters using ToxPrint fingerprints. **Figure 6** shows the number of assays in each cluster using the two kinds of fingerprints. Among the 36 clusters identified using saagar fingerprints, nine clusters showed at least five assays in the same cluster. Among the 11 clusters identified using ToxPrint fingerprints, eight clusters showed at least five assays in the same cluster.

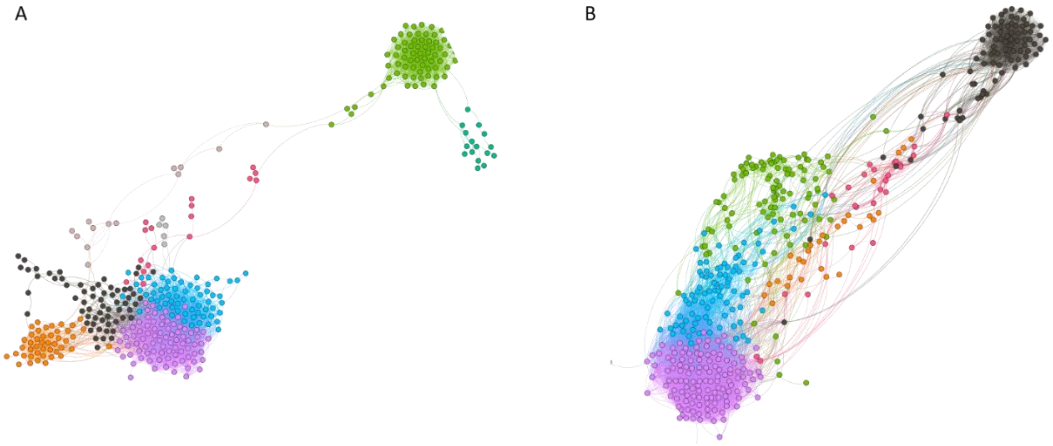

Figure 5. Subspace clustering of PubChem bioassays using the chemical-*in vitro* assay correlations. Nodes represent PubChem assays, edges represent Jaccard similarity computed based on A) saagar fingerprints B) ToxPrint fingerprints. Distinct clusters are visualized by color.

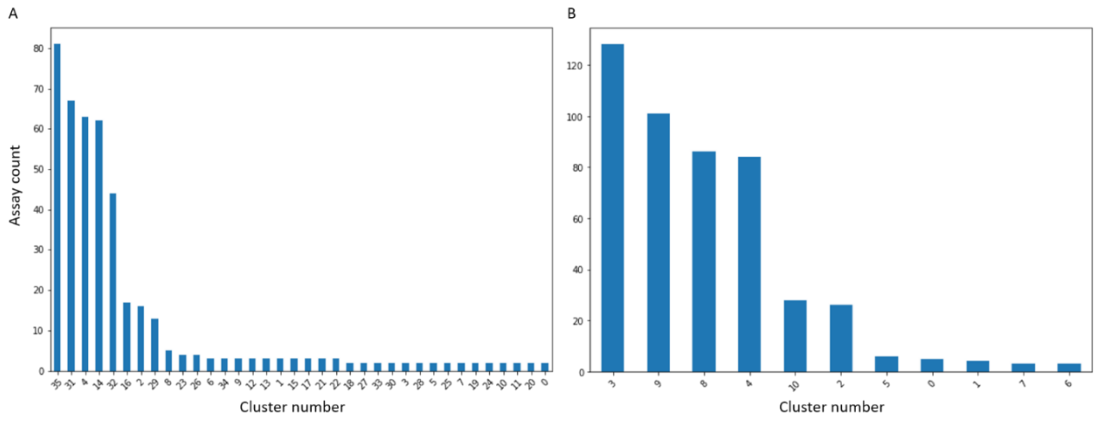

Figure 6. The number of assays in the same cluster for A) clusters identified using saagar fingerprints, B) clusters identified using ToxPrint fingerprints.

Since we are particularly interested in ARE signaling involved hepatotoxicity, we first see which cluster the ARE assay falls into. For clusters identified using saagar fingerprints, ARE assay falls into cluster#32, which has 44 assays. For clusters identified using ToxPrint fingerprints, ARE assay falls into cluster#9, which has 101 assays. After reading the assay description of the two clusters that contain the ARE assays, I classified the assays into different groups. In both cases, more than half of the assays are testing cell viability, anticancer activity, or a summary of the qHTS. The summary of the qHTS usually summarizes the results of primary HTS and cell viability counter screen and sometimes the auto-fluorescence. Other groups including nuclear receptor activation, CYP enzyme interaction, mitochondria activity, DNA damage, cell repair, and transcriptional alteration may be related to an outcome pathway lead to DILI. These two assay clusters represent relevant biological data for compounds of interest that can be integrated into oxidative stress involved DILI pathways.

Approach: Developing vAOP Models using Biosimilarity Search and Read-Across

The assay cluster containing the ARE assay represents relevant biological data for compounds of interest that can be integrated to ARE involved DILI pathways. As shown

in **Figure 3**, many compounds often do not have experimental results against all assays. Missing data can hamper the modeling and predicting procedure. Therefore, we will develop machine learning and deep learning models to fill in the missing data for new compounds under selected assays, like what we did in predicting ARE activation for untested hepatotoxicity compounds. Deep Neuron Network (DNN) is a popular deep learning approach with many layers, which consist of many nodes (Goh et al., 2017; LeCun et al., 2015). The outputs from the neurons of the previous layer serve as the inputs to the neurons in the next layer, creating a highly interconnected network. DNNs have been widely used in the field of speech and image recognition, drug discovery, and genomics data analysis (Zhang et al., 2017). Multi-task learning, which is based on DNNs, is a modeling approach that allows for multiple related tasks to be modeled simultaneously. Modeling several biologically related endpoints through multi-task learning has shown superior performance to traditional QSAR models by reducing overfitting, solving issues of biased data, and identifying variables from related tasks (Xu et al., 2017). In this project, the multi-task learning is suitable for model development and filling the missing biological data of assays in a cluster for target compounds. The input data will be the chemical descriptors of all compounds in the hepatotoxicity database and the DNN models will be developed simultaneously for several bioassays within a pathway to fill all the missing data. We will use standard five-fold cross validation techniques to determine the optimum neural network architecture, including the 1) number of hidden layers, 2) number of neurons per layer, and 3) activation function for neurons.

Our lab previously described a novel strategy of using the bioprofile-based read-across to predict chemical toxicity (Russo et al., 2019; Zhao et al., 2020). In a read-across study, the toxicity potential of a new compound will be evaluated by its most "similar" compound that has an experimental toxicity result (Ball et al., 2016). Traditional read-across is based on chemical similarity calculation, which has proved to be error-prone for predicting complex toxicity endpoints due to "activity cliffs" (ie, structurally similar compounds have different toxicity). The inclusion of biosimilarity rankings based on biological data adds extra strength. The biosimilarity between two compounds in a cluster *c* can be calculated by:

$$Biosimilarity\ (A,B)_c = \frac{|A_a \cap B_a| + |A_i \cap B_i| \cdot w}{|A_a \cap B_a| + |A_i \cap B_i| \cdot w + |A_a \cap B_i| + |A_i \cap B_a|}$$

Where A_a and B_a represent the sets of active responses, A_i and B_i represent the sets of inactive responses in PubChem bioassays within a cluster *c* for compounds A and B. The term *w* weights the inactive responses less than active responses since the proportion of active data, which indicates more significant chemical-biological interactions, is much lower than inactive data.

Approach: Developing vAOP Models using Knowledgebase DNNs Modeling

The drawback of the above biosimilarity search approach is that the linear similarity search does not draw distinction and mechanism association between assays. To rectify this, we will develop Knowledgebase DNNs (K-DNNs) inspired by the nature of complex biological systems and the combined effect of assays testing for protein interactions, gene activation, and growth inhibition on resulting adverse outcomes (e.g., hepatotoxicity) (**Figure 7**). The input of all compounds for K-DNNs modeling will be the ToxPrint or saagar fingerprints initially generated. This chemical information will be fed

to the assays in the second layer of the network system, which in turn are connected to the bioassays in the third layer and then induce the organism toxicity. The numbers on the edges indicate the weights between connections and the labels on the neurons (active/inactive) indicate the status of the activation functions as a result of training. As shown in **Figure 7A**, the path connecting Fragment 1 to Bioassay 1 and then to Bioassay 3 represents a potential vAOP that predicts organism toxicity (i.e., hepatotoxicity). By comparison, Fragment 2 and Bioassay 2 are determined to be irrelevant (Figure 6a). "Activation" of a neuron in the K-DNN is dependent on the observed outcome of a compound in that assay. In this manner, during the training process, the connections between neurons can offer insights into perturbed biological pathways contributing to hepatotoxicity. **Figure 7B** shows a vAOP model based on preliminary data, that includes the initial chemical fragment structure, a protein-binding assay, a cell stress assay, a cytotoxicity assay, and four other "off-target" assays (two protein-related and two cell stress-related). The connections between neurons have been optimized to show the predicated vAOP highlighted in red.

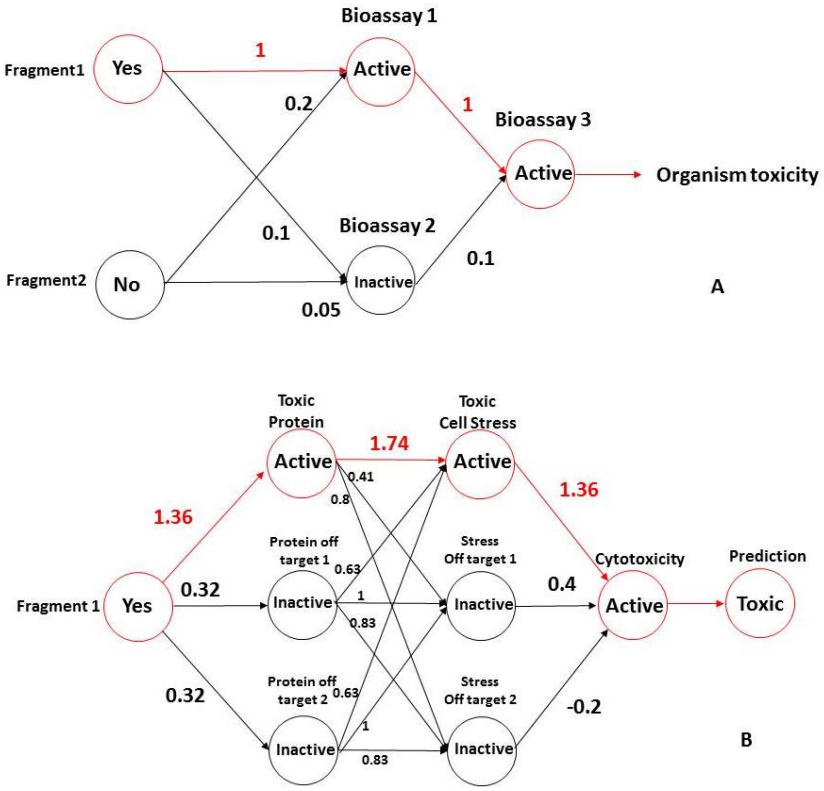

Figure 7. K-DNN modeling. **A)** The concept of K-DNN modeling; **B)** A representative vAOP model resulting from K-DNN modeling.

Compilation and organization of a large amount of high-quality data are key for proper use for modeling. Besides *in vitro* assays data, we also plan to integrate other types of data like toxicogenomic data and dosage data into the vAOP modeling process. Toxicogenomic data from CTD or L1000 project could yield novel insights into toxicity mechanisms. Drug dosage is also reported as an important feature in DILI (Chen et al., 2013; Xu et al., 2015). Overall, we expect to develop predictive vAOP models based on the combined conventional chemical descriptors, bioprofiles, and toxicogenomic information through vAOP modeling workflow.

Section III. Share the datasets and modeling framework via a web portal

The goal of this project is to make use of currently available *in vitro* data and provide methodologies and models that can predict more complicated hepatotoxicity. This will greatly advance the field of alternative methods to animal testing. The data collected (Section I) and models developed (Section II) in this study will be made be accessible through an in-house web portal (http://ciipro.rutgers.edu/). The workflows and tools can be adapted accordingly by chemists, biologists, toxicologists, and information scientists to model other complex animal toxicity endpoints.

A key step after the development of the vAOP model is to validate the model predictivity using an external dataset. Drug compounds with known hepatotoxicity activities and not in the modeling set will be collected from literature or LiverTox website. And it is possible that some external test compounds may not have sufficient bioassay data that needs for vAOP model prediction. This issue can be partially solved by building computational models to predict *in vitro* assay activity (e.g., by QSAR modeling using machine learning or multi-task learning). This process undoubtedly introduces some additional uncertainty into the prediction. External compounds that have little bioactivity data or are very dissimilar to our modeling compounds need to be subject to applicability domain assessments to eliminate possible unreliable predictions. Furthermore, our preliminary data demonstrate that the potential prediction errors of models in this step can be corrected by further experimental validation at a low cost.

The ultimate goal of this project is to make use of currently available *in vitro* data and develop methodologies and models that can predict more complicated toxicity endpoints. The Chemical In Vitro-In Vivo Profiling (CIIPro) portal (http://ciipro.rutgers.edu/) was

designed in our lab to disseminate the models that we develop to the toxicology research community. The CIIPro portal will be the final deliverable of this project and will provide users with access to the hepatotoxicity database, bioprofiles, and the vAOP models developed under **Section I** and **II**. By processing the data and the models in a web portal rather than a toolkit, research groups and communities will be able to obtain real-time results that can be readily shared with and accessed by other groups. Most importantly, the CIIPro portal will greatly save resources by reducing the use of animals in toxicity testing and provide toxicologists worldwide with a computational tool to evaluate the risk of toxicity for new compounds.

EVALUATION PLAN OF SCIENTIFIC RESULTS

The one common ground to all computational modeling approaches is obtaining standardized and curated high-quality data in large quantity enough to ensure predictivity and accuracy. Publicly available data are often not standardized, and some compounds may have different results among different data sources, especially for hepatotoxicity classification. For example, desmopressin and ethotoin (PubChem CID 27991, 3292) were classified as non-hepatotoxic by multiple research papers but were classified as ambiguous-DILI-concern in the DILIrank dataset because of liver injury reports without confirmed causality. Such compounds will be carefully examined and excluded from further computational modeling. Besides that, we have developed methods to detect structural errors, structural duplicates, and standardization of chemical representations. To evaluate the significance of the resulting models developed in this project, we will compare the results to standard machine learning algorithms. All models will undergo rigorous technical testing, such as Y-randomization testing, to ensure robustness. The biological relevance of the vAOP models will be validated by experimental testing of compounds not included in the model training set due to a lack of assay testing results in online repositories. If toxicants for the specified endpoint, containing the identified MIEs, show active responses in the identified assays, then the vAOP models will be proven valid.

REFERENCE

Andrade, R. J., Chalasani, N., Bjornsson, E. S., Suzuki, A., Kullak-Ublick, G. A., Watkins, P. B., . . . Aithal, G. P. (2019). Drug-induced liver injury. *Nat Rev Dis Primers, 5*(1), 58. doi:10.1038/s41572-019-0105-0

Attene-Ramos, M. S., Miller, N., Huang, R., Michael, S., Itkin, M., Kavlock, R. J., . . . Xia, M. (2013). The Tox21 robotic platform for the assessment of environmental chemicals--from vision to reality. *Drug Discov Today, 18*(15-16), 716-723. doi:10.1016/j.drudis.2013.05.015

Bale, S. S., Vernetti, L., Senutovitch, N., Jindal, R., Hegde, M., Gough, A., . . . Yarmush, M. L. (2014). In vitro platforms for evaluating liver toxicity. *Exp Biol Med (Maywood), 239*(9), 1180-1191. doi:10.1177/1535370214531872

Ball, N., Cronin, M. T., Shen, J., Blackburn, K., Booth, E. D., Bouhifd, M., . . . Hartung, T. (2016). Toward Good Read-Across Practice (GRAP) guidance. *ALTEX, 33*(2), 149-166. doi:10.14573/altex.1601251

Betts, K. S. (2013). Tox21 to date: steps toward modernizing human hazard characterization. *Environ Health Perspect, 121*(7), A228. doi:10.1289/ehp.121-a228

Chen, M., Borlak, J., & Tong, W. (2013). High lipophilicity and high daily dose of oral medications are associated with significant risk for drug-induced liver injury. *Hepatology, 58*(1), 388-396. doi:10.1002/hep.26208

Chen, M., Suzuki, A., Thakkar, S., Yu, K., Hu, C., & Tong, W. (2016). DILIrank: the largest reference drug list ranked by the risk for developing drug-induced liver injury in humans. *Drug Discov Today, 21*(4), 648-653. doi:10.1016/j.drudis.2016.02.015

Chen, M., Vijay, V., Shi, Q., Liu, Z., Fang, H., & Tong, W. (2011). FDA-approved drug labeling for the study of drug-induced liver injury. *Drug Discov Today, 16*(15-16), 697-703. doi:10.1016/j.drudis.2011.05.007

Ciallella, H. L., Russo, D. P., Aleksunes, L. M., Grimm, F. A., & Zhu, H. (2020). Predictive modeling of estrogen receptor agonism, antagonism, and binding activities using machine- and deep-learning approaches. *Lab Invest*. doi:10.1038/s41374-020-00477-2

Durant, J. L., Leland, B. A., Henry, D. R., & Nourse, J. G. (2002). Reoptimization of MDL keys for use in drug discovery. *J Chem Inf Comput Sci, 42*(6), 1273-1280. doi:10.1021/ci010132r

Ekins, S., Williams, A. J., & Xu, J. J. (2010). A predictive ligand-based Bayesian model for human drug-induced liver injury. *Drug Metab Dispos, 38*(12), 2302-2308. doi:10.1124/dmd.110.035113

Fourches, D., Barnes, J. C., Day, N. C., Bradley, P., Reed, J. Z., & Tropsha, A. (2010). Cheminformatics analysis of assertions mined from literature that describe drug-induced liver injury in different species. *Chem Res Toxicol, 23*(1), 171-183. doi:10.1021/tx900326k

Gadaleta, D., Manganelli, S., Roncaglioni, A., Toma, C., Benfenati, E., & Mombelli, E. (2018). QSAR Modeling of ToxCast Assays Relevant to the Molecular Initiating Events of AOPs Leading to Hepatic Steatosis. *J Chem Inf Model, 58*(8), 1501-1517. doi:10.1021/acs.jcim.8b00297

Gibb, S. (2008). Toxicity testing in the 21st century: a vision and a strategy. *Reprod Toxicol, 25*(1), 136-138. doi:10.1016/j.reprotox.2007.10.013

Gijbels, E., Vilas-Boas, V., Annaert, P., Vanhaecke, T., Devisscher, L., & Vinken, M. (2020). Robustness testing and optimization of an adverse outcome pathway on cholestatic liver injury. *Arch Toxicol, 94*(4), 1151-1172. doi:10.1007/s00204-020-02691-9

Gijbels, E., & Vinken, M. (2019). Mechanisms of Drug-Induced Cholestasis. *Methods Mol Biol, 1981*, 1-14. doi:10.1007/978-1-4939-9420-51

Goh, G. B., Hodas, N. O., & Vishnu, A. (2017). Deep learning for computational chemistry. *J Comput Chem, 38*(16), 1291-1307. doi:10.1002/jcc.24764

Golbraikh, A., Wang, X. S., Zhu, H., & Tropsha, A. (2017). Predictive QSAR Modeling: Methods and Applications in Drug Discovery and Chemical Risk Assessment. In *Handbook of Computational Chemistry* (pp. 2303-2340). doi:10.1007/978-3-319-27282-5_37

Hoofnagle, J. H., Serrano, J., Knoben, J. E., & Navarro, V. J. (2013). LiverTox: a website on drug-induced liver injury. *Hepatology, 57*(3), 873-874. doi:10.1002/hep.26175

Judson, R. S., Houck, K. A., Kavlock, R. J., Knudsen, T. B., Martin, M. T., Mortensen, H. M., . . . Dix, D. J. (2010). In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. *Environ Health Perspect, 118*(4), 485-492. doi:10.1289/ehp.0901392

Judson, R. S., Martin, M. T., Egeghy, P., Gangwal, S., Reif, D. M., Kothiya, P., . . . Richard, A. M. (2012). Aggregating data for computational toxicology applications: The U.S. Environmental Protection Agency (EPA) Aggregated Computational Toxicology Resource (ACToR) System. *Int J Mol Sci, 13*(2), 1805-1831. doi:10.3390/ijms13021805

Kavlock, R., Chandler, K., Houck, K., Hunter, S., Judson, R., Kleinstreuer, N., . . . Dix, D. (2012). Update on EPA's ToxCast program: providing high throughput decision support tools for chemical risk management. *Chem Res Toxicol, 25*(7), 1287-1302. doi:10.1021/tx3000939

Kim, M. T., Huang, R., Sedykh, A., Wang, W., Xia, M., & Zhu, H. (2016). Mechanism Profiling of Hepatotoxicity Caused by Oxidative Stress Using Antioxidant Response Element Reporter Gene Assay Models and Big Data. *Environ Health Perspect, 124*(5), 634-641. doi:10.1289/ehp.1509763

Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., . . . Bryant, S. H. (2016a). PubChem Substance and Compound databases. *Nucleic Acids Res., 44*(D1), D1202-D1213. doi:10.1093/nar/gkv951

Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., . . . Bryant, S. H. (2016b). PubChem Substance and Compound databases. *Nucleic Acids Res, 44*(D1), D1202-1213. doi:10.1093/nar/gkv951

Kotsampasakou, E., & Ecker, G. F. (2017). Predicting Drug-Induced Cholestasis with the Help of Hepatic Transporters-An in Silico Modeling Approach. *J Chem Inf Model, 57*(3), 608-615. doi:10.1021/acs.jcim.6b00518

Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., . . . Golub, T. R. (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science, 313*(5795), 1929-1935. doi:10.1126/science.1132939

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444. doi:10.1038/nature14539

Li, T., Tong, W., Roberts, R., Liu, Z., & Thakkar, S. (2021). DeepDILI: Deep Learning-Powered Drug-Induced Liver Injury Prediction Using Model-Level Representation. *Chem Res Toxicol, 34*(2), 550-565. doi:10.1021/acs.chemrestox.0c00374

Liew, C. Y., Lim, Y. C., & Yap, C. W. (2011). Mixed learning algorithms and features ensemble in hepatotoxicity prediction. *J Comput Aided Mol Des, 25*(9), 855-871. doi:10.1007/s10822-011-9468-3

Liu, R., Yu, X., & Wallqvist, A. (2015). Data-driven identification of structural alerts for mitigating the risk of drug-induced human liver injuries. *J Cheminform, 7*, 4. doi:10.1186/s13321-015-0053-y

Ma, Q. (2013). Role of nrf2 in oxidative stress and toxicity. *Annu Rev Pharmacol Toxicol, 53*, 401-426. doi:10.1146/annurev-pharmtox-011112-140320

Mattingly, C. J., Colby, G. T., Forrest, J. N., & Boyer, J. L. (2003). The Comparative Toxicogenomics Database (CTD). *Environ Health Perspect, 111*(6), 793-795. doi:10.1289/ehp.6028

Mattingly, C. J., Colby, G. T., Rosenstein, M. C., Forrest, J. N., Jr., & Boyer, J. L. (2004). Promoting comparative molecular studies in environmental health research: an overview of the comparative toxicogenomics database (CTD). *Pharmacogenomics J, 4*(1), 5-8. doi:10.1038/sj.tpj.6500225

Mattingly, C. J., Rosenstein, M. C., Davis, A. P., Colby, G. T., Forrest, J. N., Jr., & Boyer, J. L. (2006). The comparative toxicogenomics database: a cross-species resource for building chemical-gene interaction networks. *Toxicol Sci, 92*(2), 587-595. doi:10.1093/toxsci/kfl008

Mulliner, D., Schmidt, F., Stolte, M., Spirkl, H. P., Czich, A., & Amberg, A. (2016). Computational Models for Human and Animal Hepatotoxicity with a Global Application Scope. *Chem Res Toxicol, 29*(5), 757-767. doi:10.1021/acs.chemrestox.5b00465

Park, B. K., Boobis, A., Clarke, S., Goldring, C. E., Jones, D., Kenna, J. G., . . . Baillie, T. A. (2011). Managing the challenge of chemically reactive metabolites in drug development. *Nat Rev Drug Discov, 10*(4), 292-306. doi:10.1038/nrd3408

Rogers, D., & Hahn, M. (2010). Extended-connectivity fingerprints. *J Chem Inf Model, 50*(5), 742-754. doi:10.1021/ci100050t

Russo, D. P., Strickland, J., Karmaus, A. L., Wang, W., Shende, S., Hartung, T., . . . Zhu, H. (2019). Nonanimal Models for Acute Toxicity Evaluations: Applying Data-Driven Profiling and Read-Across. *Environ Health Perspect, 127*(4), 47001. doi:10.1289/EHP3614

Sedykh, A. Y., Shah, R. R., Kleinstreuer, N. C., Auerbach, S. S., & Gombar, V. K. (2021). Saagar-A New, Extensible Set of Molecular Substructures for QSAR/QSPR and Read-Across Predictions. *Chem Res Toxicol, 34*(2), 634-640. doi:10.1021/acs.chemrestox.0c00464

Shukla, S. J., Huang, R., Simmons, S. O., Tice, R. R., Witt, K. L., Vanleer, D., . . . Xia, M. (2012). Profiling environmental chemicals for activity in the antioxidant response element signaling pathway using a high throughput screening approach. *Environ Health Perspect, 120*(8), 1150-1156. doi:10.1289/ehp.1104709

Stevens, J. L., & Baker, T. K. (2009). The future of drug safety testing: expanding the view and narrowing the focus. *Drug Discov Today, 14*(3-4), 162-167. doi:10.1016/j.drudis.2008.11.009

Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natoli, T. E., Lu, X., . . . Golub, T. R. (2017). A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell, 171*(6), 1437-1452 e1417. doi:10.1016/j.cell.2017.10.049

Tice, R. R., Austin, C. P., Kavlock, R. J., & Bucher, J. R. (2013). Improving the human hazard characterization of chemicals: a Tox21 update. *Environ Health Perspect, 121*(7), 756-765. doi:10.1289/ehp.1205784

Vinken, M. (2018). In vitro prediction of drug-induced cholestatic liver injury: a challenge for the toxicologist. *Arch Toxicol, 92*(5), 1909-1912. doi:10.1007/s00204-018-2201-4

Vorrink, S. U., Zhou, Y., Ingelman-Sundberg, M., & Lauschke, V. M. (2018). Prediction of Drug-Induced Hepatotoxicity Using Long-Term Stable Primary Hepatic 3D Spheroid Cultures in Chemically Defined Conditions. *Toxicol Sci, 163*(2), 655-665. doi:10.1093/toxsci/kfy058

Wang, W., Kim, M. T., Sedykh, A., & Zhu, H. (2015). Developing Enhanced Blood-Brain Barrier Permeability Models: Integrating External Bio-Assay Data in QSAR Modeling. *Pharm Res, 32*(9), 3055-3065. doi:10.1007/s11095-015-1687-1

Wang, Y., Bryant, S. H., Cheng, T., Wang, J., Gindulyte, A., Shoemaker, B. A., . . . Zhang, J. (2017a). PubChem BioAssay: 2017 update. *Nucleic Acids Res., 45*(D1), D955-D963. doi:10.1093/nar/gkw1118

Wang, Y., Bryant, S. H., Cheng, T., Wang, J., Gindulyte, A., Shoemaker, B. A., . . . Zhang, J. (2017b). PubChem BioAssay: 2017 update. *Nucleic Acids Res, 45*(D1), D955-D963. doi:10.1093/nar/gkw1118

Xu, Y., Dai, Z., Chen, F., Gao, S., Pei, J., & Lai, L. (2015). Deep Learning for Drug-Induced Liver Injury. *J Chem Inf Model, 55*(10), 2085-2093. doi:10.1021/acs.jcim.5b00238

Xu, Y., Pei, J., & Lai, L. (2017). Deep Learning Based Regression and Multiclass Models for Acute Oral Toxicity Prediction with Automatic Chemical Feature Extraction. *J Chem Inf Model, 57*(11), 2672-2685. doi:10.1021/acs.jcim.7b00244

Yang, C., Tarkhov, A., Marusczyk, J., Bienfait, B., Gasteiger, J., Kleinoeder, T., . . . Rathman, J. (2015). New publicly available chemical query language, CSRML, to support chemotype representations for application to data mining and modeling. *J Chem Inf Model, 55*(3), 510-528. doi:10.1021/ci500667v

Zhang, L., Tan, J., Han, D., & Zhu, H. (2017). From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug Discov Today, 22*(11), 1680-1685. doi:10.1016/j.drudis.2017.08.010

Zhao, L., Russo, D. P., Wang, W., Aleksunes, L. M., & Zhu, H. (2020). Mechanism-Driven Read-Across of Chemical Hepatotoxicants Based on Chemical Structures and Biological Data. *Toxicol Sci, 174*(2), 178-188. doi:10.1093/toxsci/kfaa005

Zhu, H. (2020). Big Data and Artificial Intelligence Modeling for Drug Discovery. *Annu. Rev. Pharmacol. Toxicol., 60*, 573-589. doi:10.1146/annurev-pharmtox-010919-023324