From QSAR to Big Data: Developing Mechanism-Driven Predictive Models for Animal Toxicity

Hao Zhu

Department of Chemistry
The Rutgers Center for Computational and Integrative Biology
Rutgers University-Camden
Email: hao.zhu99@rutgers.edu

September 24, 2015

Acknowledgements

Rutgers:

- PhD students: **Marlene Kim**, **Wenyi Wang**, Daniel Russo, Linlin Zhao
- Master students: Kathryn Ribay, Joe Hess
- Visiting Scholar: **Dr. Aleck Sedykh, Dr. Jun Zhang**

John Hopkins University:

- Dr. Thomas Hartung

Shandong University:

- Dr. Bing Yan

NCATS:

- **Dr. Menghang Xia**, **Dr. Ruili Huang**

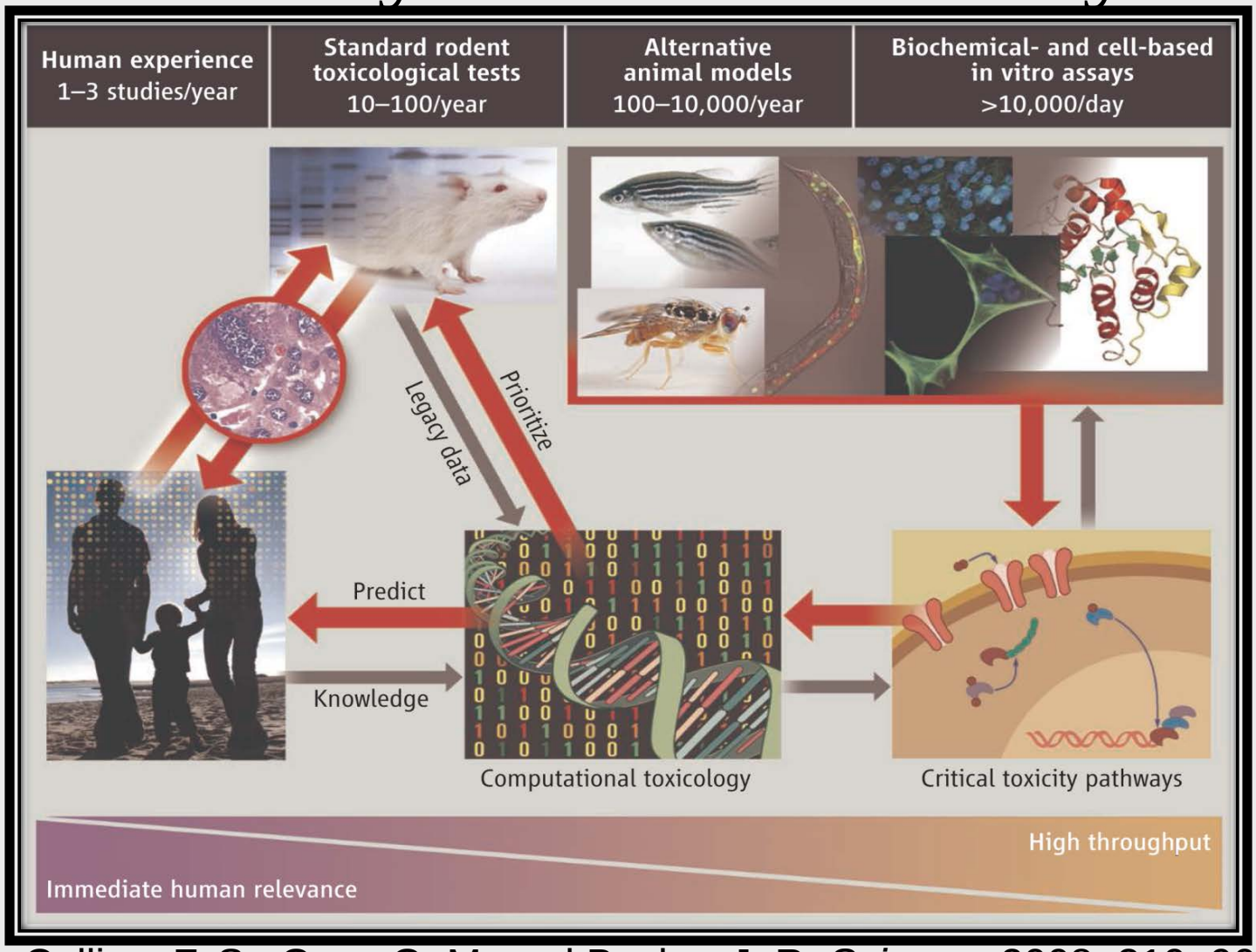
ICCVAM:

- Dr. Judy Strickland

Funding resource:

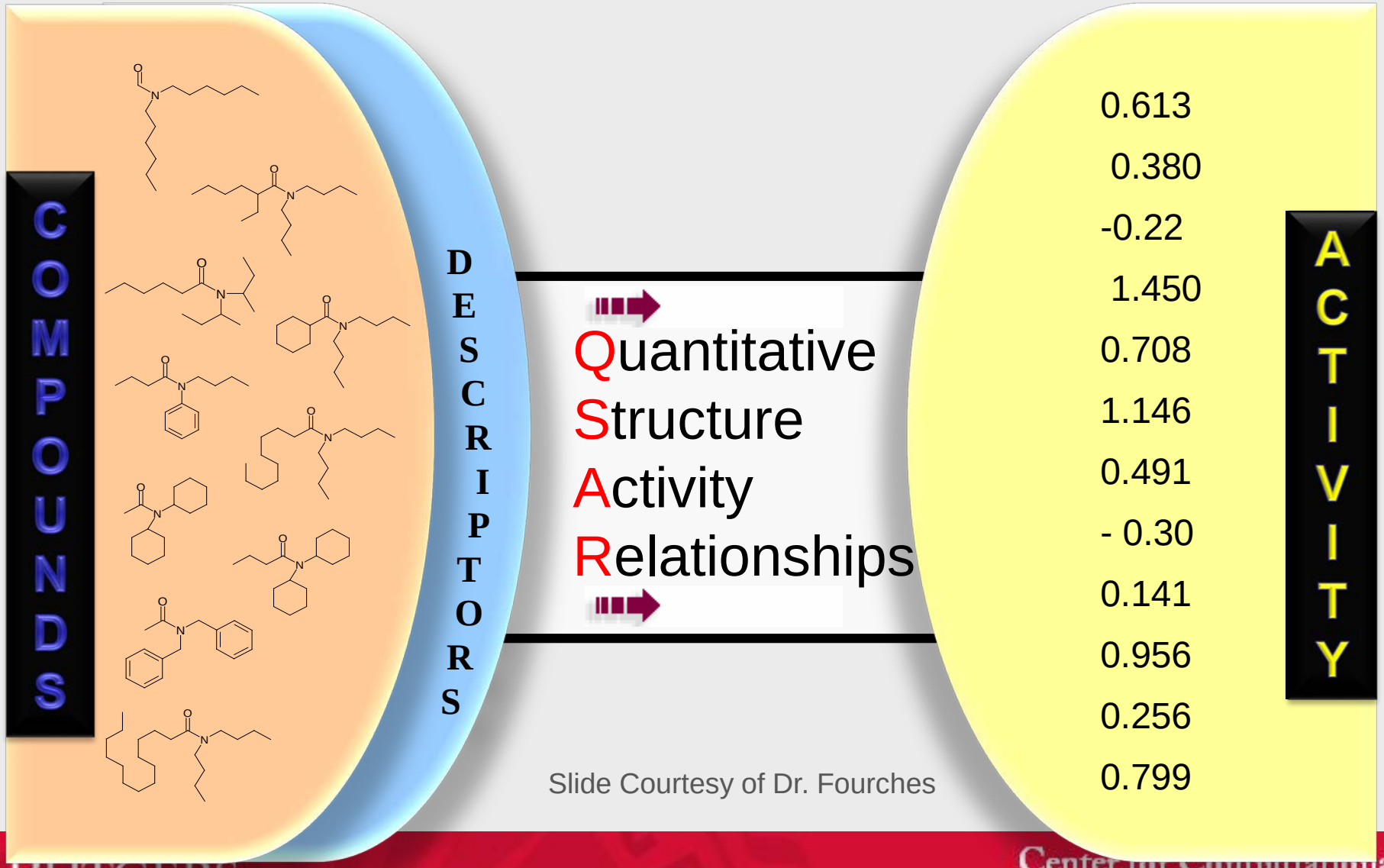
- National Institute of Health: 1R15ES023148
- Society of Toxicology: Colgate-Palmolive Grant for Alternative Research

Toxicity evaluation today



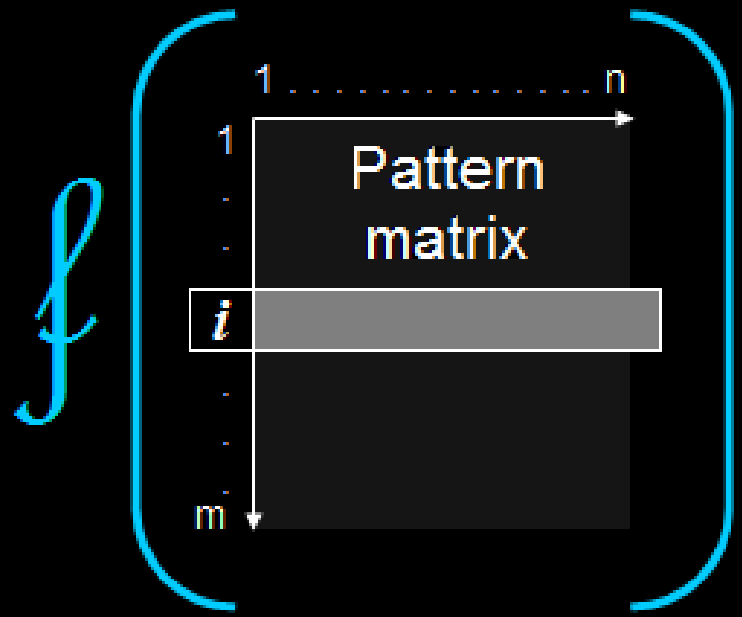
Collins, F. S., Gray, G. M. and Bucher J. R. *Science*, 2008, 319, 906-907

Principles of QSAR modeling



Slide Courtesy of Dr. Fourches

Principles of QSAR modeling



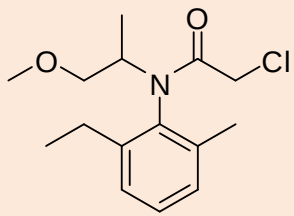
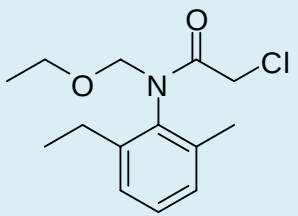
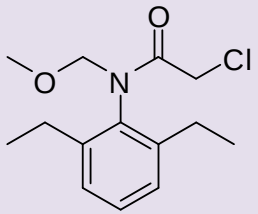
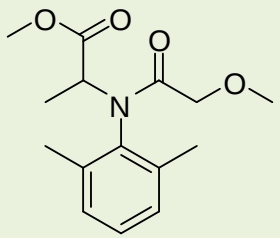
f = **PROPERTY** *(i)*

With m molecules and n descriptors

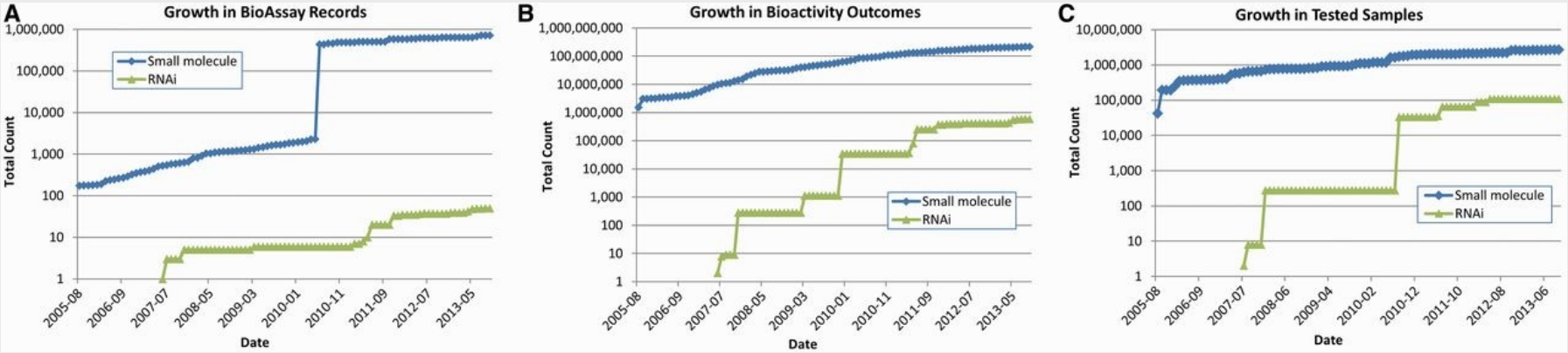
0.799

The “similar” compounds that have “dissimilar” toxicity profiles

NOTES:
0: non-toxic/inactive
1: toxic/active
-: not tested
NN: Nearest Neighbor

		IN VIVO ASSAYS					
		MOUSE_KIDNEY	RAT_SKELETAL_AXIAL	MGR_RAT_LIVER	MGR_RAT_KIDNEY		
METOLACHLOR		0	0	0	0		NN1-ACETOCHLOR
		1	1	1	1		
NN2-ALACHLOR		-	0	-	-		NN3-METALAXYL
		0	1	1	0		

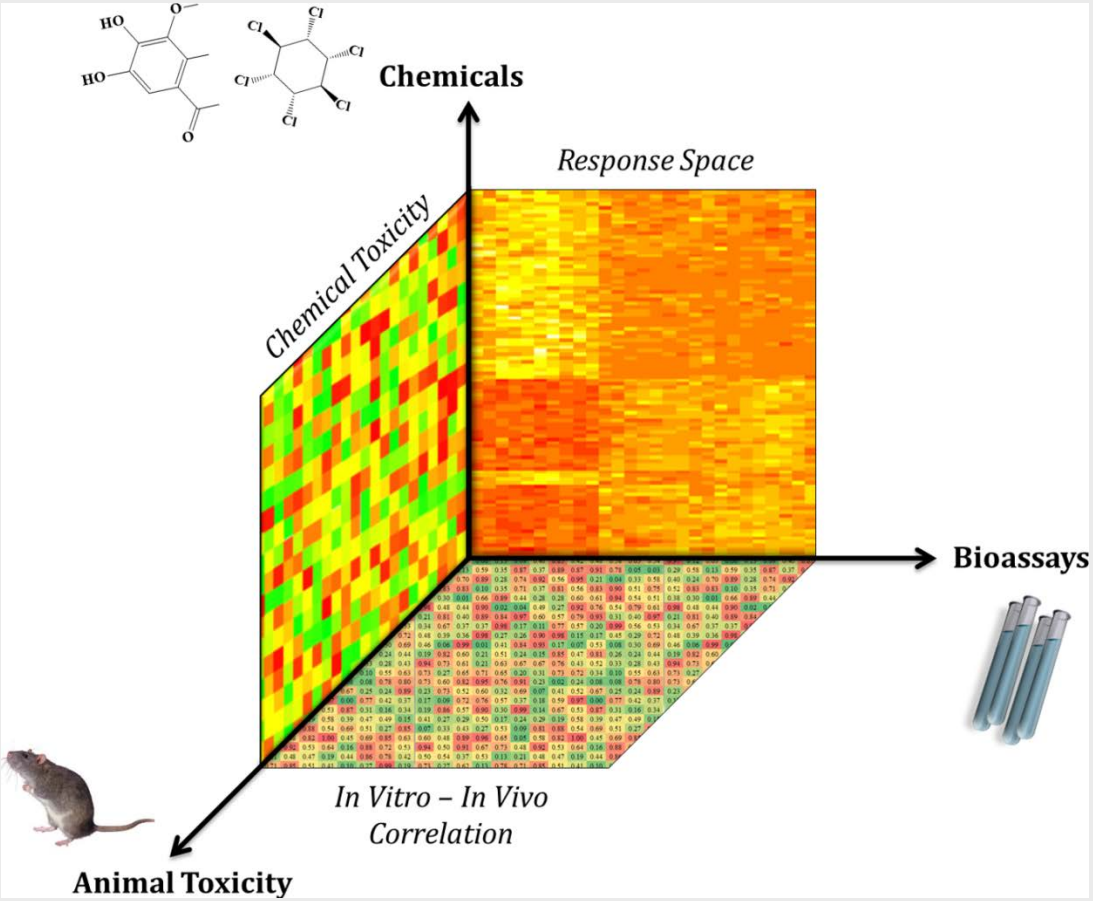
PubChem data in 2014



- >700,000 bioassays
- >200,000,000 bioactivity outcomes
- >1,200,000,000 data points
- >2,800,000 small molecule samples
- >1,900,000 chemical structures
- >108,000 RNAi reagents

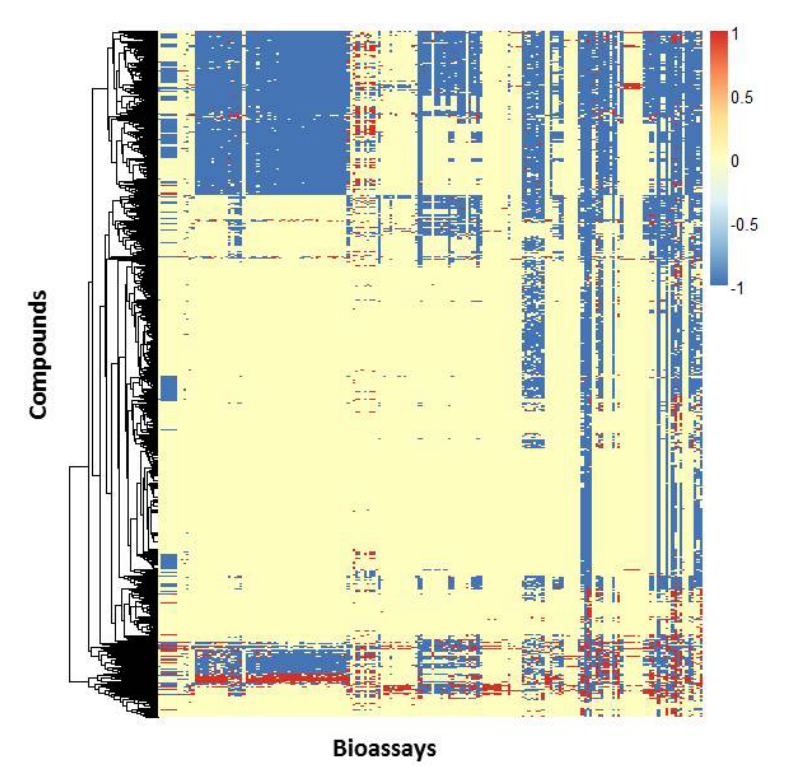
Yang et al. *Nucleic Acids Res.* 2014 Jan;42: D1075-82

Chemical-in vitro-in vivo profiles in big data era



Chem. Res. Tox. 2014; (27) 1643-1651

Before the ToxCast project, data already existed



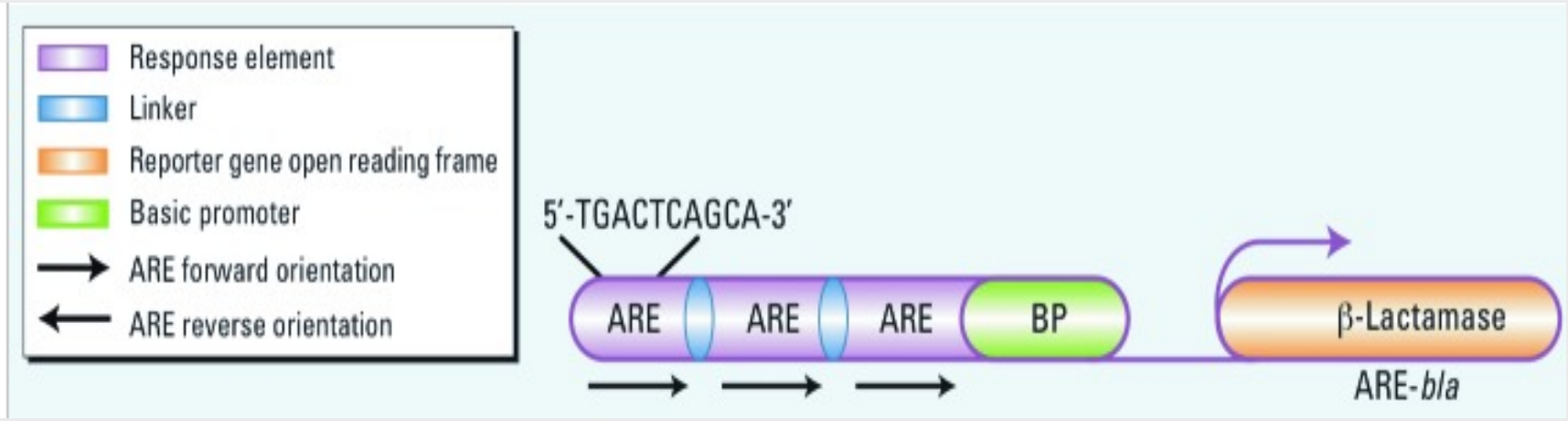
Obtained from PubChem on Aug. 1, 2013, before the ToxCast phase II data was released.

The current question is:

- What can we do if we have limited in-house data available for the compounds of interest?

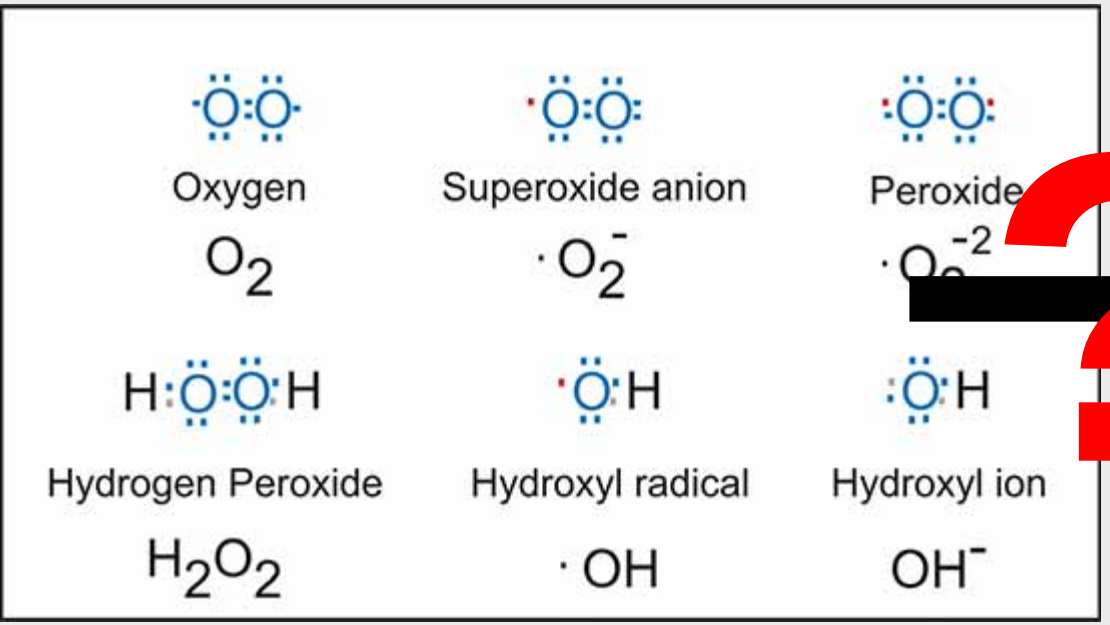
Antioxidant Response Element β-lactamase reporter gene assay (ARE-*bla*)

- Recognized by the Tox21 program as one of the most important toxicity assays
- ARE genes play a role in alleviating oxidative stress



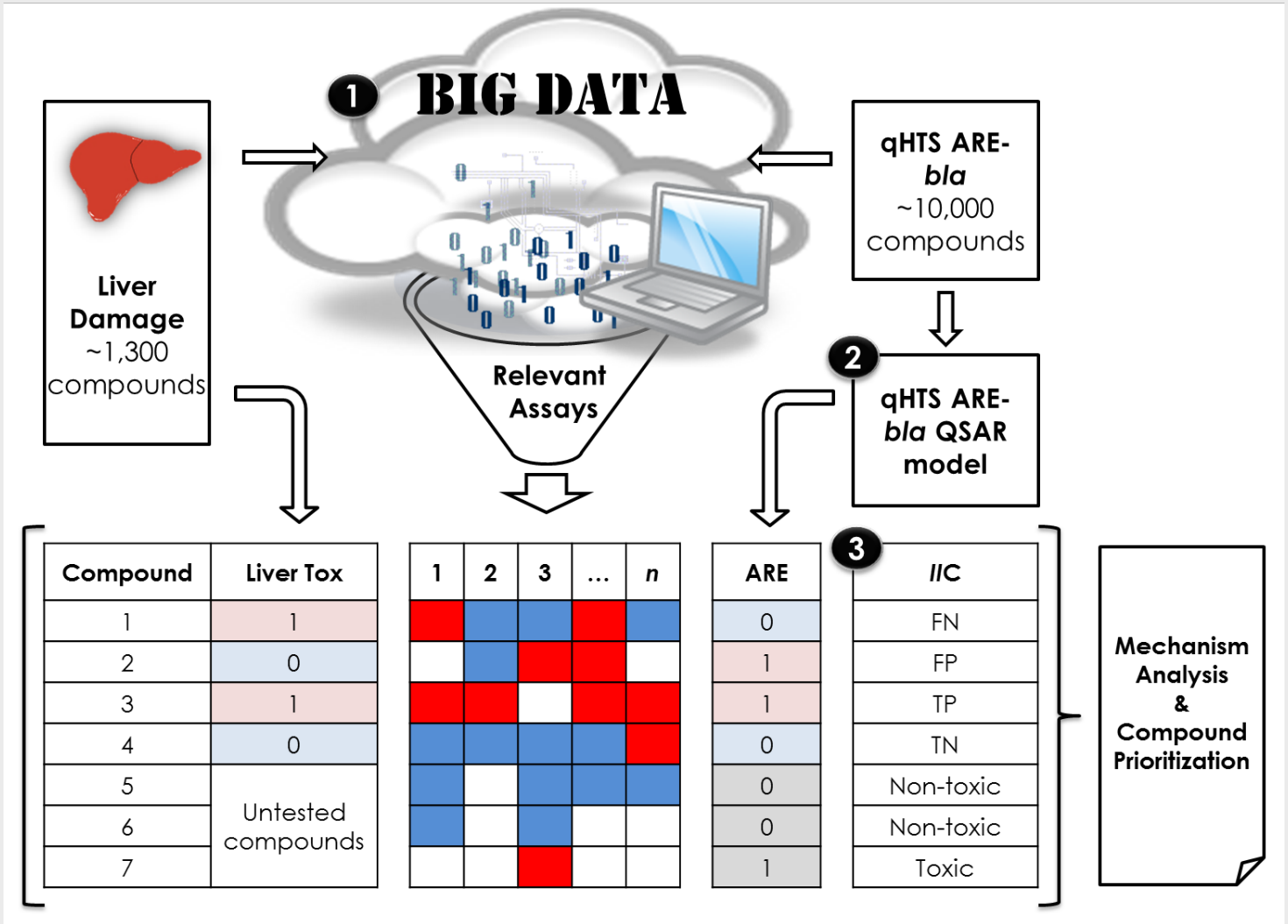
Shukla SJ, *et al.* Environ Health Perspect. 2012, 120(8):1150-6.

Reactive Oxygen Species (ROS)



Liver damage

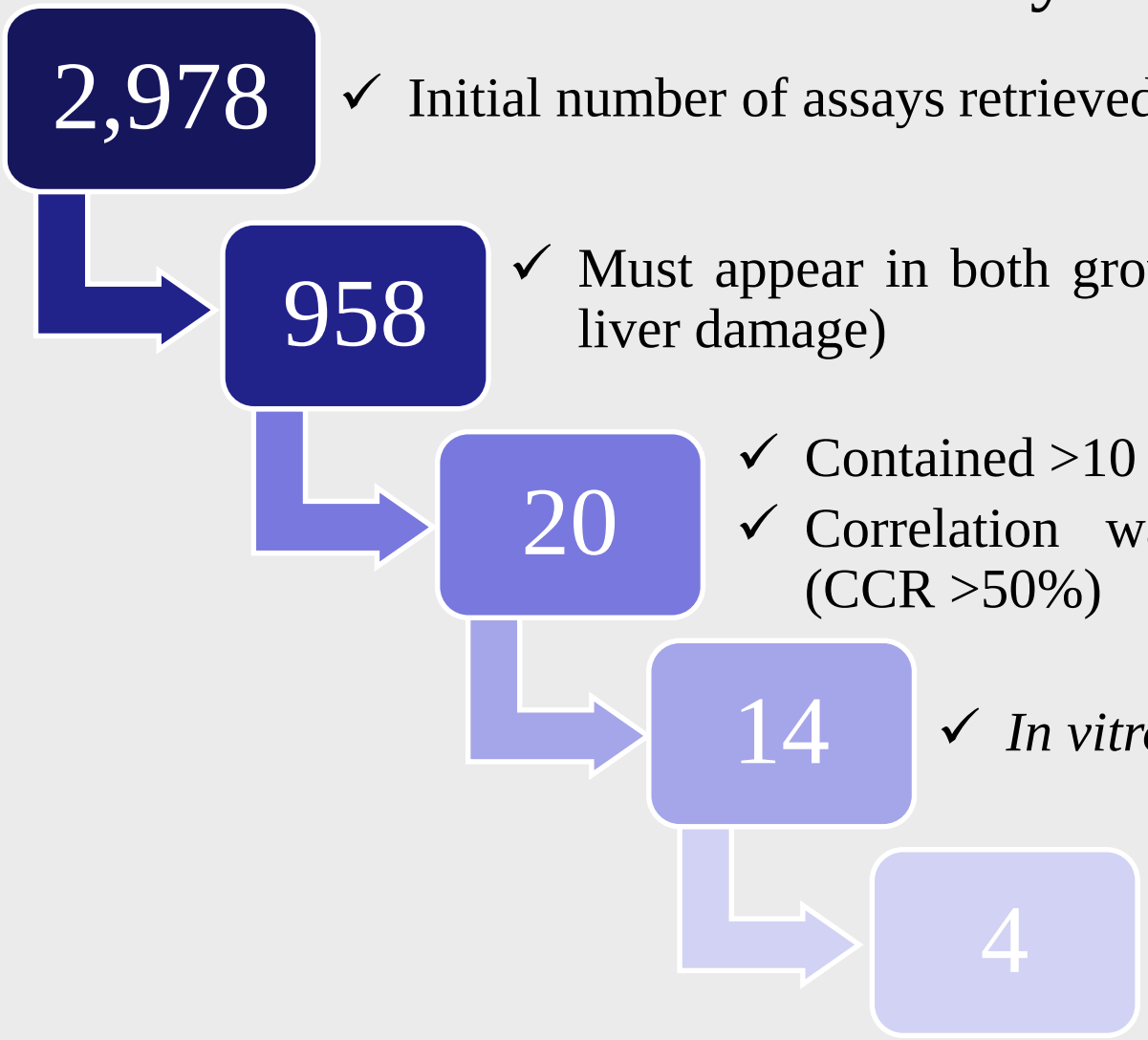
Workflow for profiling liver toxicants



Profiling target compounds with biological responses using automated tool

- Input – target compounds:
  1. qHTS ARE-*bla* dataset (10,928 compounds)
  2. FDA liver damage dataset (1,314 compounds)
- Output – assays related to:
  1. qHTS ARE-*bla* activation (1,819 assays)
  2. Liver damage (1,159 assays)

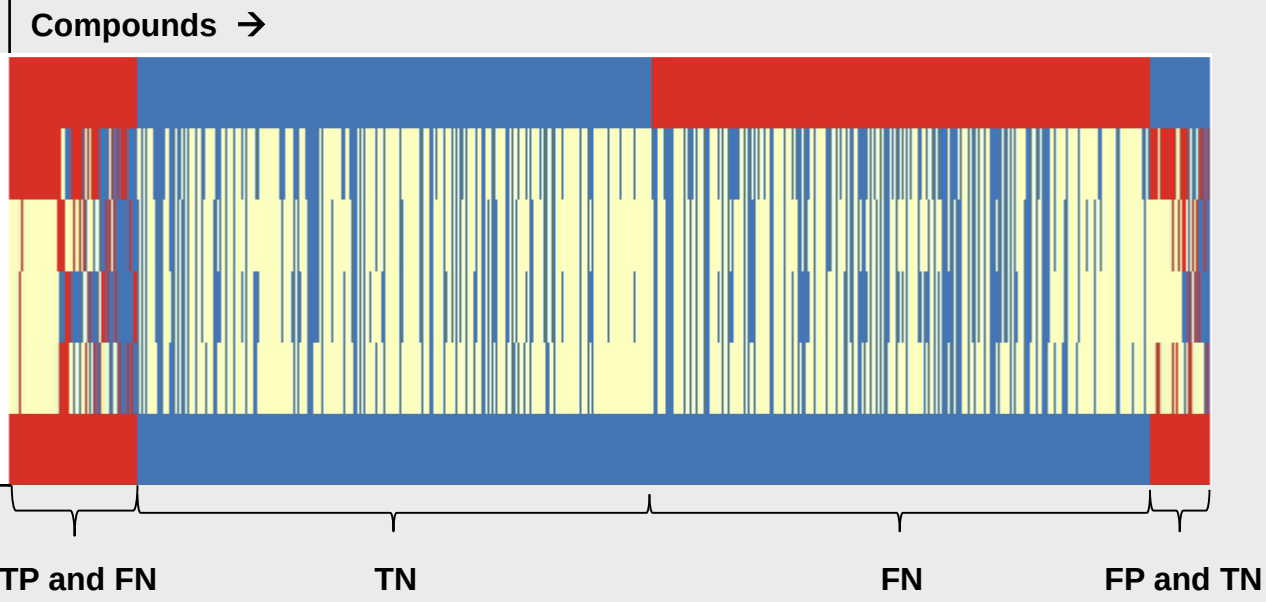
Criteria for filtering inadequate and finding relevant assays



- ✓ Initial number of assays retrieved
- ✓ Must appear in both groups (qHTS ARE-*bla* and liver damage)
- ✓ Contained >10 true positive responses
- ✓ Correlation was better than random (CCR >50%)
- ✓ *In vitro* assay
- ✓ Evidence supported by reliable literature

Individual assays showed poor IIC, but the combined response using RA>0.25 show statistical significance

Bioassay	p-value
Liver damage activity	--
686978	0.59
743067	0.42
743140	0.24
(ARE-*bla*) 743202	0.31
Combined	**4.25×10^{-4}**



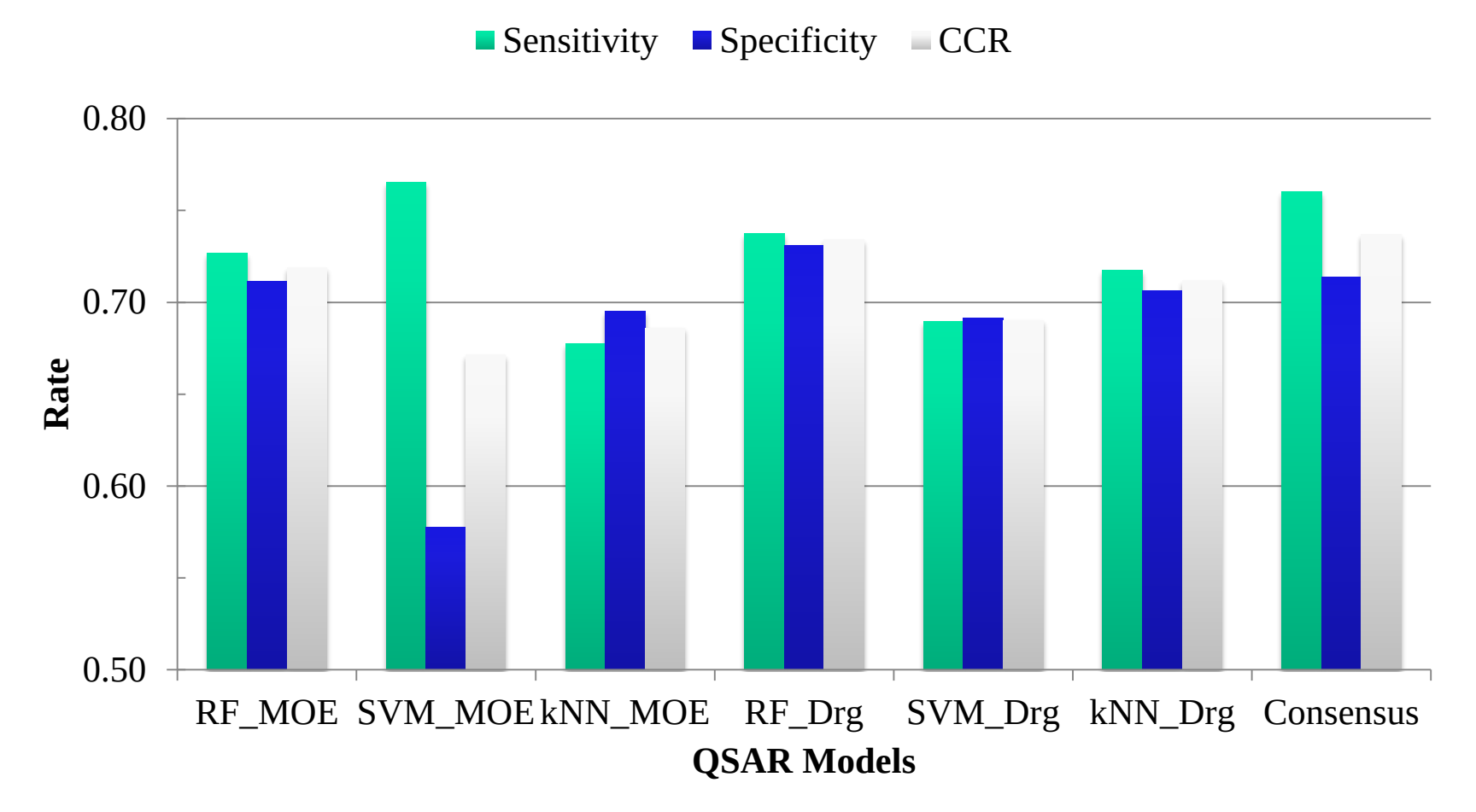
- Active or toxic
- Inconclusive or untested
- Inactive or non-toxic

Rate of actives (RA)

$$RA = \frac{A}{A + I}$$

A = no. of active responses
I = no. in active responses

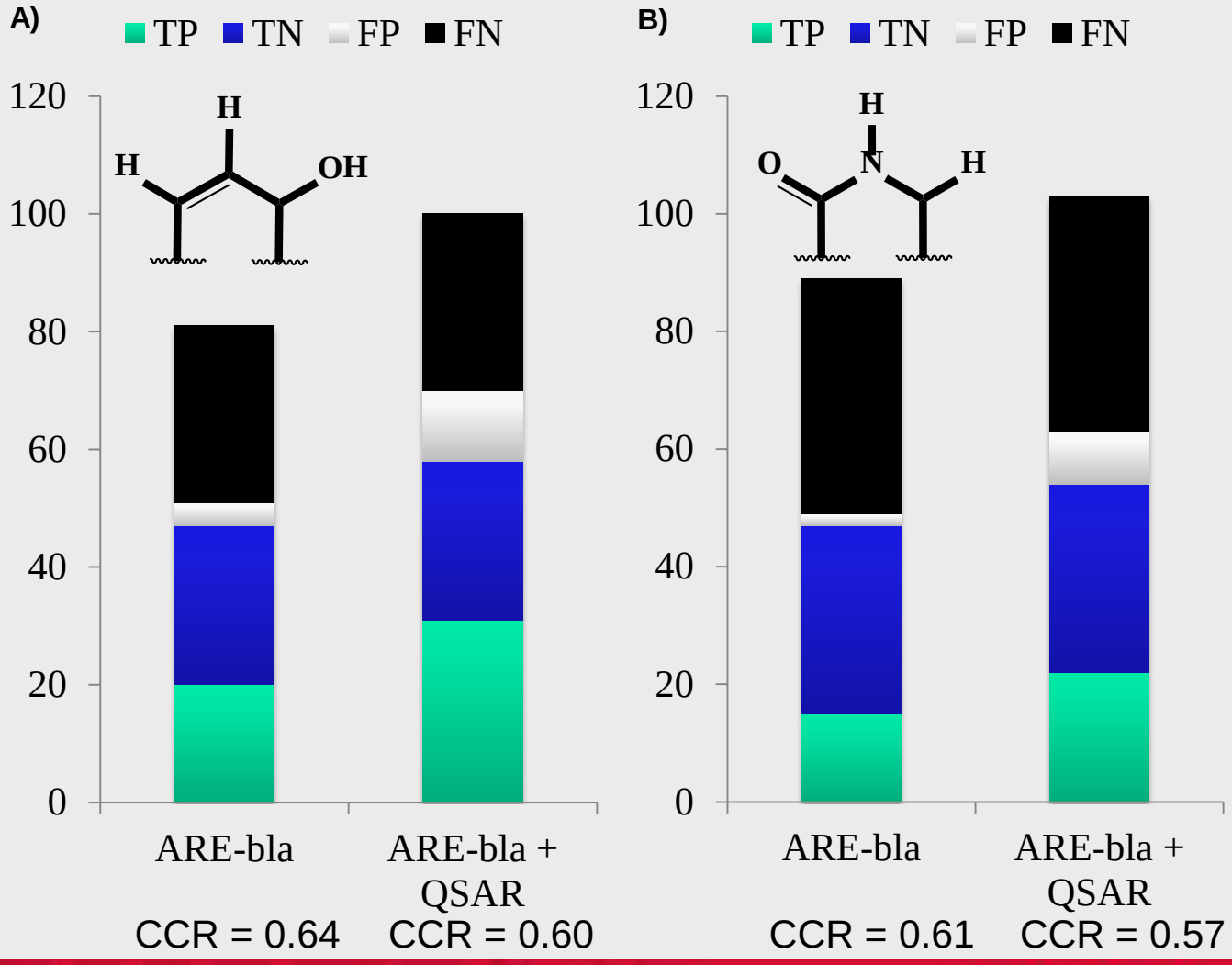
Modeling qHTS ARE-*bla* activation using QSAR approaches: 5-fold cross validation for all individual models



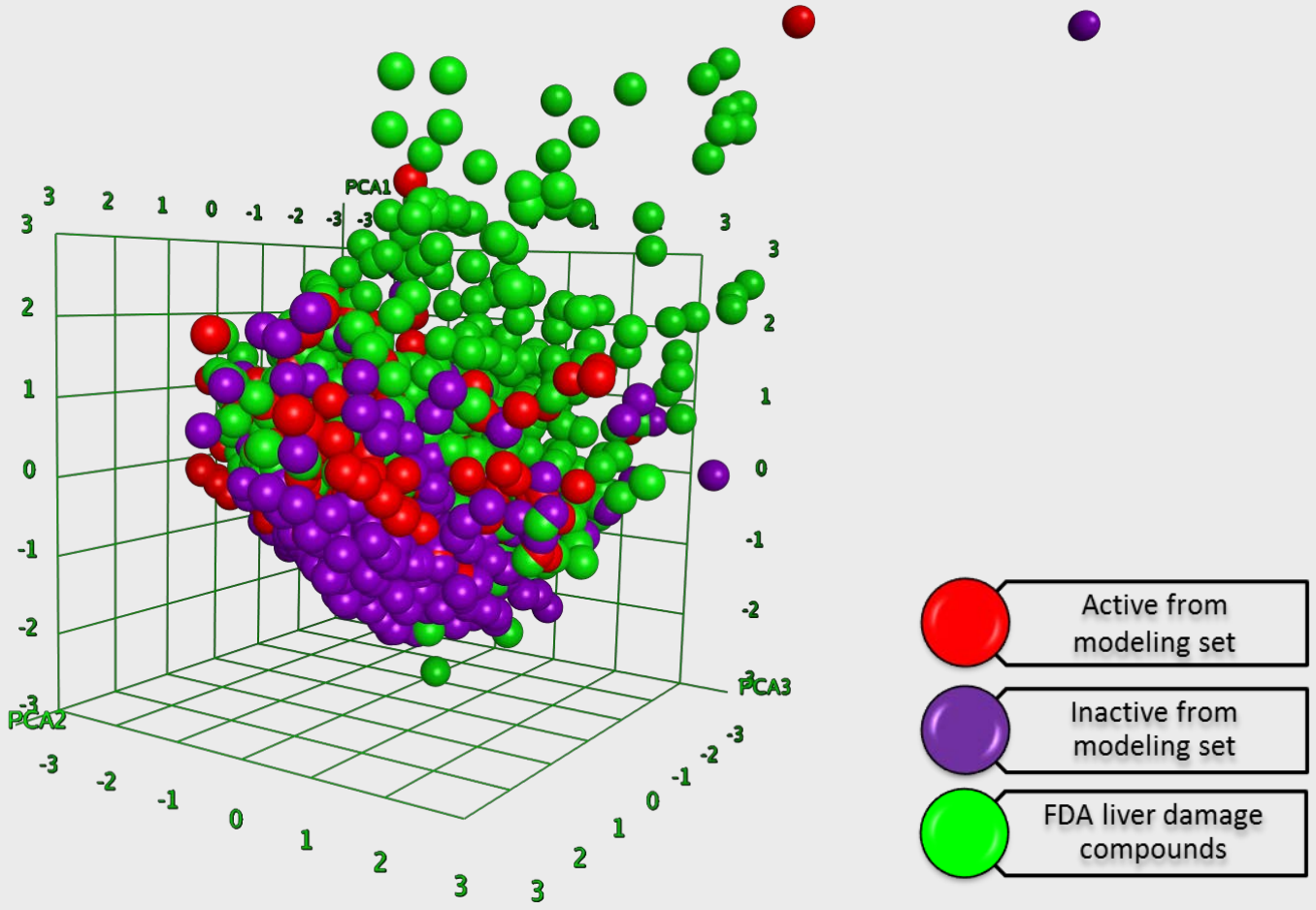
Evaluating In vitro-In vivo Correlations (IICs)

- Focused on compounds that were active in qHTS ARE-*bla* and liver toxic
- Searched for common chemical features
- Evaluated IICs (sensitivity, specificity, CCR, and X^2)

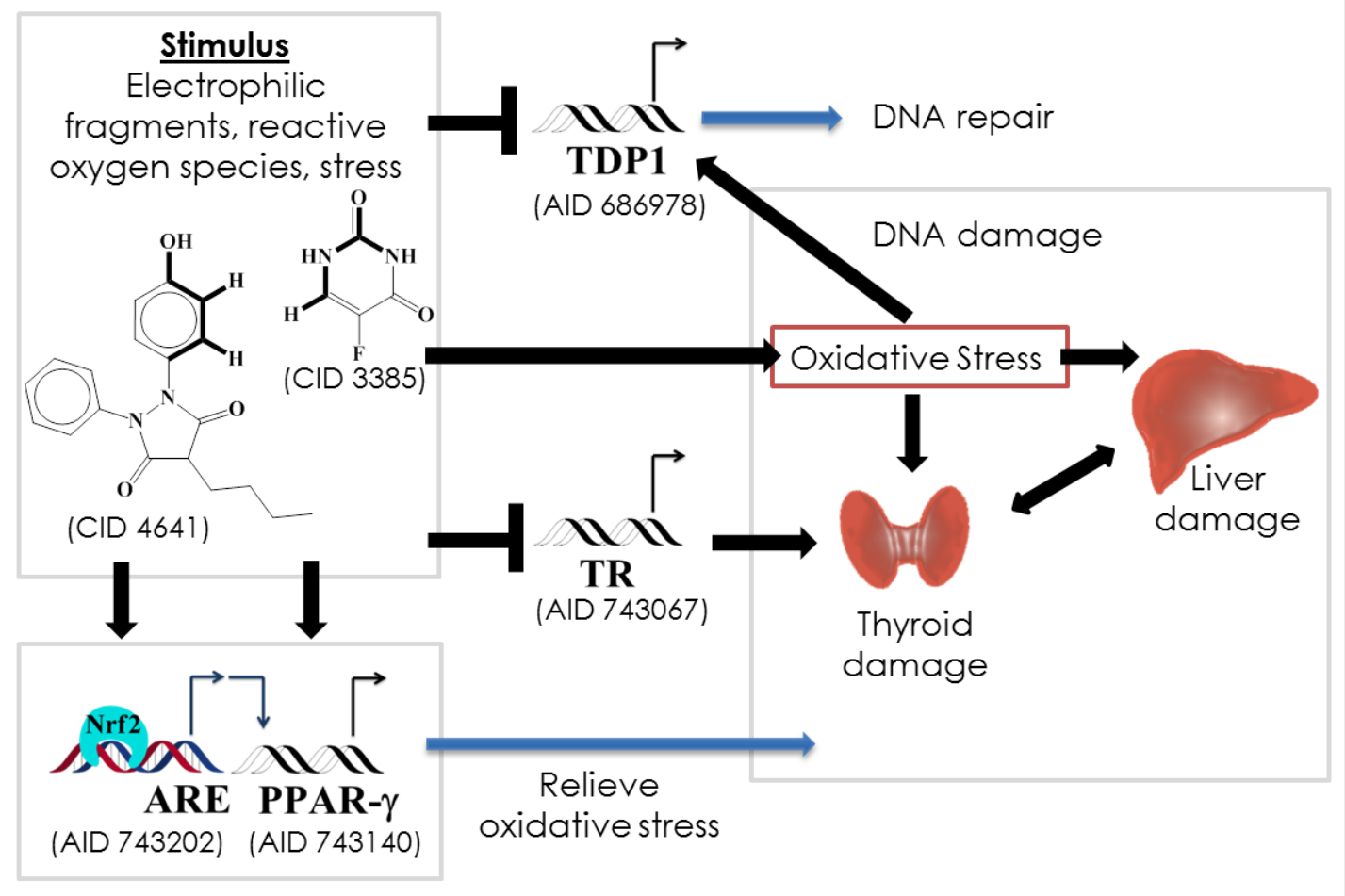
IIC between qHTS ARE-*bla* activation and liver damage for overlapping compounds containing the toxicophores



3-D plot of Tox21 phase II modeling set vs FDA liver damage dataset using principal components analysis



Liver toxicity mechanism analysis involving ARE pathway perturbations



Conclusions

- Developed a workflow
  - Profiles biological responses from big data
  - Incorporates QSAR models to fill-in missing data
  - Evaluates the chemical IIC
- Identified toxicophores and assays that can be used to assess liver damage induced by oxidative stress
- Workflow can be adapted to model or assess other complex animal toxicity endpoints

Mechanism profiling liver toxicants by using antioxidant response element assay data model and public big data. *Environ. Health Perspect.* In press

Take home message

- Reliable information exists, but it is difficult to locate
- Good data may not guarantee good decisions