United States Environmental Protection Agency

Predicting hepatotoxicity using ToxCast *in vitro* bioactivity and chemical structure

Jie Liu

*UALR/UAMS Joint Bioinformatics Program
‡National Center for Computational Toxicology, US EPA
†Oak Ridge Institute for Science and Education (ORISE)

MCBIOS
Mar 13, 2015

Office of Research and Development
National Center for Computational Toxicology
www.epa.gov/ncct

Background

- Over 10,000 chemicals are currently in commercial use, of which only a small fraction of chemicals have been adequately assessed for potential hazard.
- Humans are exposed to over 6,000 environmental chemicals.
- The liver is usually the first site of chemical-induced toxicity in animal studies.
- Evaluating the risk of liver toxicity due to xenobiotics is critical for protecting public health.

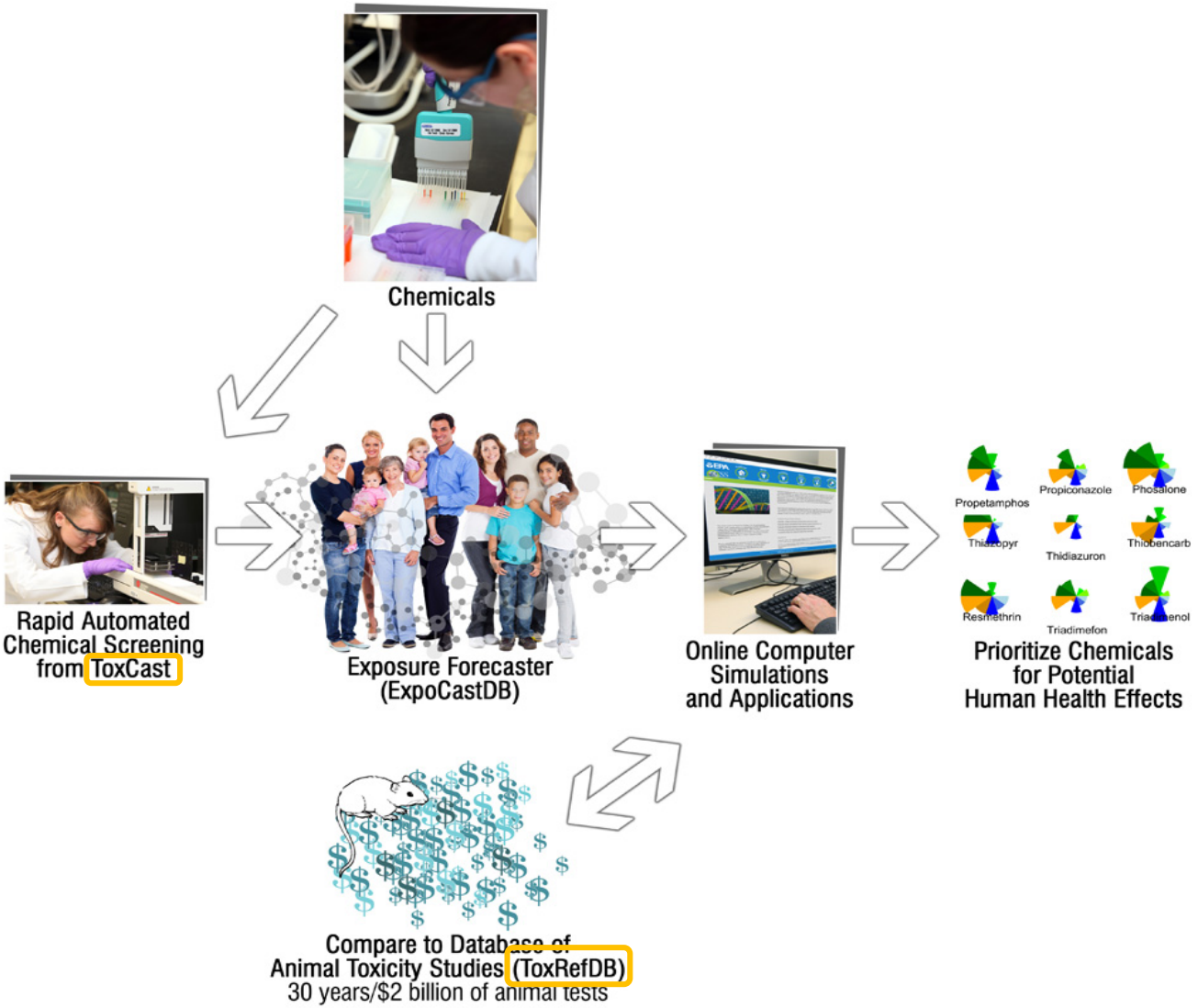
Toxicity Testing Challenges

- Animal testing
  - cost, time, animal welfare
  - European Union Cosmetics Directive banned animal testing in cosmetic products in 2013.*
- *In vitro* toxicity testing
  - US National Research Council (NRC, 2007) report envisions a future change in toxicity testing from use of laboratory animals to *in vitro* methods using human-relevant cells or tissues.**

* Raunio, Hannu. (2011) In Silico Toxicology – Non-Testing Methods. Front Pharmacol., 2, 33.

**National Research Council (2007) *Toxicity Testing in the 21st Century: A Vision and a Strategy.* The National Academies Press, Washington, DC.

Computational Toxicology



http://www.epa.gov/ncct/toxcast/

Data sources

Data	Source	# Chemicals
Histopathological effects from animal testing studies	ToxRefDB	1014
HTS assay results	ToxCast Phases I & II	1068
Chemical structure descriptors	QikProp, OpenBabel, PaDEL, PubChem	903

Toxicity Reference Database (ToxRefDB)

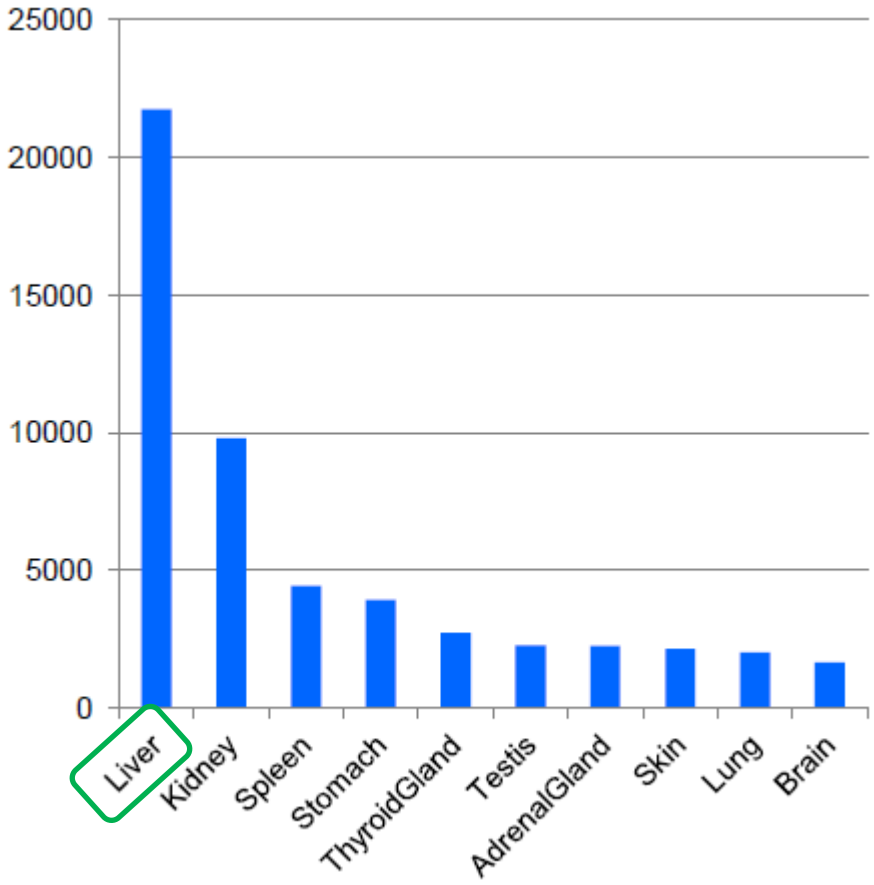
- Captures over 30 years of animal testing results (~ $2 billion).
- Contains more than 6,000 studies including National Toxicology Program (NTP), public literature and pharmaceutical studies.
- Covers 1,014 chemicals (version Aug. 2014)
- Publicly available (http://www.epa.gov/ncct/toxcast/data.html)

chemical_casrn	chemical_name	guideline_name	study_type	species	effect_target	effect_desc
103-33-3	Azobenzene	Carcinogenicity	CHR	rat	Liver	Hemosiderosis
104-76-7	2-Ethyl-1-hexanol	Carcinogenicity	CHR	rat	Liver	Congestion
104-76-7	2-Ethyl-1-hexanol	Carcinogenicity	CHR	rat	Liver	Relative to Body Weight
106-93-4	1,2-Dibromoethane	Carcinogenicity	CHR	rat	Liver	Carcinoma
…	…	…	…	…	…	…

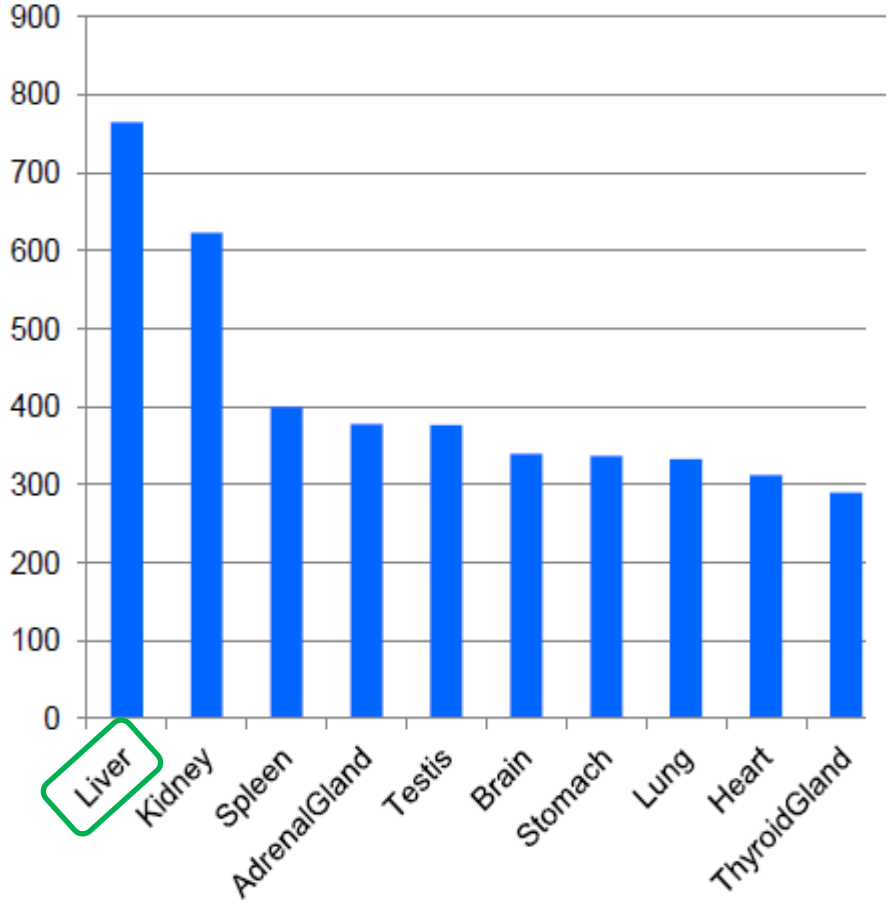
ToxRefDB *in vivo* Data Overview

United States Environmental Protection Agency

Effect Counts for Top 10 Organs



Chemical Counts for Top 10 Organs



K McLaurin, S Sneed, I Thillainadarajah, L Truong, M Martin. Expansion of Toxicity Reference Database (ToxRefDB) Study Data for In Vivo Phenotypic Modeling. Poster, SOT2013.

Toxicity Forecaster (ToxCast)

ToxCast phases I and II: 1,068 chemicals

- more than 800 HTS assay endpoints
- publicly available (http://www.epa.gov/ncct/toxcast/data.html, version Nov. 2014)

Set	Chemicals	Assays	Endpoints	Completion	Available
ToxCast Phase I	309	~600	~700	2011	Now
ToxCast Phase II	776	~600	~800	03/2013	Now
ToxCast Phase IIIa	1001	~100	~100	Ongoing	Ongoing

ToxCast HTS Assays

United States Environmental Protection Agency

Biochemical Assays

- Protein families
  - GPCR
  - NR
  - Kinase
  - Phosphatase
  - Protease
  - Other enzyme
  - Ion channel
  - Transporter
- Assay formats
  - Radioligand binding
  - Enzyme activity
  - Co-activator recruitment

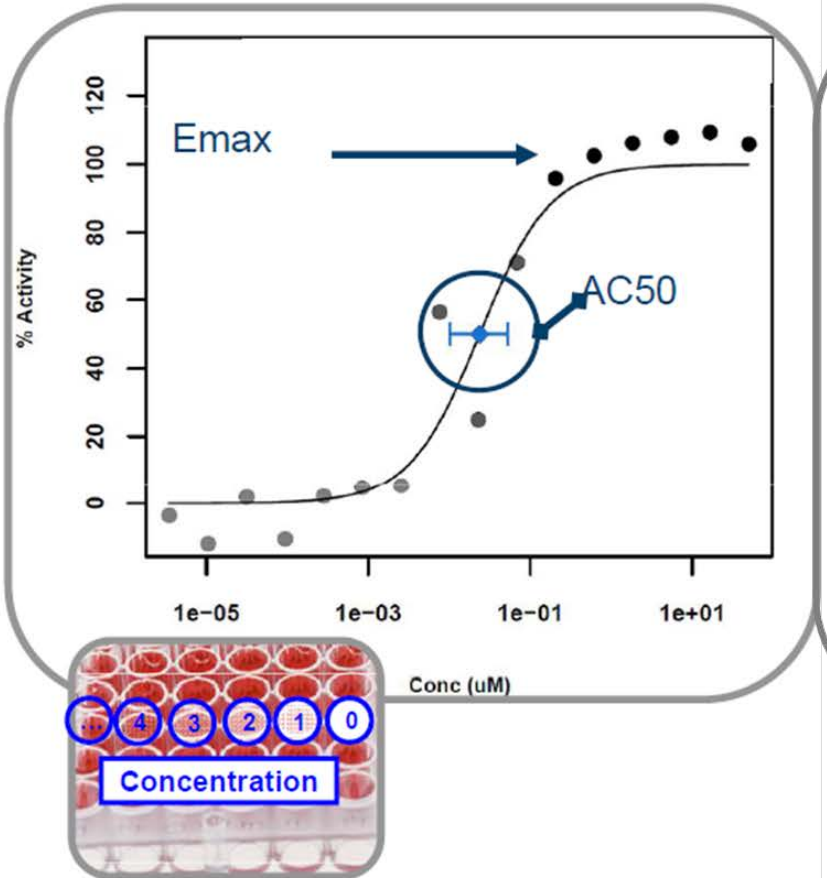
Cellular Assays

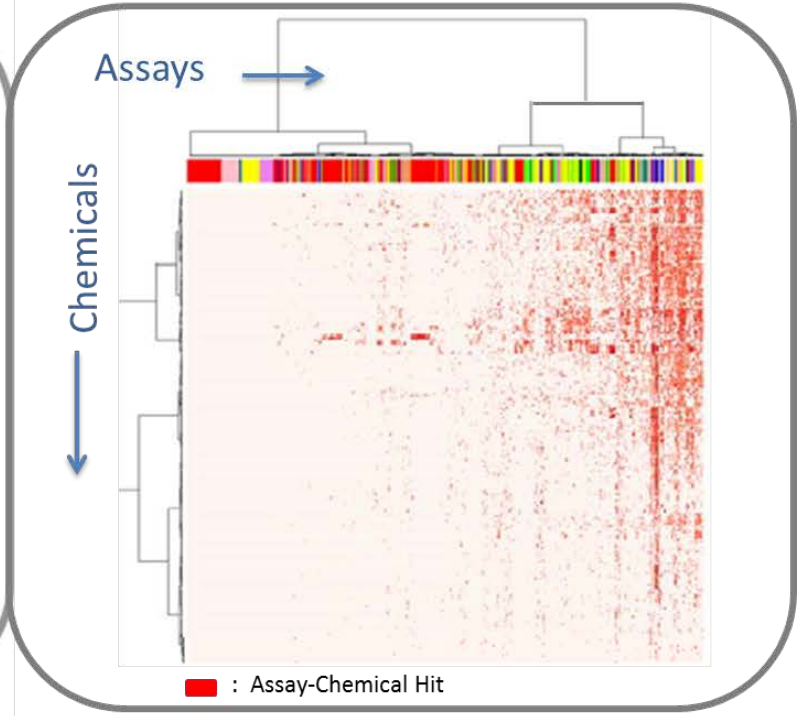
- Cell lines
  - HepG2 human hepatoblastoma
  - A549 human lung carcinoma
  - HEK 293 human embryonic kidney
- Primary cells
  - Human endothelial cells
  - Human monocytes
  - Human keratinocytes
  - Human fibroblasts
  - Human proximal tubule kidney cells
  - Human small airway epithelial cells
  - Rat hepatocytes
  - Mouse embryonic stem cells (Sid Hunter)
- Biotransformation competent cells
  - Primary rat hepatocytes
  - Primary human hepatocytes
- Assay formats
  - Cytotoxicity
  - Reporter gene
  - Gene expression

13

Judson RS, Houck KA, Kavlock RJ, Knudsen TB, Martin MT, Mortensen HM, et al. (2010). In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. Environ Health Perspect. 118, 485–492.

ToxCast Data Analysis

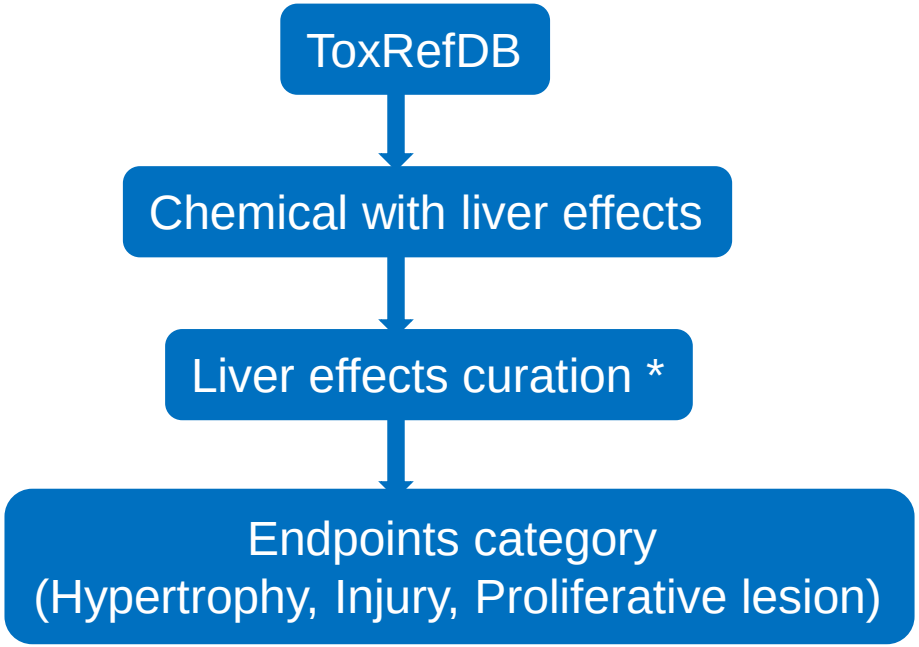




AC50 : chemical concentration (micromolar) at half maximal efficacy.

Judson RS, Houck KA, Kavlock RJ, Knudsen TB, Martin MT, Mortensen HM, et al. (2010). In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. Environ Health Perspect. 118, 485–492.

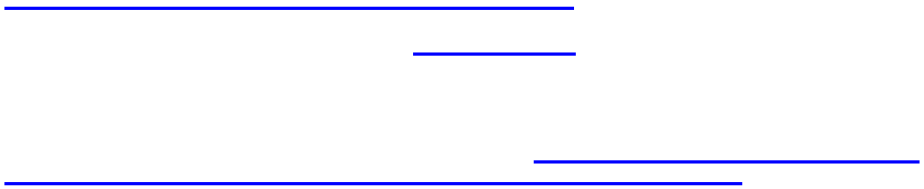
Liver effects curation



Effect	Category
Absolute	Hypertrophy
Accentuated Lobular Pattern	Injury
Adenocarcinoma	Proliferative lesion
Adenoma	Proliferative lesion
Adenoma/Carcinoma Combined	Proliferative lesion
Angiectasis	Injury
Apoptosis	Injury
areas of collapse	Injury
Arteritis	Injury
Atrophy	Injury
…	…

* Thoolen B, Maronpot RR, Harada T, …, Ward JM. Proliferative and nonproliferative lesions of the rat and mouse hepatobiliarysystem. Toxicol Pathol. 2010; 38(7 Suppl):5S-81S.

Predicting Rat Chronic Hepatotoxicity



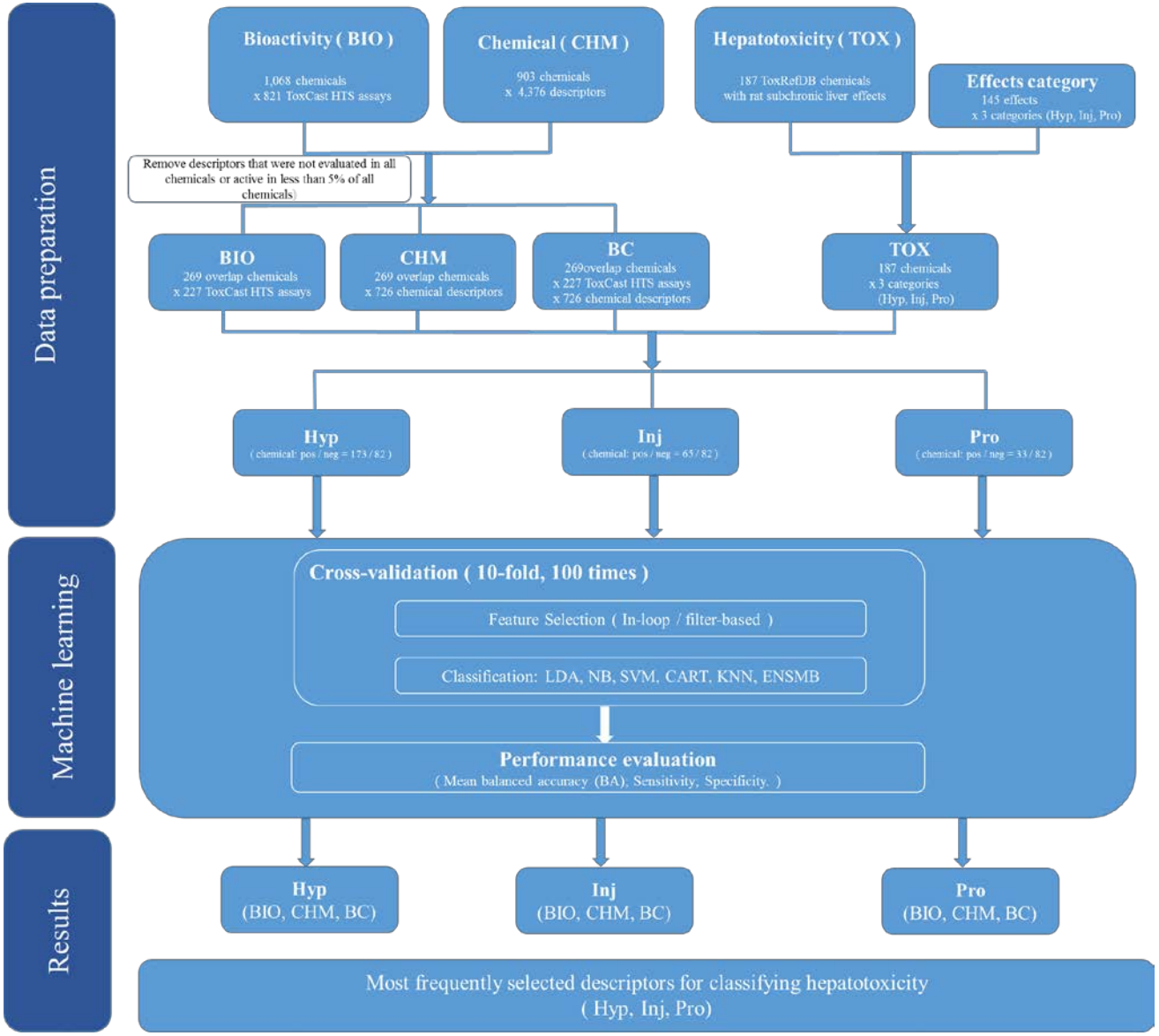
Supervised Machine Learning

- Data sets (**Rat Subchronic**):

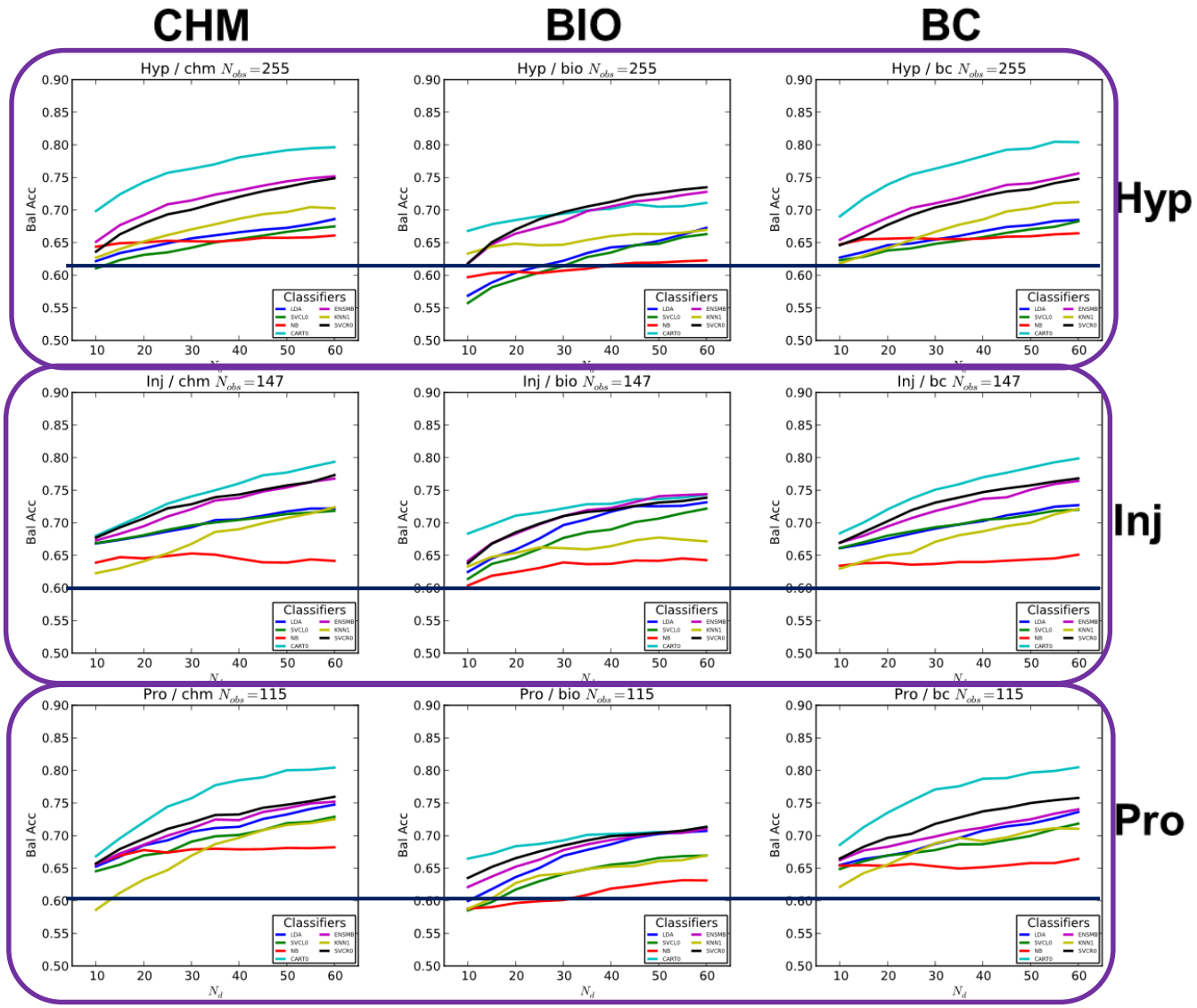
Data sets	Total chemicals	Hypertrophy	Injury	Proliferative lesions	Negative set	Descriptors
Bioactivity	269	173	–	–	82	227 ToxCast HTS assay endpoints
		–	65	–	82	
		–	–	33	82	
Chemical	269	173	–	–	82	726 chemical structure descriptors
		–	65	–	82	
		–	–	33	82	
Bioactivity & Chemical	269	173	–	–	82	227 ToxCast HTS assay endpoints & 726 chemical structure descriptors
		–	65	–	82	
		–	–	33	82	

- Feature (X / inputs)
  - Bioactivity descriptors
  - Chemical structure descriptors
  - Bioactivity and Chemical structure descriptors
- Class labels (Y, outputs)
  - Hypertrophy
  - Injury
  - Proliferative lesions

Workflow for the whole classification process



Classification Performance Results



Hyp: hypertrophy;

Inj: injury;

Pro: proliferative lesions;

BIO: bioactivity descriptors;

CHM: chemical structure descriptors;

BC: bioactivity & chemical structure descriptors;

BA: balanced accuracy;

Bal Acc: balanced accuracy;

Desc: descriptors;

N_d : number of descriptors;

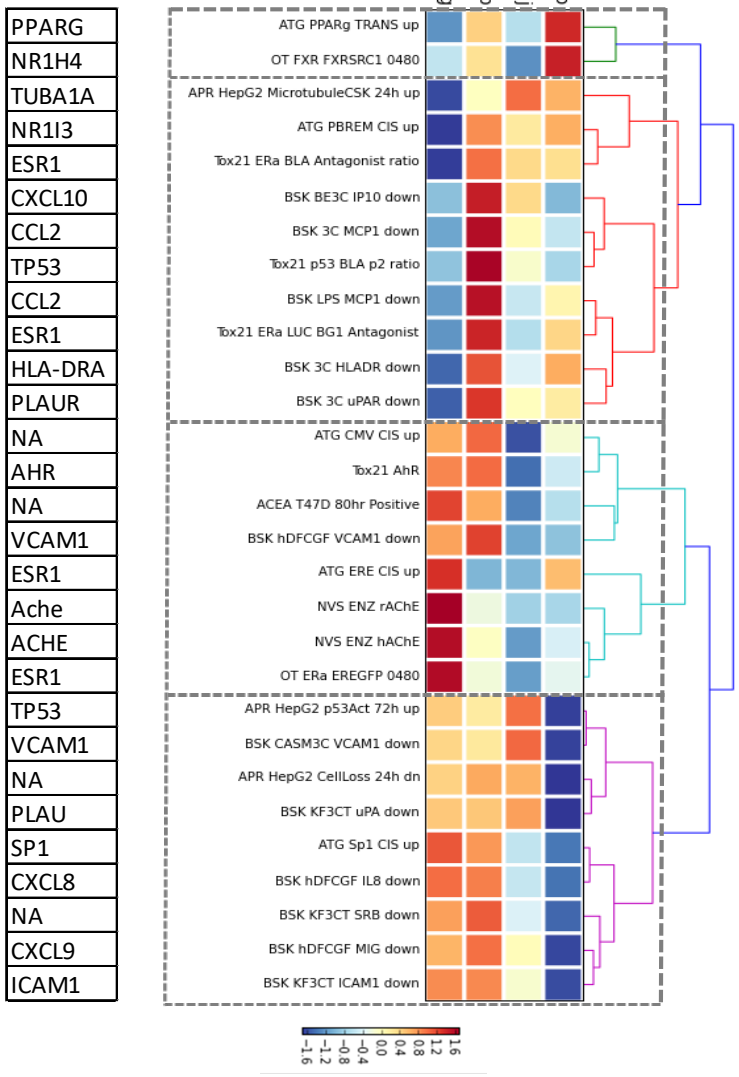
N_{obs} : number of observations.

The maximum predictive performance of different classification methods

		#desc.			BA			Sensitivity			Specificity		
Toxicity	Classifier	BIO	CHM	BC	BIO	CHM	BC	BIO	CHM	BC	BIO	CHM	BC
Hyp	CART0	60	60	55	0.71 (0.10)	**0.80 (0.13)**	**0.80 (0.12)**	0.92 (0.09)	0.86 (0.12)	0.87 (0.12)	0.52 (0.20)	0.73 (0.26)	0.74 (0.24)
	ENSMB	60	60	60	**0.73 (0.10)**	0.75 (0.11)	0.76 (0.10)	0.89 (0.10)	0.84 (0.12)	0.84 (0.12)	0.58 (0.20)	0.67 (0.24)	0.67 (0.22)
	KNN1	60	55	60	0.67 (0.10)	0.70 (0.12)	0.71 (0.12)	0.84 (0.19)	0.87 (0.13)	0.88 (0.13)	0.50 (0.25)	0.56 (0.29)	0.57 (0.28)
	LDA	60	60	60	0.67 (0.10)	0.69 (0.10)	0.68 (0.10)	0.89 (0.10)	0.85 (0.12)	0.85 (0.11)	0.51 (0.19)	0.54 (0.21)	0.54 (0.21)
	NB	60	60	60	0.62 (0.10)	0.66 (0.10)	0.66 (0.10)	0.54 (0.24)	0.71 (0.21)	0.71 (0.20)	0.79 (0.28)	0.65 (0.27)	0.64 (0.25)
	SVCL0	60	60	60	0.66 (0.10)	0.67 (0.10)	0.68 (0.10)	0.92 (0.09)	0.86 (0.12)	0.85 (0.12)	0.43 (0.19)	0.51 (0.23)	0.52 (0.22)
	SVCR0	60	60	60	**0.73 (0.09)**	0.75 (0.11)	0.75 (0.10)	0.97 (0.08)	0.92 (0.12)	0.93 (0.11)	0.50 (0.19)	0.58 (0.23)	0.57 (0.22)
Inj	CART0	60	60	60	**0.74 (0.12)**	**0.79 (0.12)**	**0.80 (0.13)**	0.64 (0.27)	0.74 (0.23)	0.76 (0.24)	0.85 (0.23)	0.85 (0.19)	0.84 (0.18)
	ENSMB	60	60	60	**0.74 (0.13)**	0.77 (0.12)	0.76 (0.13)	0.60 (0.27)	0.69 (0.23)	0.68 (0.23)	0.89 (0.22)	0.84 (0.19)	0.85 (0.18)
	KNN1	50	60	60	0.68 (0.12)	0.72 (0.13)	0.72 (0.14)	0.71 (0.28)	0.78 (0.28)	0.76 (0.26)	0.67 (0.25)	0.69 (0.33)	0.70 (0.31)
	LDA	60	60	60	0.73 (0.13)	0.72 (0.12)	0.73 (0.13)	0.67 (0.27)	0.65 (0.23)	0.65 (0.24)	0.80 (0.22)	0.81 (0.19)	0.81 (0.18)
	NB	55	30	60	0.65 (0.11)	0.65 (0.11)	0.65 (0.11)	0.81 (0.28)	0.73 (0.30)	0.71 (0.30)	0.49 (0.27)	0.77 (0.34)	0.76 (0.32)
	SVCL0	60	60	60	0.72 (0.12)	0.72 (0.12)	0.72 (0.13)	0.60 (0.26)	0.65 (0.23)	0.65 (0.23)	0.85 (0.23)	0.79 (0.19)	0.80 (0.18)
	SVCR0	60	60	60	**0.74 (0.12)**	0.77 (0.12)	0.77 (0.12)	0.54 (0.26)	0.74 (0.24)	0.71 (0.24)	0.94 (0.22)	0.81 (0.19)	0.82 (0.18)
Pro	CART0	60	60	60	**0.71 (0.15)**	**0.80 (0.15)**	**0.80 (0.16)**	0.52 (0.29)	0.71 (0.30)	0.72 (0.32)	0.93 (0.12)	0.91 (0.12)	0.91 (0.12)
	ENSMB	60	60	60	0.71 (0.15)	0.75 (0.14)	0.74 (0.15)	0.46 (0.29)	0.56 (0.28)	0.53 (0.30)	0.96 (0.09)	0.95 (0.11)	0.95 (0.11)
	KNN1	60	60	55	0.67 (0.16)	0.73 (0.17)	0.71 (0.18)	0.71 (0.30)	0.69 (0.33)	0.64 (0.32)	0.66 (0.26)	0.87 (0.39)	0.87 (0.36)
	LDA	60	60	60	0.71 (0.16)	0.75 (0.16)	0.74 (0.16)	0.57 (0.29)	0.61 (0.29)	0.59 (0.29)	0.90 (0.14)	0.89 (0.12)	0.90 (0.12)
	NB	55	60	60	0.63 (0.15)	0.68 (0.15)	0.66 (0.16)	0.76 (0.35)	0.53 (0.30)	0.46 (0.31)	0.51 (0.32)	0.86 (0.19)	0.88 (0.17)
	SVCL0	60	60	60	0.67 (0.15)	0.73 (0.15)	0.72 (0.16)	0.40 (0.29)	0.55 (0.29)	0.54 (0.31)	0.94 (0.10)	0.91 (0.12)	0.90 (0.12)
	SVCR0	60	60	60	0.71 (0.14)	0.76 (0.15)	0.76 (0.16)	0.44 (0.28)	0.54 (0.29)	0.53 (0.31)	0.99 (0.07)	0.98 (0.12)	0.98 (0.12)

Visualizing bioactivity descriptors most frequently selected in classifying hepatotoxicity

Target / gene Assay endpoints



Conclusions

- The results show the utility of high-throughput assays for characterizing the hepatotoxic liability in rodents.
- The results suggest the advantage of using hybrid representations that integrate bioactivity and chemical structure descriptors for hepatotoxicity prediction.
- Provided linkages between the *in vitro* bioactivity of environmental chemicals and their adverse histopathological outcomes.

Acknowledgements

Committee:

Dr. Imran Shah	NCCT, U.S. EPA
Dr. Xiaowei Xu	UALR
Dr. Elizabeth Pierce	UALR
Dr. Huixiao Hong	NCTR, U.S. FDA
Dr. Minjun Chen	NCTR, U.S. FDA

Organizations:

UALR/UAMS Joint Bioinformatics program

National Center for Computational Toxicology (NCCT), U.S. EPA

Oak Ridge Institute for Science and Education (ORISE)

United States
Environmental Protection
Agency

Thank you!