Understanding Significant Polypharmacology in Chemical-induced Toxicity Using Toxcast and Tox21 Data

UNIVERSITY OF CAMBRIDGE

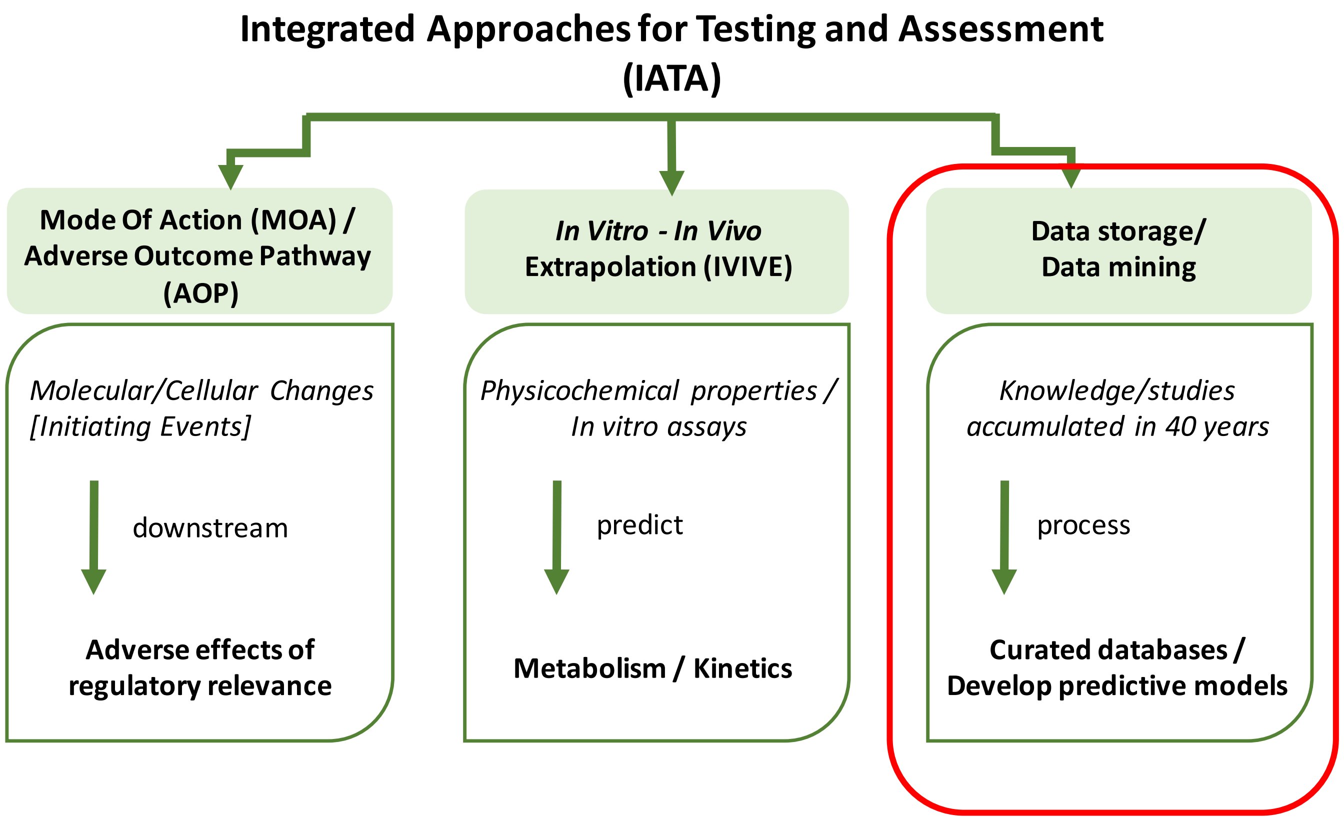
Samar Mahmoud
symm2@cam.ac.uk
s.mahmoud75@yahoo.com

22 Aug 19

Alternative approaches for toxicity assessment

Recent efforts shifted towards interpretable models which can explain the mechanistic background, helping in:

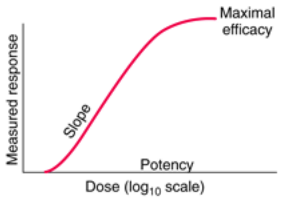
- Understanding mechanisms of toxicity
- Rationalizing decision making in toxicity assessment
- Prioritizing *in vitro* models for the relevant biological activities
- Variable, and slow, regulatory acceptance of black box QSAR models



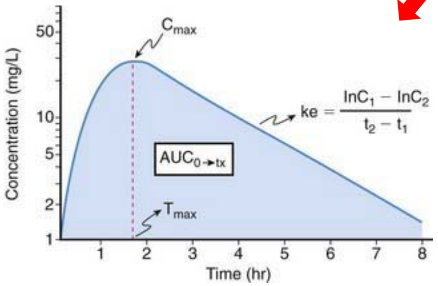
QSAR models to predict chemical properties

Toxicity

Bioactivity

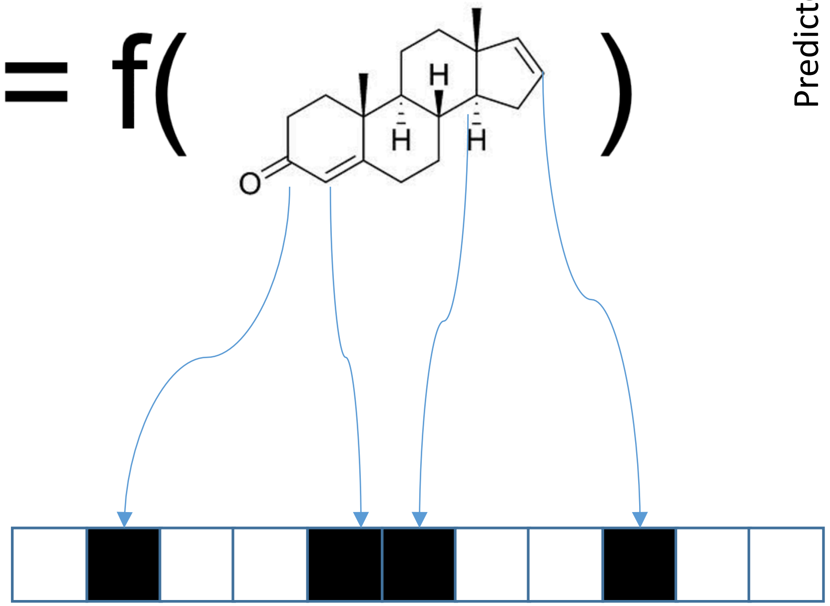


ADME

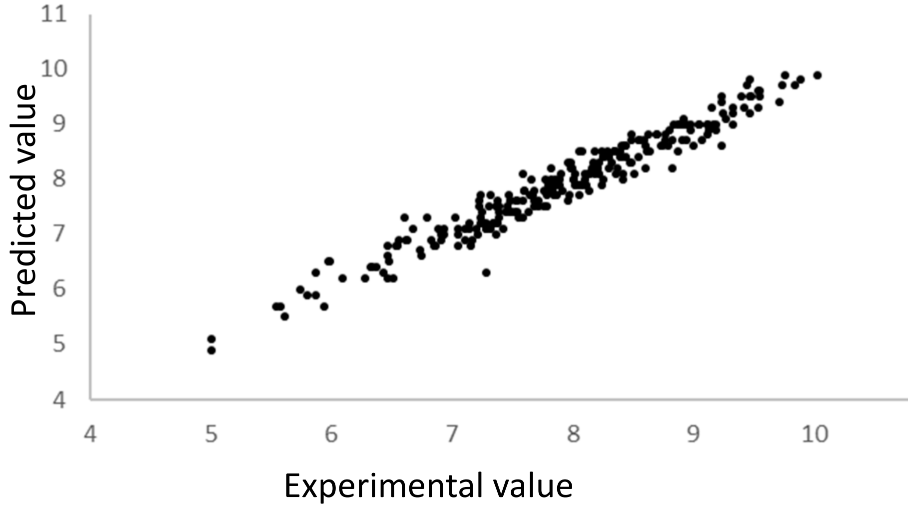


$y = f($

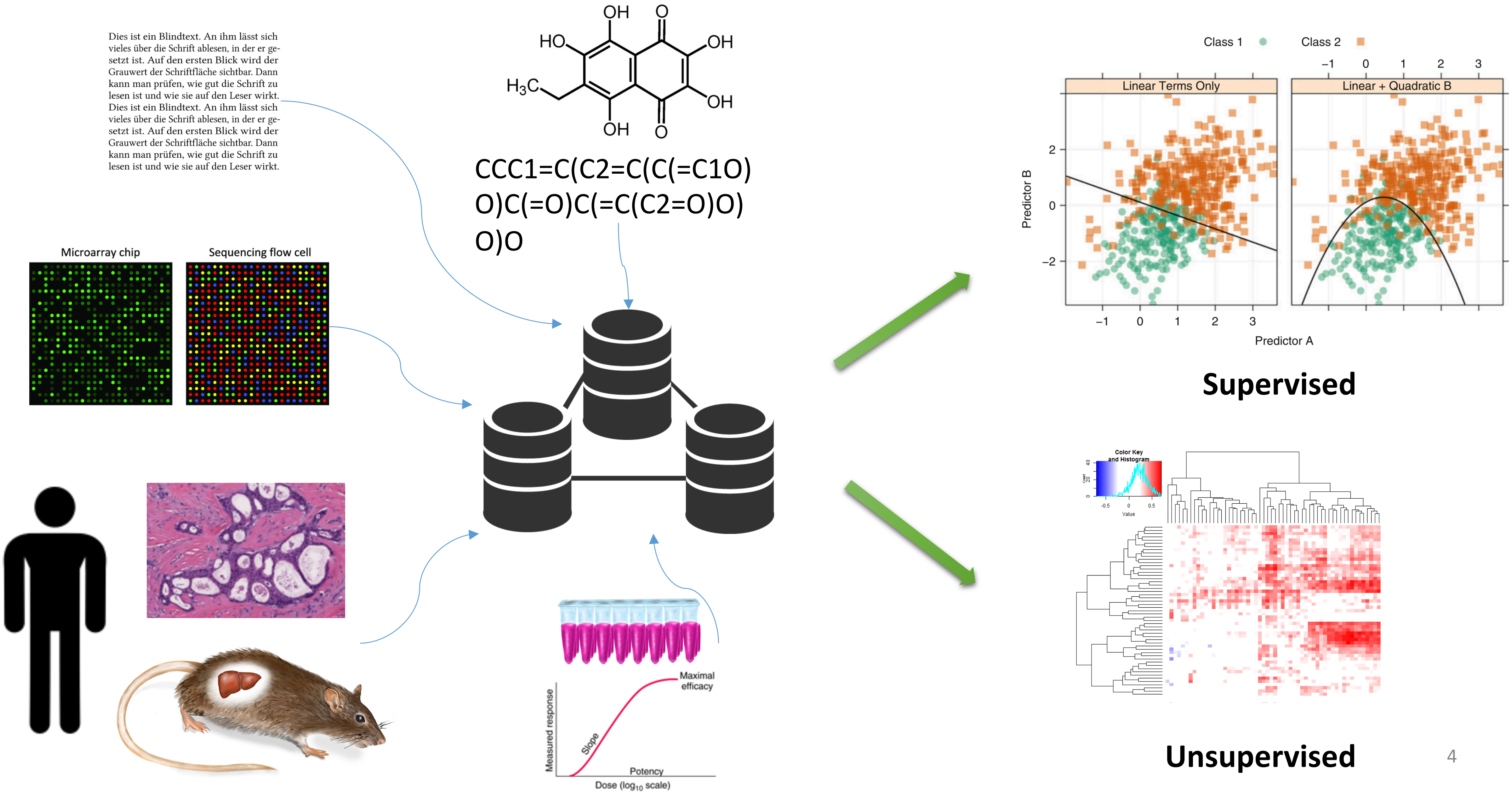
CHEMICAL STRUCTURE



$)$



Data is used to build predictive models and extract patterns



Machine learning models are used to predict *in vivo* and clinical adverse effects

Chemical Research in Toxicology

Article

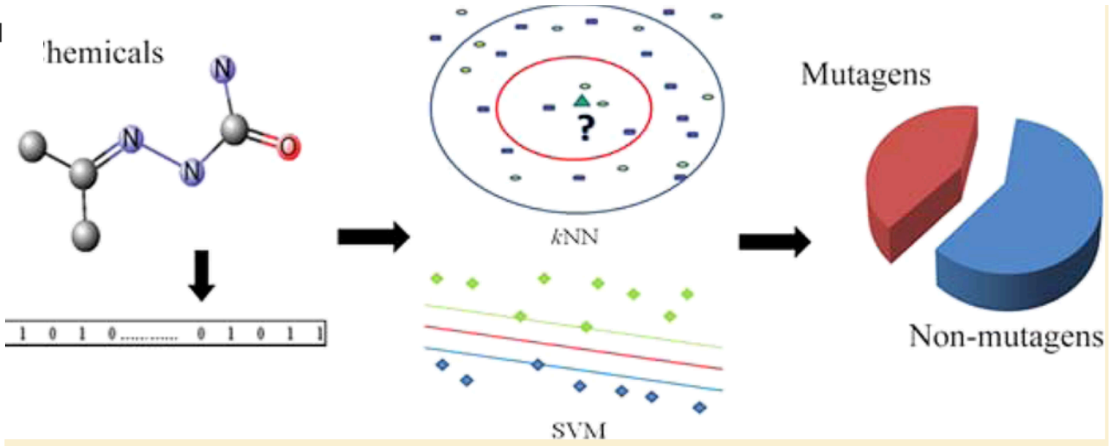
pubs.acs.org/crt

Predicting Hepatotoxicity Using ToxCast *in Vitro* Bioactivity and Chemical Structure

Jie Liu,[†,‡,§] Kamel Mansouri,[†,§] Richard S. Judson,[†] Matthew T. Martin,[†] Huixiao Hong,[∥] Minjun Chen,[∥] Xiaowei Xu,[‡,∥] Russell S. Thomas,[†] and Imran Shah*,[†]

In silico Prediction of Chemical Ames Mutagenicity

Congying Xu,[†] Feixiong Cheng,[†] Lei Chen,[†] Zheng Du,[†] Weihua Li,[†] Guixia Liu,*,[†,‡] Philip W. Lee,[†] and Yun Tang*,[†]



Deep Learning-based Prediction of Drug-induced Cardiotoxicity

Chuipu Cai[1,2], Pengfei Guo[1], Yadi Zhou[3], Jingwei Zhou[1], Qi Wang[1], Fengxue Zhang[2], Jiansong Fang[1,*], and Feixiong Cheng[4,5,6,*]

Gene Expression Data Based Deep Learning Model for Accurate Prediction of Drug-Induced Liver Injury in Advance

Chunlai Feng,*,# Hengwei Chen,# Xianqin Yuan, Mengqiu Sun, Kexin Chu, Hanqin Liu, and Mengjie Rui*

PAPER

View Article Online
View Journal | View Issue

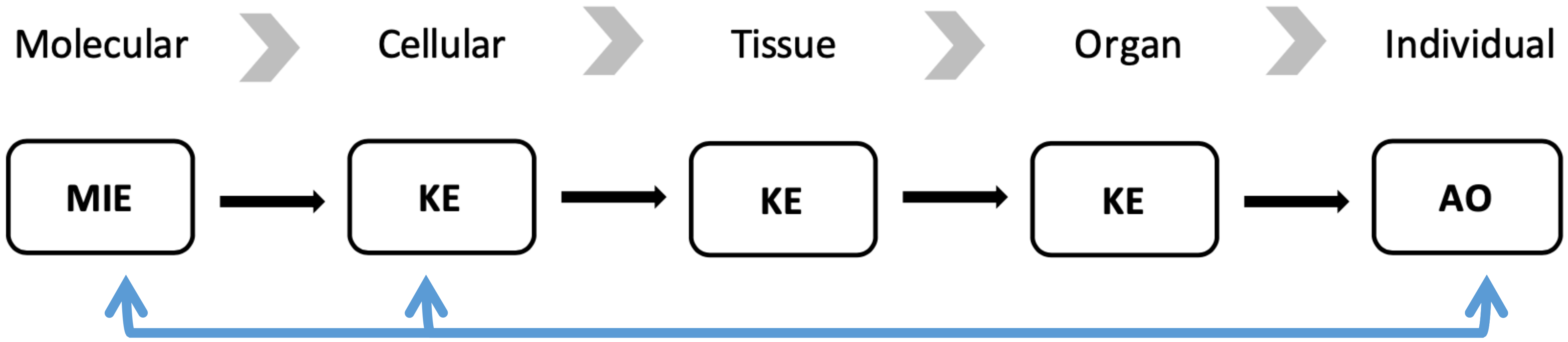
CrossMark
click for updates

Cite this: *Toxicol. Res.*, 2016, **5**, 570

In silico prediction of hERG potassium channel blockage by chemical category approaches†

Chen Zhang, Yuan Zhou, Shikai Gu, Zengrui Wu, Wenjie Wu, Changming Liu, Kaidong Wang, Guixia Liu, Weihua Li, Philip W. Lee and Yun Tang*

Adverse outcome pathway framework is used for mechanistic interpretation of adverse effects



MIE: **Molecular Initiating Event**

KE: **Key Event**

AO: **Adverse Outcome**

By mining the statistical associations between changes at molecular and cellular level against toxicity in animal or human, we can generate hypothesis about mechanisms of toxicity

How data-based methods can be used to derive hypotheses about toxicity modes of action

Interpretable Features

Chemical Properties

-Physicochemical properties

-Structural alerts/Toxicophores

Biological activity

-Molecular

-Cellular

-Tissue/Organ

Interpretable Algorithms

-Correlation/association analysis

- Similarity over chemical or biological space (analogues)

- multivariate models (regression and decision trees)

Integrating clinical reports and *in vitro* data can be mined to derive mechanistic hypotheses

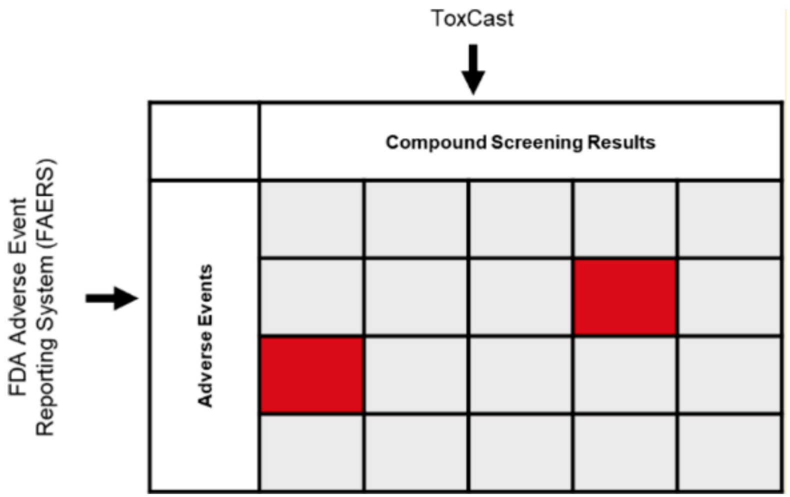
Chemical Research in Toxicology

Article

Cite This: *Chem. Res. Toxicol.* 2018, 31, 1119–1127 pubs.acs.org/crt

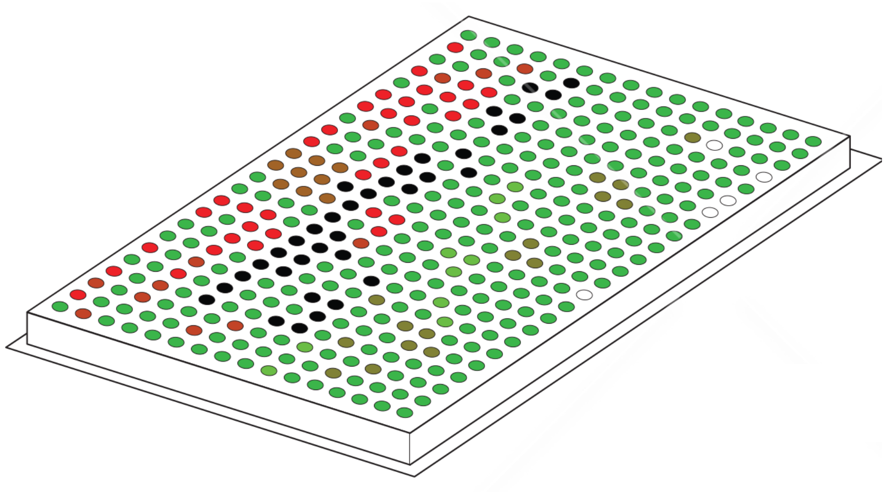
Information-Derived Mechanistic Hypotheses for Structural Cardiotoxicity

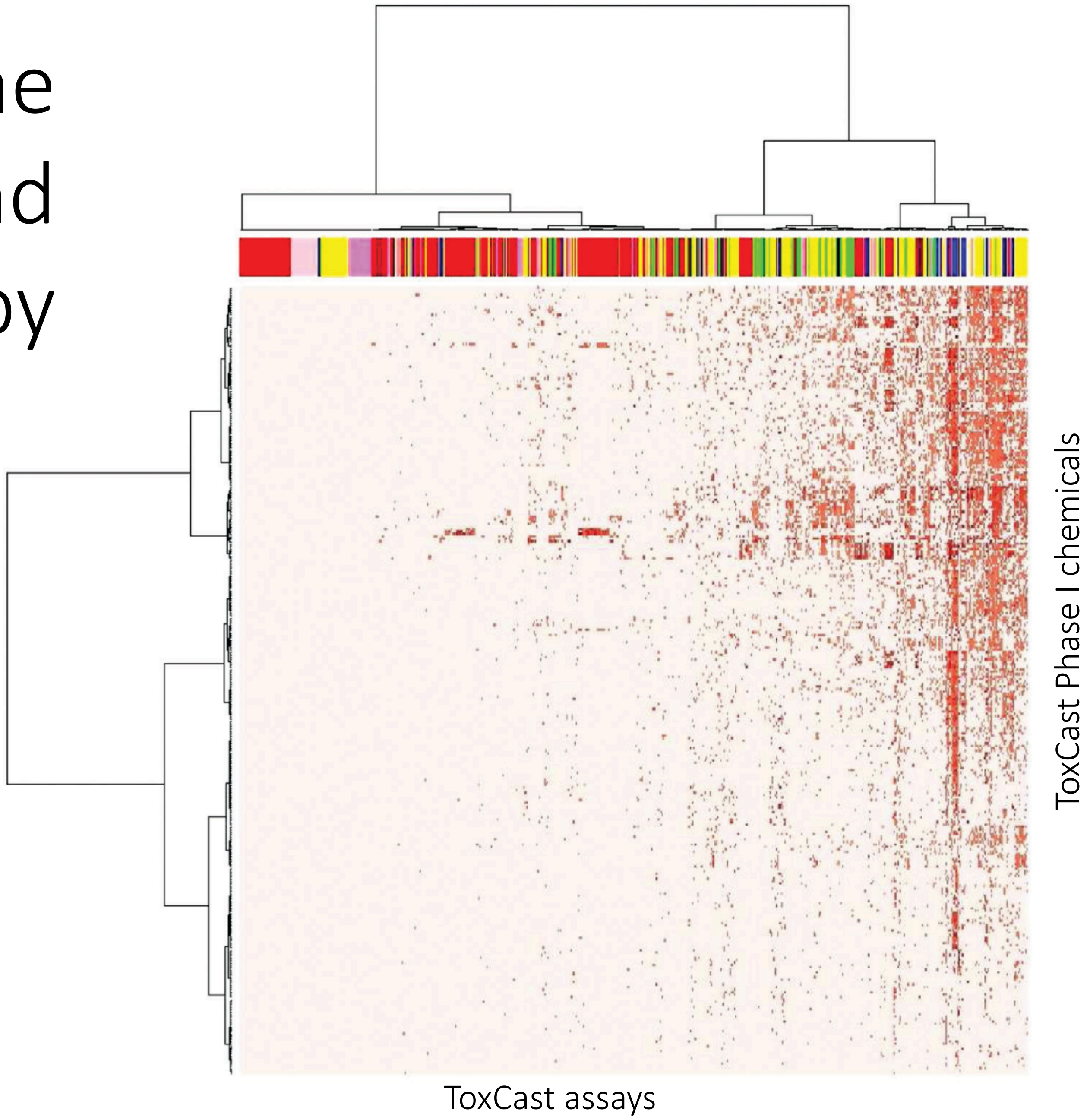
Fredrik Svensson,*,† Azedine Zoufir,† Samar Mahmoud,† Avid M. Afzal,† Ines Smit,† Kathryn A. Giblin,† Peter J. Clements,‡ Jerome T. Mettetal,§ Amy Pointon,‖ James S. Harvey,‡ Nigel Greene,⊥ Richard V. Williams,# and Andreas Bender*,†



Modulation of translocation protein (TSPO) → Production of ROS and increased oxidative stress (ROS) → (Prolonged) opening of mitochondrial permeability transition pore (mPTP) → Cardiomyocyte apoptosis (APOP) → Dilated cardiomyopathy → Valvular annular stretching → Mitral valve incompetence (MVI)

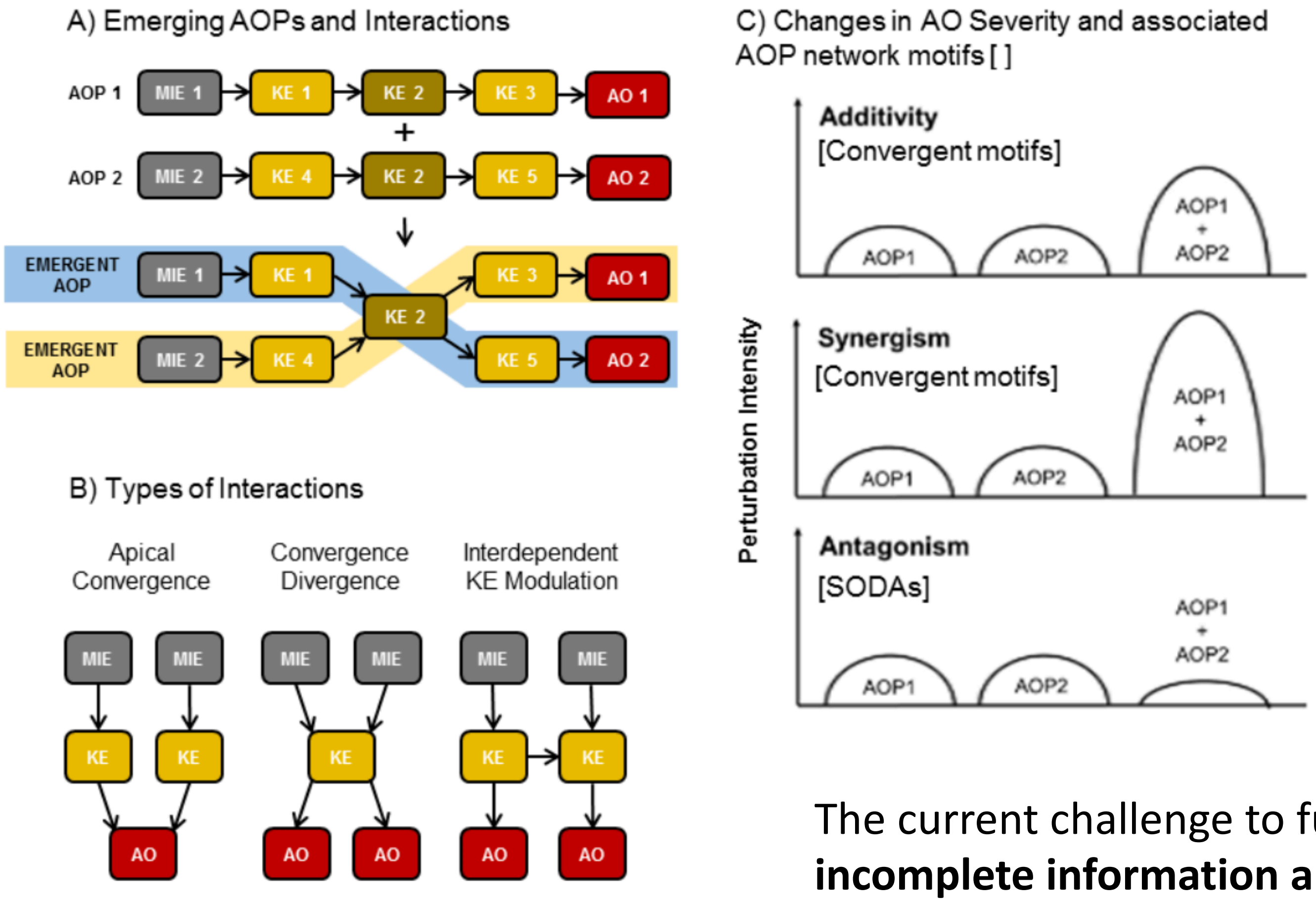
Compounds rarely have one single biological action and are rather characterized by polypharmacology profiles

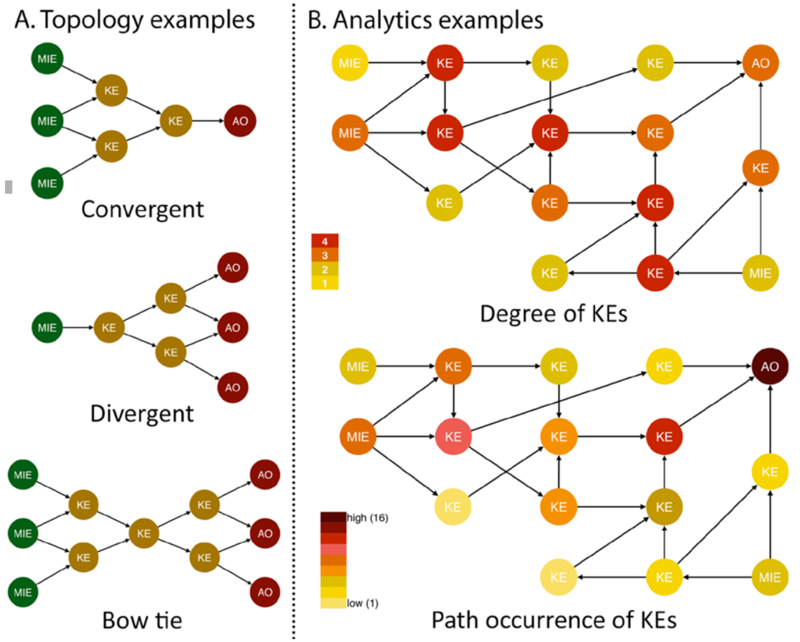




Judson, R. S. *et al* (2010). *Environ. Health Perspect. 118*, 485–492.

AOP networks are useful to analyze interactions, but limited by incomplete information





The current challenge to fully utilize the AOP network framework is the **incomplete information about KE/MIE** linked by KER towards adverse outcomes

Daniel Villeneuve *et al Environ. (2018) Toxicol. Chem*

Toxicity is explained by combinations of features represented as chemical and bioactivity properties

Assay$_i$ + *Chemical Feature***

Assay$_k$ + *Chemical Feature* $_m$

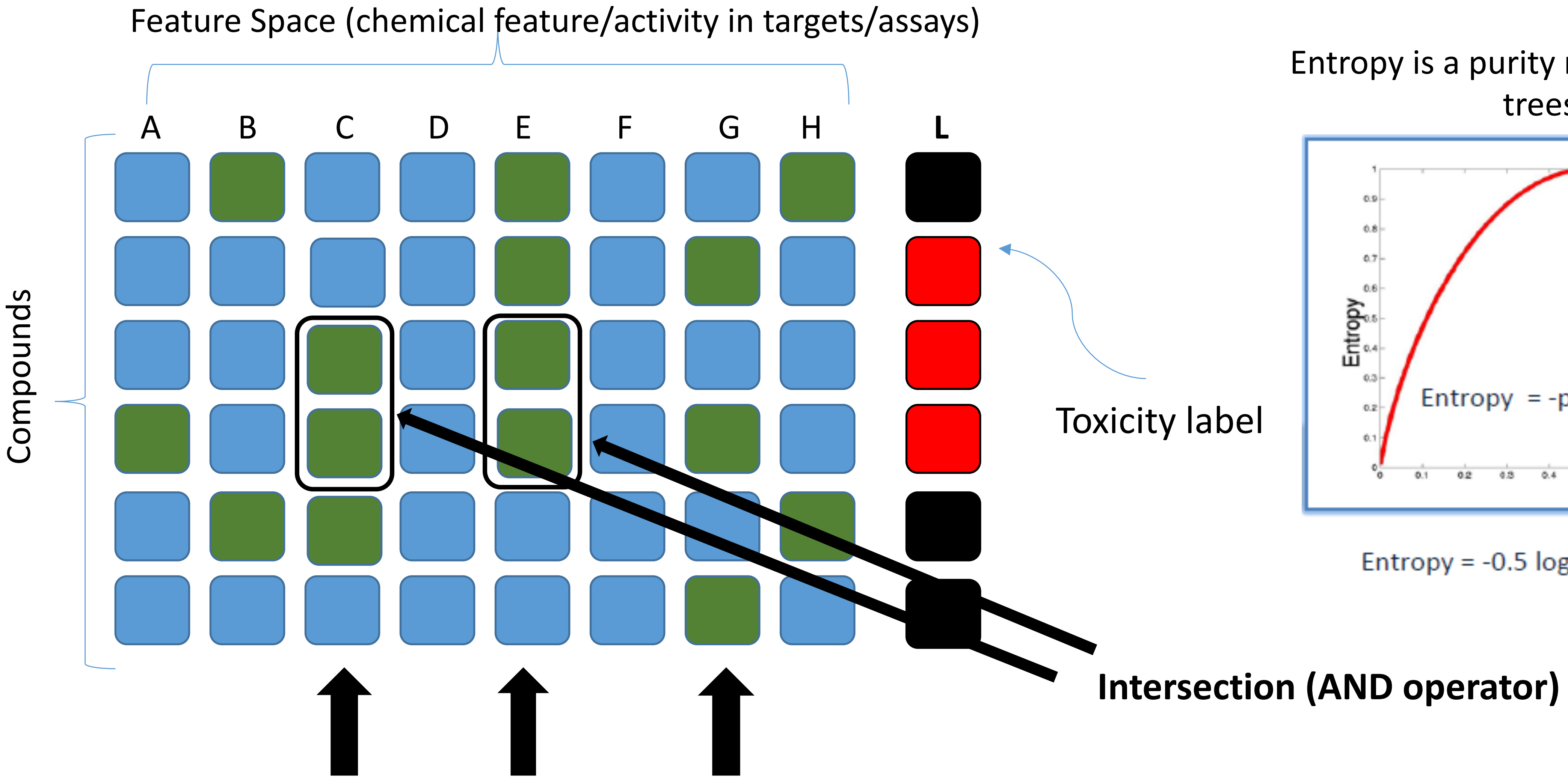
…

→ Toxicity

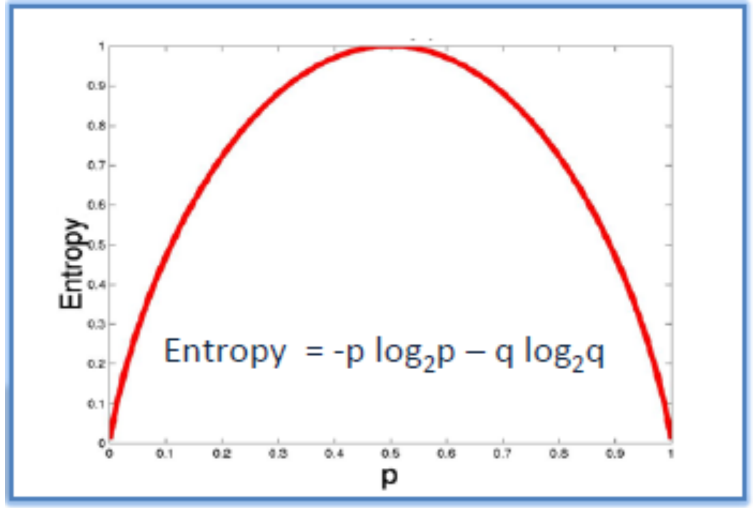
***Each combination constitutes of one or more assay endpoint and can contain one or more chemical properties*

l, m $\in$ F, where F : physicochemical properties or structural alerts

Data-based approach: **Rules**



Entropy is a purity measure to construct trees/rules



$\text{Entropy} = -0.5 \log_2 0.5 - 0.5 \log_2 0.5 = 1$

RULE : If (E) is green **AND** (C) is green → **then** (L) is red

Conventional rules do not respect direction of association

Assumption **Positive Activity in Assay (key event)** → **Toxicity**

Example

Mitochondrial impairment → Toxicity - - - - - - - **Biologically meaningful**

Not cytotoxic → Toxicity - - - - - - - **Biologically irrelevant**

Molecular ❯ Cellular ❯ Tissue ❯ Organ ❯ Individual

MIE → **KE** → **KE** → **KE** → **AO**

Key Event Relationships (KERs) in Adverse Outcome Pathways are directional

Biologically meaningful information can be generated by controlling direction of data-derived associations

Assumption **Positive Activity in Assay (key event)** → **Toxicity**

Example

Mitochondrial impairment → Toxicity - - - - - - - **Biologically meaningful**

Not cytotoxic → Toxicity - - - - - - - **Biologically irrelevant**

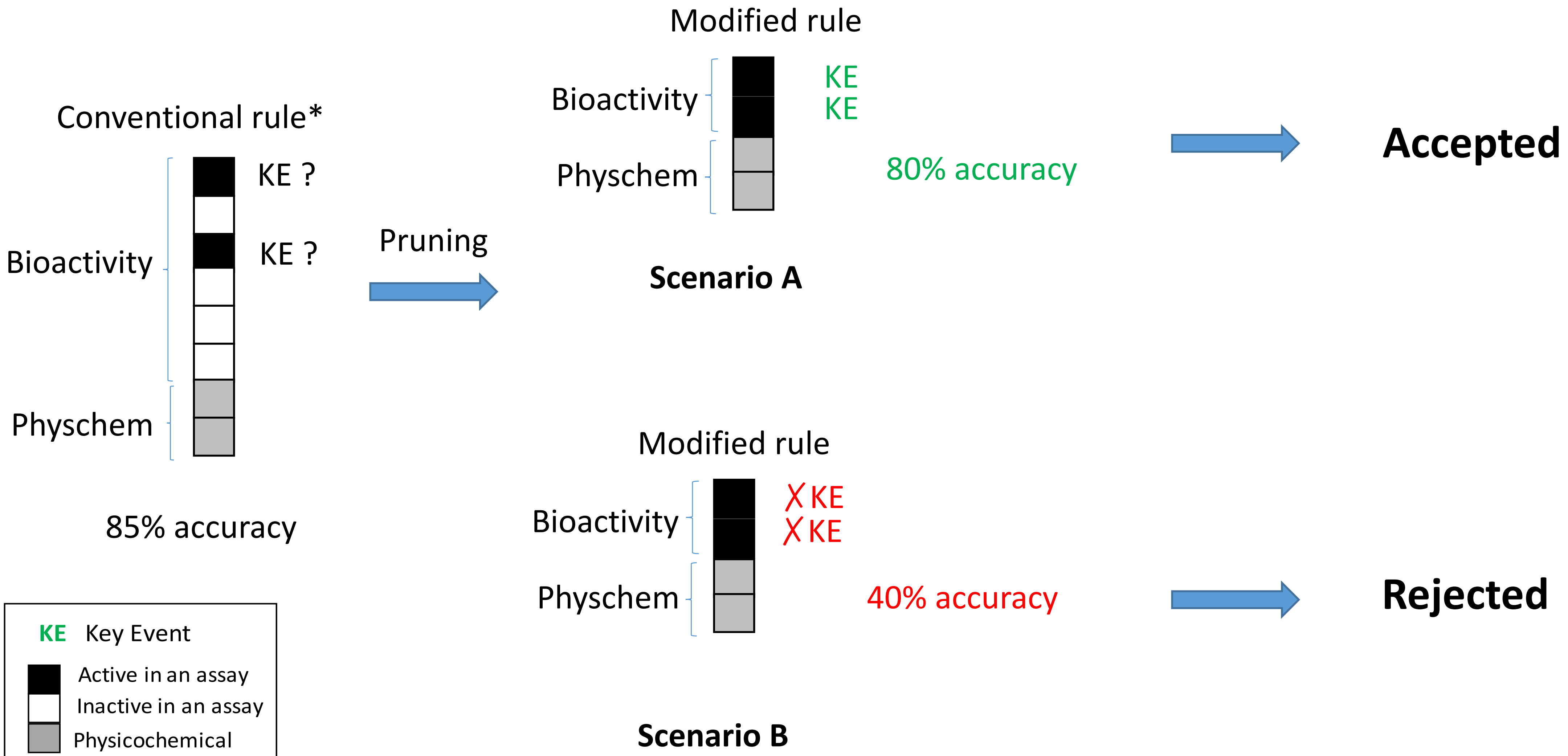
$$\left(\begin{array}{l} \textbf{Assay}_i \ + \textit{chemical property} \\ \textbf{Assay}_k \ + \textit{chemical property} \\ \quad \ldots \end{array} \right) \rightarrow \text{Toxicity}$$

Direction of in vitro-in vivo association is essential for interpretability!

Two rule-based workflows are proposed to mine associations with constraints

- 1- Rule models on **continuous data inputs** via modification of conventional rule models
- 2- Rules models on **binary/categorical** variables using controlled emerging patterns

Approach 1: Rule Pruning is applied to satisfy directional association



* Underlying algorithm is C5.0

Approach 2: Controlled Emerging Patterns using binary features

These represent frequent itemsets that are more common in one class in comparison to the other (discriminating itemsets)
The pattern can be composed of one or more discriminating features

JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Emerging Pattern Mining To Aid Toxicological Knowledge Discovery

Richard Sherhod,[†,‖] Philip N. Judson,[‡] Thierry Hanser,[§] Jonathan D. Vessey,[§] Samuel J. Webb,[§] and Valerie J. Gillet*[,†]

Hypothetical dataset containing the pattern {a,c} emerging in D1.

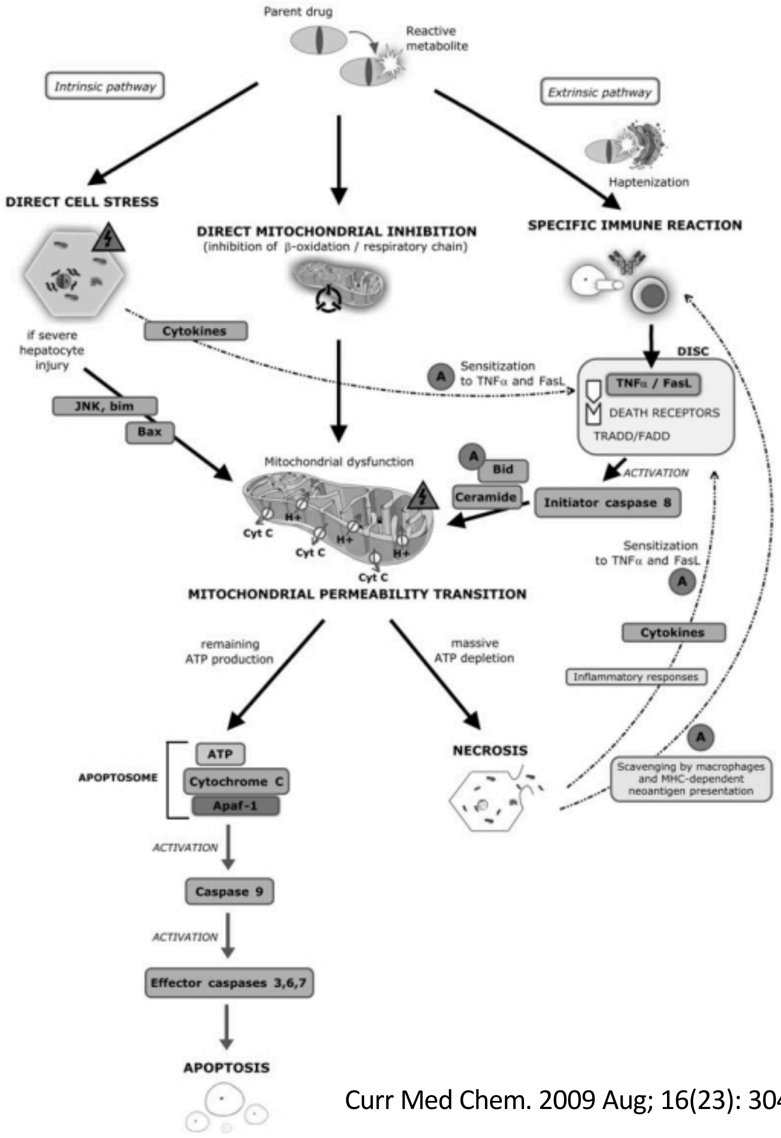
	data entry	properties			
D_1	1	**a**	b	**c**	d
	2	**a**	b	**c**	
	3	**a**		**c**	
	4	a	b		d
	5		b	c	d
D_2	6	**a**		**c**	d
	7			c	d
	8		b		d
	9			c	
	10	a			

J. Chem. Inf. Model. 2014, 54, 1864−1879

Example (1)

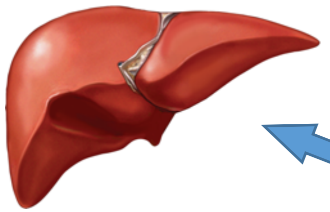
prioritizing hepatotoxic *in vitro* endpoints (modified rules on continuous features)

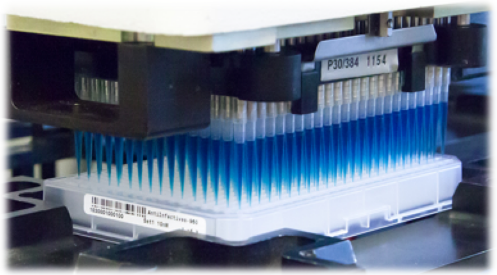
Hepatotoxicity involves complex pathological pathways which are difficult to capture



Curr Med Chem. 2009 Aug; 16(23): 3041–3053.

In vitro models for hepatotoxicity have high specificity but **low sensitivity**





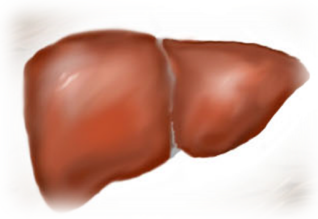
- **Which bioactivities in lab are predictive of hepatotoxicity *in vivo*?**
- **How chemical properties affect the concordance between *in vitro* measurements and *in vivo* observations?**

DATASET

ToxCast *in vitro* readouts in AC_{50}*
~8000 compounds against over 800 assays

Open-Source Cheminformatics and Machine Learning

29 calculated physicochemical properties
(lipophilicity, molecular weight, number of rings, etc.)

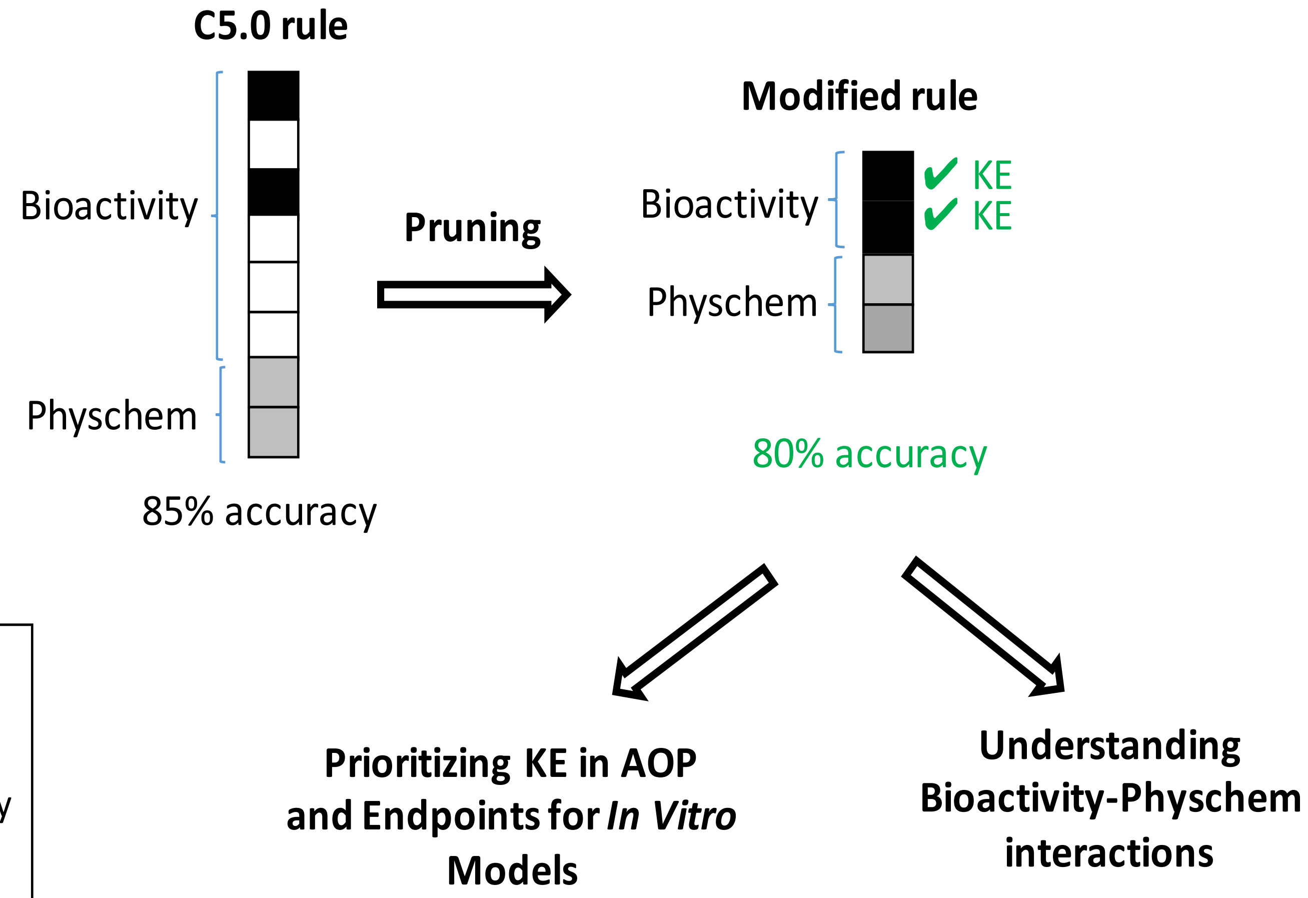


ToxRefDB rat liver observations
~900 compounds against 17 liver measurements
Discretized into LEL** of 15mg/kg/day and 500mg/kg/day

Data integration of multiple sources: ~6million data points of 673 compounds

* Concentration at half maximum activity
** Lowest Effect Level

Analysis performed using modified rule workflow



Average number of conditions per toxic rule in the original set

	Condition type in toxic rules		
Toxicity threshold	*Active in an assay*	*Inactive in an assay*	*physicochemical*
15mg/kg/day	1	3.8	0.9
500mg/kg/day	1	3.6	0.6

Overall per rule, there is one positive bioactivity, four negative bioactivities and one physicochemical property. The abundance of inactive assay conditions and physicochemical conditions is slightly lower at toxicity threshold 500mg/kg/day.

Rule modification to improve interpretability

Original rule:

APR_HepG2_MitoMembPot_72h_up > 2.036928
Tox21_HSE_BLA_agonist_ratio > 2.40309
Tox21_p53_BLA_p5_viability <= 0.02595656
AMW > 192.001
NumAromaticHeterocycles <= 1

>> class toxic

Pruning →

Modified rule:

Tox21_p53_BLA_p5_viability <= 0.02595656
AMW > 192.001
NumAromaticHeterocycles <= 1

>> class toxic

Workflow overview

Rule modification

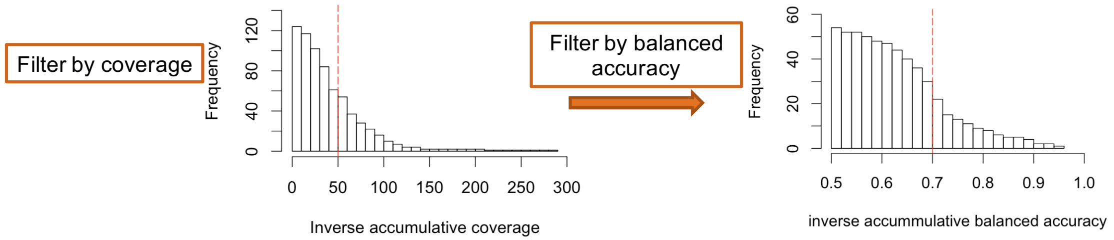
APR_HepG2_NuclearSize_24h_up {Active}
Strongest.basic.pKa <= 6.48 → Hepatotoxic

Modifications makes rules simple and interpretable

↓

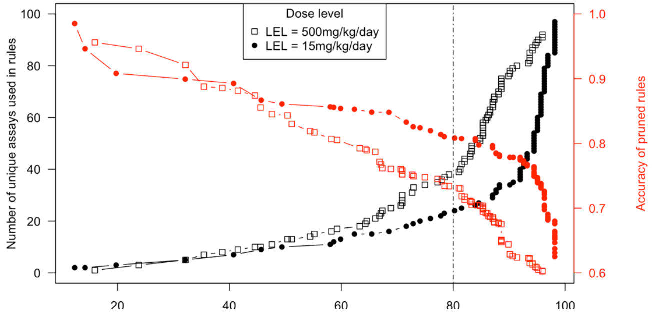
Rule prioritization

i) By performance



This selection retains rules with top accuracy (above 70%) and high rule coverage (above 50 and 20 for 500mg/kg/day and 15mg/kg/day, respectively)

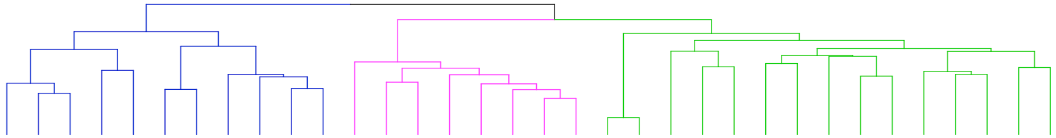
ii) By overall coverage of toxic compounds



The combination of top accurate rules, which represents 80% of all toxic compounds, is prioritized. This resulted in 34 and 20 rules at 500mg/kg/day and 15mg/kg/day, respectively

↓

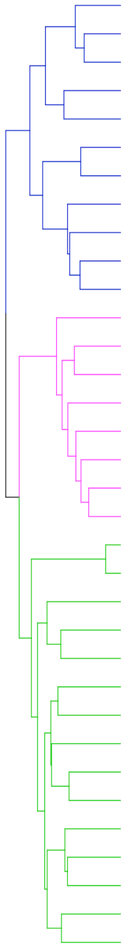
Clustering



Prioritized rules are clustered according to similarity in toxic compound coverage

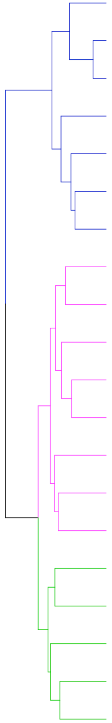
Bioactivity space for different levels of toxic potency is similar

500 mg/kg/day



Bioactivity class	Associated assay
Activity against Cytochrome P	APR_HepG2_MitoMass_24h_up
	ATG_PPARg_TRANS_up
	⌊ OT_AR_ARSRC1_0480
	NVS_ADME_hCYP2C18
	NVS_ADME_hCYP2C19
	⌊ NVS_TR_hDAT
	NVS_ADME_rCYP3A1
	NVS_ADME_rCYP3A2
	NVS_MP_hPBR
	NVS_NR_hCAR_Antagonist
	OT_FXR_FXRSRC1_0480
Immunological activity	APR_HepG2_CellCycleArrest_72h_dn
	⌊ Tox21_FXR_BLA_antagonist_ratio
	BSK_BE3C_uPA_down
	BSK_KF3CT_IP10_down
	BSK_KF3CT_MMP9_down
	BSK_LPS_CD40_down
	⌊ BSK_3C_IL8_down
	BSK_LPS_MCP1_down
	BSK_SAg_CD40_down
	BSK_SAg_SRB_down
Nuclear receptor activity/ phenotypic readouts	APR_HepG2_MitoMembPot_72h_up
	APR_HepG2_MitoMembPot_1h_dn
	⌊ Tox21_AR_BLA_Antagonist_ratio
	APR_HepG2_NuclearSize_24h_up
	APR_HepG2_OxidativeStress_1h_up
	APR_HepG2_StressKinase_1h_up
	ATG_BRE_CIS_up
	ATG_C_EBP_CIS_up
	⌊ ATG_HIF1a_CIS_up
	ATG_CRE_CIS_up
	ATG_FoxA2_CIS_up
	BSK_SAg_PBMCCytotoxicity_up
	Tox21_ERa_LUC_BG1_Agonist
	Tox21_GR_BLA_Antagonist_ratio
	Tox21_MitochondrialToxicity_viability
	⌊ ATG_p53_CIS_up

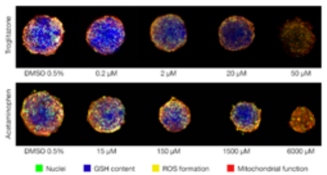
15 mg/kg/day



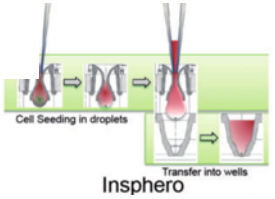
Rule cluster/key mechanism	Associated assay
Activity against Cytochrome P	BSK_LPS_PGE2_up
	NVS_ADME_hCYP2C19
	NVS_ADME_rCYP2A1
	⌊ OT_AR_ARSRC1_0480
	NVS_ADME_rCYP2C12
	NVS_ADME_rCYP2C13
	NVS_ADME_rCYP2C6
	⌊ ATG_VDRE_CIS_up
Immunological activity/Endocrine disruption	BSK_3C_ICAM1_down
	BSK_4H_MCP1_down
	BSK_BE3C_MIG_down
	⌊ Tox21_ERa_BLA_Agonist_ratio
	BSK_hDFCGF_IP10_down
	⌊ Tox21_MitochondrialToxicity_viability
	BSK_SAg_CD40_down
	Tox21_AR_BLA_Antagonist_viability
	⌊ Tox21_PPARd_BLA_antagonist_ratio
	Tox21_ERa_BLA_Antagonist_ratio
Nuclear receptor activity	APR_HepG2_MitoMass_72h_up
	⌊ APR_HepG2_NuclearSize_24h_up
	ATG_LXRb_TRANS_up
	NVS_TR_hDAT
	OT_ER_ERbERb_0480
	Tox21_FXR_BLA_agonist_ratio

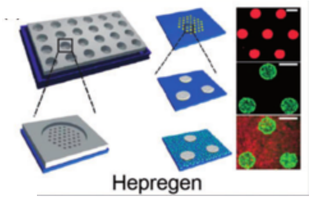
- Three key clusters; Cytochrome P, immunological responses and nuclear receptor activities
- Multiple bioactivities were described in rules

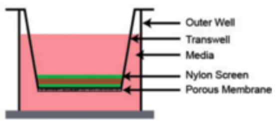
Endpoints used commercial setups are captured in rules, except for endocrine disruption



Cyprotex







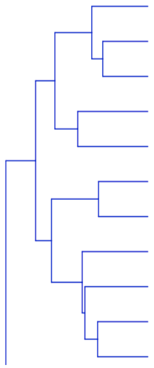
Regenemed

***In vitro* systems**	Metabolism	Viability/phenotypic changes	Cell stress	Immune response	Bile transport	Inhibition of protein synthesis	Mitochondrial impairment	Endocrine activity
Cyprotex CellCiphr®		Apoptosis Cell cycle arrest Cell loss Cytoskeletal disruption DNA fragmentation and damage response Mitosis marker Nuclear size Phospholipidosis Steatosis	Glutathione depletion Oxidative stress Stress kinase activation Reactive oxygen species				Mitochondrial function	
InShero 3D Insight™	Cytochrome activity	Apoptosis	Glutathione depletion	IL-6 release	BSEP	Albumin	Intra-tissue ATP	
Hepregen Hepatopac	Metabolites Clearance		Glutathione levels		MRP2 CMFDA	Albumin, urea	ATP MTT	
RegeneMed	Cytochrome activity Clearance		Glutathione	Cytokine profile		Albumin, urea, fibrinogen, transferrin	ATP	
Modified rules	**Cytochrome activity**	**Cell cycle arrest Cytotoxicity Nuclear size Mitochondrial mass**	**Oxidative stress Stress kinase**	**IL-9, IL-10,CCL2 and CD40**	**FXR**	**Cellular protein content**	**Mitochondrial membrane potential/toxicity**	**Estrogen and androgen receptors activity**

Sundhar et al. In Vitro Platforms for Evaluating Liver Toxicity. *Exp Biol Med*. 2014

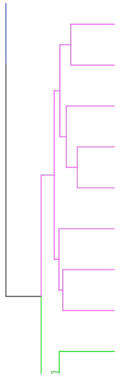
Combined bioactivity readouts in rules

500 mg/kg/day



Bioactivity class	Associated assay
Activity against Cytochrome P	APR_HepG2_MitoMass_24h_up
	ATG_PPARg_TRANS_up
	⎣ OT_AR_ARSRC1_0480
	NVS_ADME_hCYP2C18
	NVS_ADME_hCYP2C19
	⎣ NVS_TR_hDAT
	NVS_ADME_rCYP3A1
	NVS_ADME_rCYP3A2
	NVS_MP_hPBR
	NVS_NR_hCAR_Antagonist
	OT_FXR_FXRSRC1_0480

15 mg/kg/day

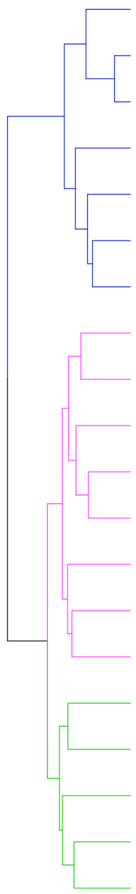


Bioactivity class	Associated assay
Immunological activity/Endocrine disruption	BSK_3C_ICAM1_down
	BSK_4H_MCP1_down
	BSK_BE3C_MIG_down
	⎣ Tox21_ERa_BLA_Agonist_ratio
	BSK_hDFCGF_IP10_down
	⎣ Tox21_MitochondrialToxicity_viability
	BSK_SAg_CD40_down
	Tox21_AR_BLA_Antagonist_viability
	⎣ Tox21_PPARd_BLA_antagonist_ratio
	Tox21_ERa_BLA_Antagonist_ratio

At both levels, rules combined the activity against AR and PPAR. There is bidirectional crosstalk between the AR and PPAR, by which each can influence the expression as well as the transcriptional activity of the other.

⎣ represent assays co-occurred in one rule

Combined bioactivity readouts in rules



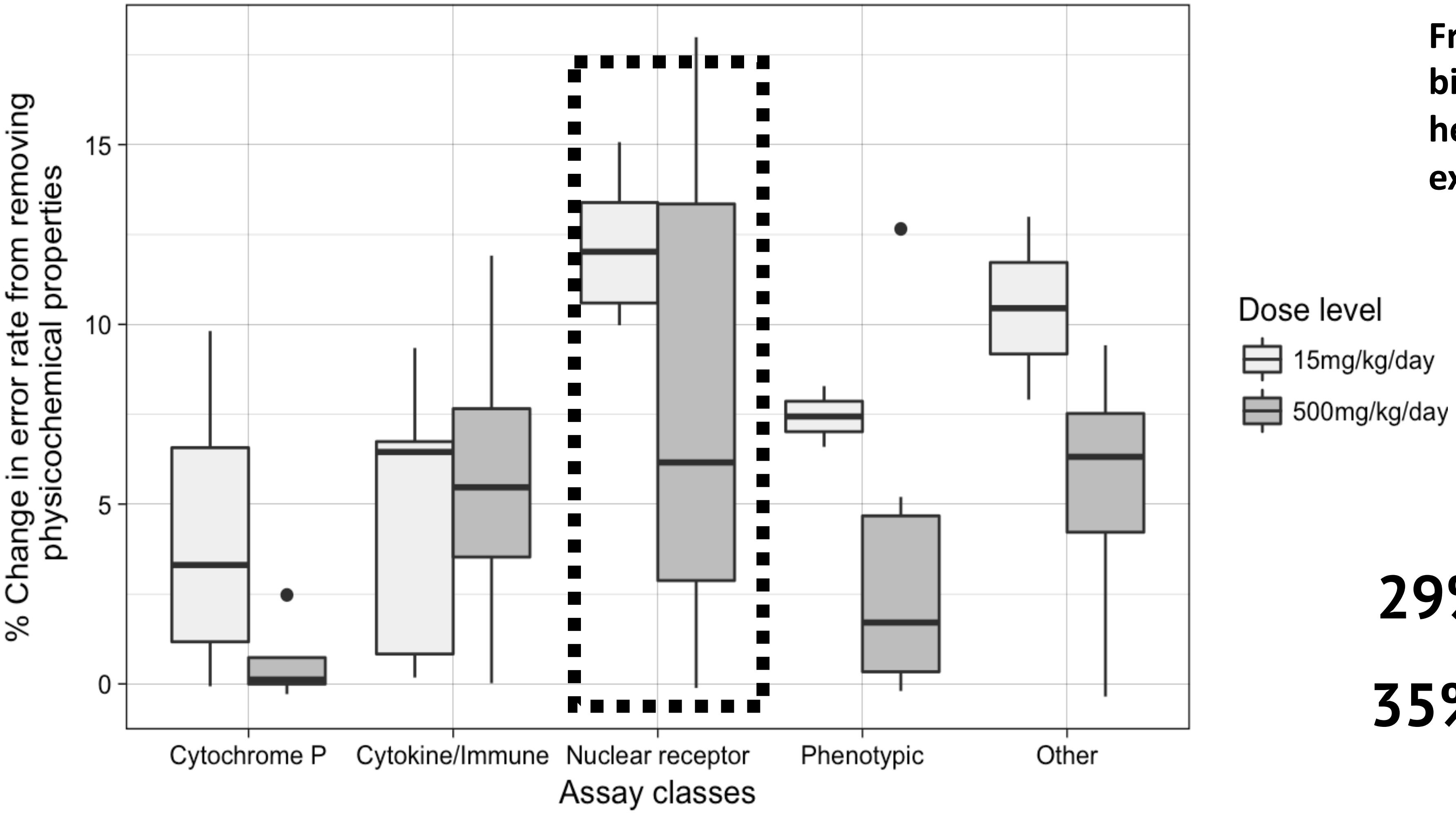
Rule cluster/key mechanism	Associated assay
Activity against Cytochrome P	BSK_LPS_PGE2_up
	NVS_ADME_hCYP2C19
	NVS_ADME_rCYP2A1
	⌊ OT_AR_ARSRC1_0480
	NVS_ADME_rCYP2C12
	NVS_ADME_rCYP2C13
	NVS_ADME_rCYP2C6
	⌊ ATG_VDRE_CIS_up
Immunological activity/Endocrine disruption	BSK_3C_ICAM1_down
	BSK_4H_MCP1_down
	BSK_BE3C_MIG_down
	⌊ Tox21_ERa_BLA_Agonist_ratio
	BSK_hDFCGF_IP10_down
	⌊ Tox21_MitochondrialToxicity_viability
	BSK_SAg_CD40_down
	Tox21_AR_BLA_Antagonist_viability
	⌊ Tox21_PPARd_BLA_antagonist_ratio
	Tox21_ERa_BLA_Antagonist_ratio
Nuclear receptor activity	APR_HepG2_MitoMass_72h_up
	⌊ APR_HepG2_NuclearSize_24h_up
	ATG_LXRb_TRANS_up
	NVS_TR_hDAT
	OT_ER_ERbERb_0480
	Tox21_FXR_BLA_agonist_ratio

- At 15mg/kg/day, multiple assay combinations predictive for hepatotoxicity can be seen including CYP2C6 with VDR and CXCL-9 with ER agonists
- In response to xenobiotics, VDR directly induces the upregulation of CYP2C6. Hence, compounds that combine activity against CYP2C6 and upregulation of VDR are likely to cause hepatotoxicity
- Studies have shown links between ER agonists and CXCL9, at which estrogen-treated mice have shown a significant reduction in the expression of CXCL9, a cytokine associated with liver fibrosis

⌊ represent assays co-occurred in one rule

Physicochemical Properties Improve Translatability Of *In Vitro* Measurements Into *In Vivo* Outcomes

Influence on rule accuracy



Frequent properties are linked to bioavailability according to literature, hence can be used as proxy for exposure/C_{max}

Most frequent

29% **Number of rings** at 500mg/kg/day

35% **Number of rotatable bonds** at 15mg/kg/day

Number of rotatable bonds is associated with permeability

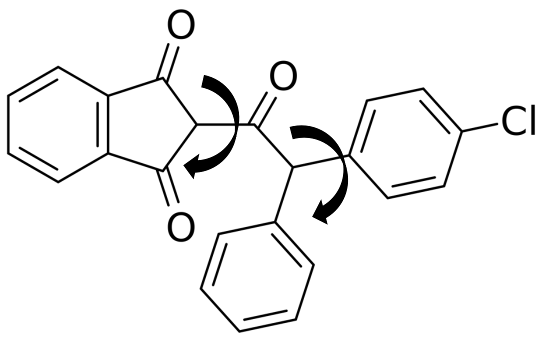
Molecular Properties That Influence the Oral Bioavailability of Drug Candidates

Daniel F. Veber,*,† Stephen R. Johnson,‡,§ Hung-Yuan Cheng,‖ Brian R. Smith,⊥ Keith W. Ward,⊥ and Kenneth D. Kopple‡,#

Departments of Medicinal Chemistry, Cheminformatics, Computational Analytical and Structural Sciences, and Drug Metabolism and Pharmacokinetics, GlaxoSmithKline, 709 Swedeland Road, P. O. Box 1539, King of Prussia, Pennsylvania 19406-0939

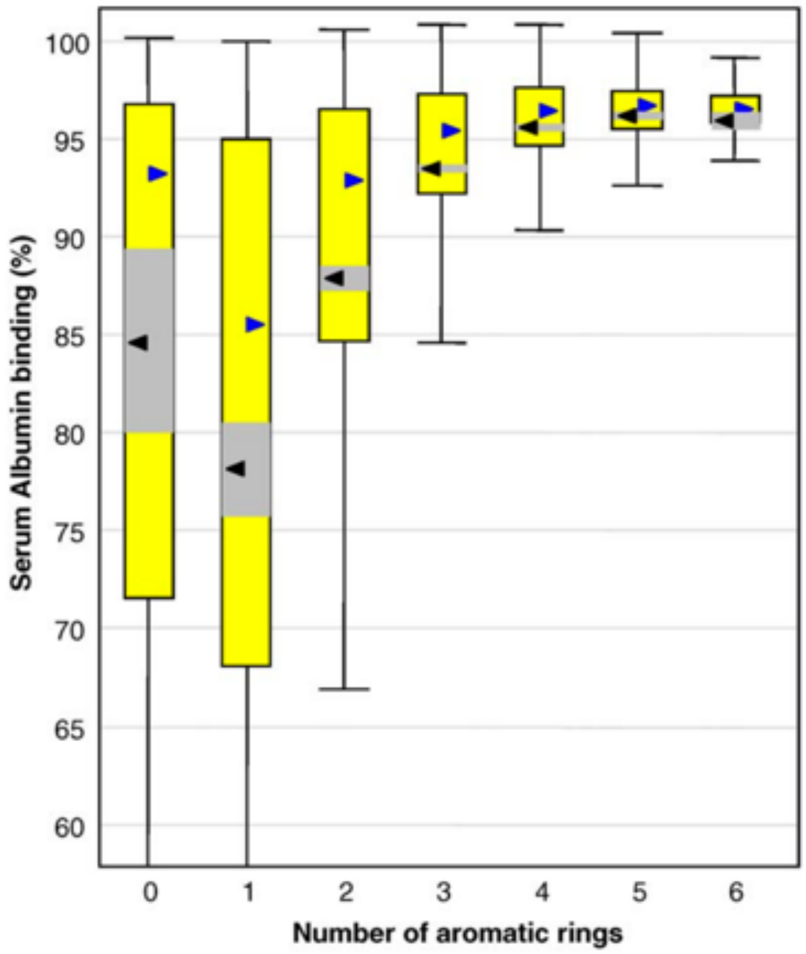
Received January 9, 2002

Oral bioavailability measurements in rats for over 1100 drug candidates studied at SmithKline Beecham Pharmaceuticals (now GlaxoSmithKline) have allowed us to analyze the relative importance of molecular properties considered to influence that drug property. Reduced molecular flexibility, as measured by the number of rotatable bonds, and low polar surface area or total hydrogen bond count (sum of donors and acceptors) are found to be important predictors of good oral bioavailability, independent of molecular weight. That on average both the number of rotatable bonds and polar surface area or hydrogen bond count tend to increase with molecular weight may in part explain the success of the molecular weight parameter in predicting oral bioavailability. The commonly applied molecular weight cutoff at 500 does not itself significantly separate compounds with poor oral bioavailability from those with acceptable values in this extensive data set. Our observations suggest that compounds which meet only the two criteria of (1) 10 or fewer rotatable bonds and (2) polar surface area equal to or less than 140 Å^2 (or 12 or fewer H-bond donors and acceptors) will have a high probability of good oral bioavailability in the rat. Data sets for the artificial membrane permeation rate and for clearance in the rat were also examined. Reduced polar surface area correlates better with increased permeation rate than does lipophilicity (*C* log *P*), and increased rotatable bond count has a negative effect on the permeation rate. A threshold permeation rate is a prerequisite of oral bioavailability. The rotatable bond count does not correlate with the data examined here for the in vivo clearance rate in the rat.



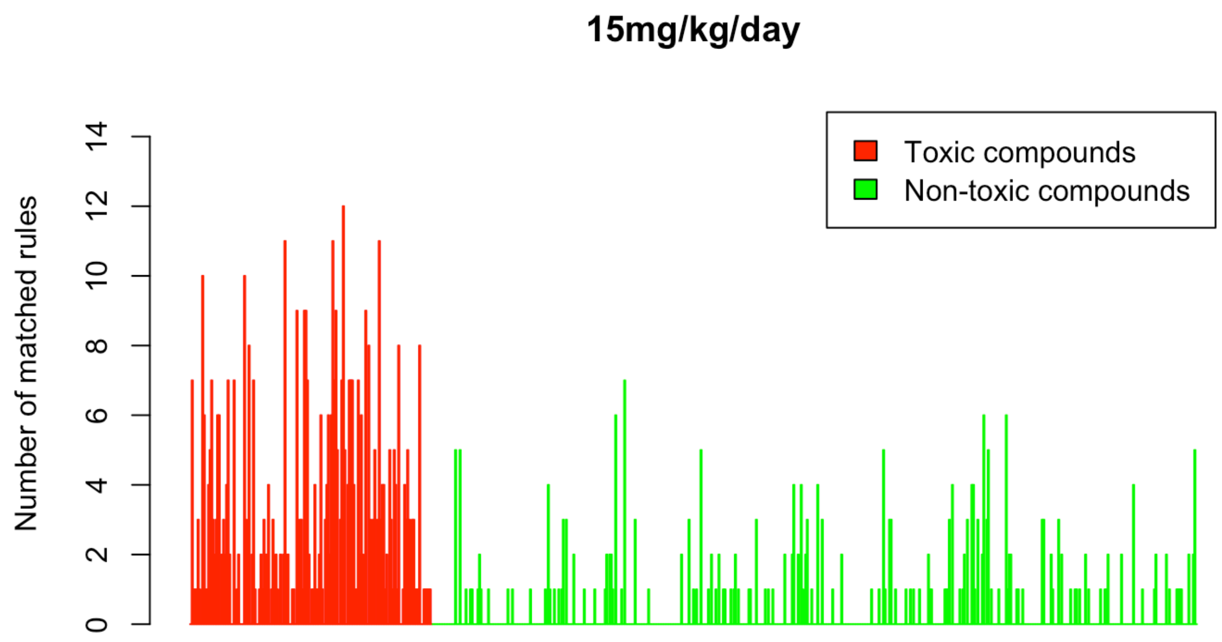
	15mg/kg/day	
Physicochemical condition	Error rate%	Frequency %
NumRotatableBonds <= 6	7.8 ± 3.2	35
NumHBD <= 0	9.2 ± 3.7	10
NumAliphaticRings <= 2	2.7 ± 0.3	10

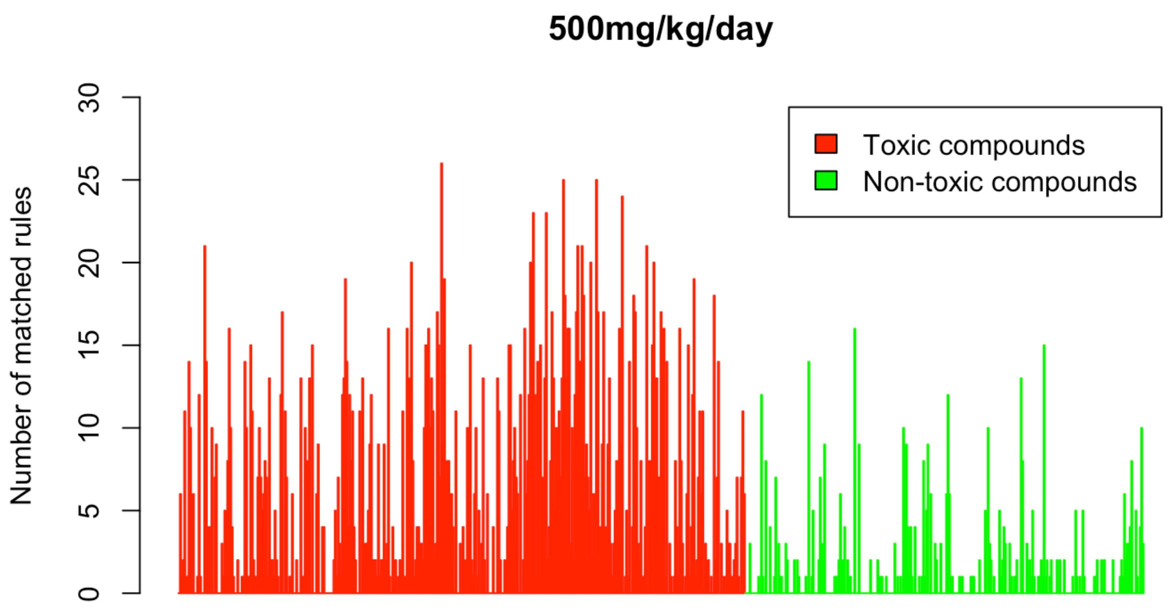
Number of rings is associated with plasma protein binding



	500mg/kg/day	
Physicochemical condition	Error rate %	Frequency %
NumRings <= 3	5.7±3.6	29
NumHeavyAtoms <= 33	3.9±0.5	11
NumAromaticCarbocycles > 0	11.5±1.9	9

Toxic compounds match significantly more rules than non-toxic compounds





Can Animal models capture effects in human?

Review

Are animal models predictive for humans?

Niall Shanks[1], Ray Greek*[2] and Jean Greek[2]

Journal of Biomedical Informatics 54 (2015) 167–173

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Prediction of clinical risks by analysis of preclinical and clinical adverse events

Matthew Clark

Elsevier Life Science Solutions, 1600 John F. Kennedy Blvd, Suite 1800, Philadelphia, PA 19103, United States

Drug Discovery Today

Volume 22, Issue 1, January 2017, Pages 127-132

Review

Post-screen

Predicting toxicities in humans by nonclinical safety testing: an update with particular reference to anticancer compounds

Varun Ahuja, Sanjay Bokan, Sharad Sharma

ATLA 43, 393–403, 2015 393

Predicting Human Drug Toxicity and Safety via Animal Tests: Can *Any* One Species Predict Drug Toxicity in Any Other, and Do Monkeys Help?

Jarrod Bailey,[1] Michelle Thew[1] and Michael Balls[2]

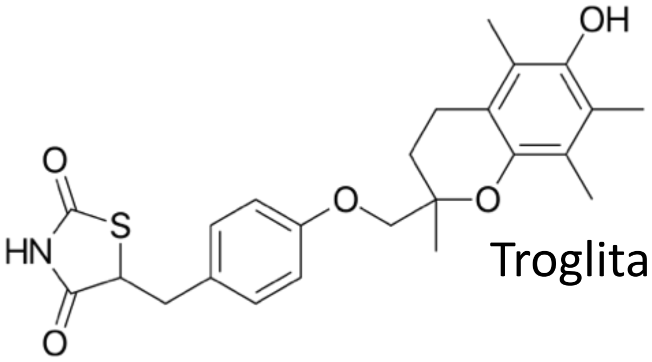
[1]Cruelty Free International, London, UK; [2]c/o Fund for the Replacement of Animals in Medical Experiments (FRAME), Nottingham, UK

Key factor : animal testing studies do not extrapolate well to human!

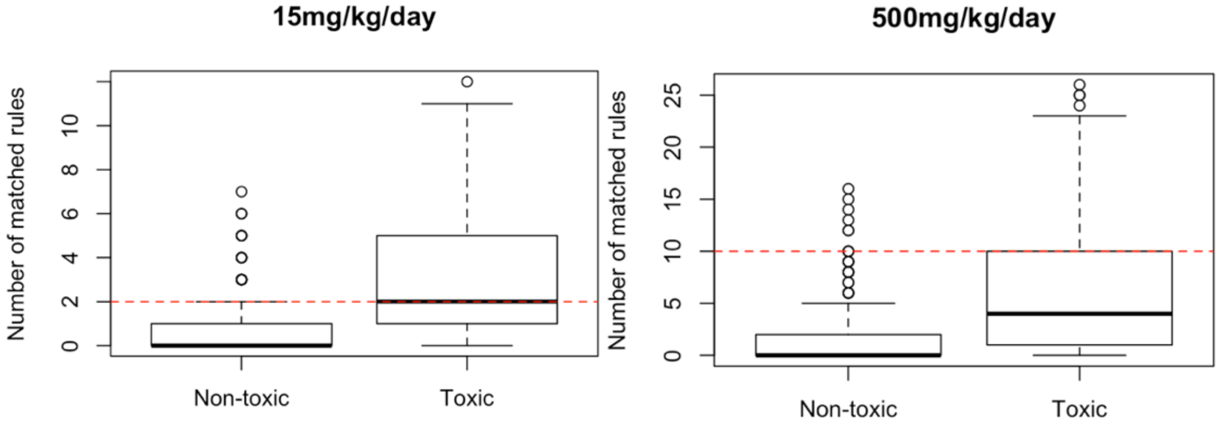
Data from animal testing can be used to infer molecular mechanisms of adverse effects in human

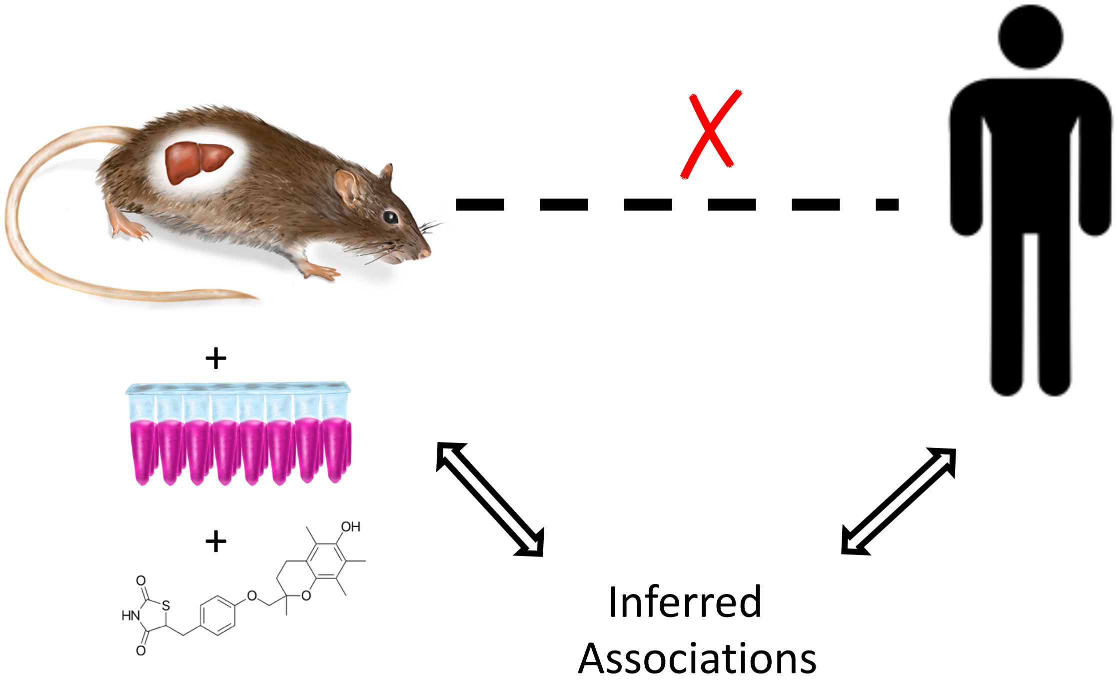
Troglitazone was withdrawn in 2000 due to hepatotoxicity

Troglitazone is labelled as toxic at 500mg/kg/day, but not at 15mg/kg/day



Troglitazone





According to rules, troglitazone has more liabilities than 75% of toxic compounds at 500mg/kg/day and equal to average at 15mg/kg/day

Example (2)

Understanding Polypharmacology in Acute Toxicity (controlled emerging patterns on binary features)

Possible mechanisms are complex and diverse

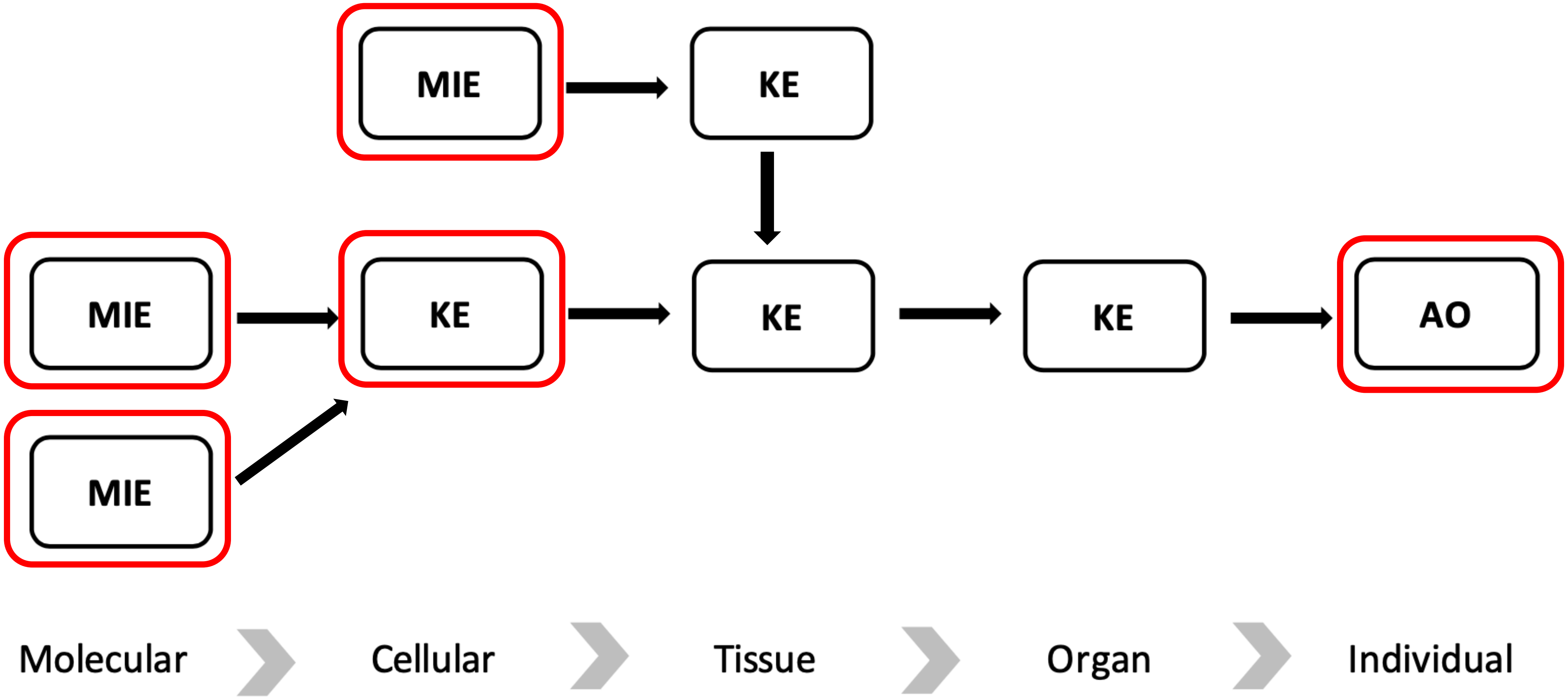
Hamm et al (2017) have suggested multiple routes for the mechanisms for acute toxicity as a result of the acute toxicity workshop in Maryland, USA in 2015

Selected mechanisms of acute toxicity.[1]

MIE or upstream key event	Example stressor	Relevant AOP
GABA receptor inhibition	Fipronil	Binding to the picrotoxin site of ionotropic GABA receptors leading to epileptic seizures[a]
Sodium channel inhibition	Pyrethroids	Axonal sodium channel modulation leading to acute mortality[b]
Protein synthesis inhibition	Ricin	
Sodium-potassium ATPase inhibition	Digoxin	
Mitochondrial inhibition	2-Buten-1-ol, 1-thenyl-4,4,4-trifluoro-3-trifluoromethyl-	
Binding of benzodiazepine sites on GABA receptor	Tetrazepam	
Acetylcholinesterase inhibition	4-(Methylamino)-3,5-xylyl methylcarbamate	Acetylcholinesterase inhibition leading to acute mortality[c]
GSH depletion followed by covalent binding of reactive metabolite to cellular proteins	Acetaminophen	
Michael acceptor reaction	Acrolein	
Voltage-gated sodium channel inhibition	Sodium valproate	
NMDA receptor antagonism	Methadone	
Anticoagulation	Coumadin	
Dopaminergic D2 receptor antagonism	Thioridazine hydrochloride	

[1] This table provides an outline of the some of the known mechanisms involved in acute systemic toxicity along with prototypical initiators. In some cases, the exact molecular initiating event (MIE) isn't known. Examples of adverse outcome pathways (AOPs) under development in the OECD AOP Wiki are noted and can be found on the web: a) https://aopwiki.org/wiki/index.php/Aop:10 b) https://aopwiki.org/wiki/index.php/Aop:96; c) https://aopwiki.org/wiki/index.php/Aop:16.

By mining the multi-conditional associations between potential MIEs and KEs against toxicity outcomes, we can generate hypothesis about significant polypharmacology



Data

❑ Toxicity data:

- **Globally Harmonized System**: acute oral toxicity PubChem class 1,2 and 3 are considered toxic, whereas classes 4 and 5 are non toxic

❑ Bioactivity data:

- i) **Tox21** %activity (~9000 compounds of ~900,000 data experimental data points)
- binned into > 20, >40 >60, >80) ~ 4million data points
- ii) **PIDGIN** (in-house tool for target prediction) annotated targets ~ 1800 for

❑ Substructures

- i) **ToxAlerts** using OCHEM server ~2300
- ii) **Frequent substructure** using MoSS in KNIME ~ 450

- **Data integration**: ~7,5millian data points of ~2000 unique compounds almost balanced toxicity label

Rule pattern generation

Binary matrix representation

Entry	*A*	*B*	*C*	*D*	*E*	*F*	*Label*
1	0	1	1	0	1	1	**0**
2	1	0	1	1	0	0	**0**
3	0	1	1	0	1	0	**0**
4	1	0	0	0	0	1	**1**
5	1	0	1	1	0	1	**1**
6	1	1	0	1	0	1	**1**

Transaction lists representation

Entry	*Feature set*	*Label*
1	*B,C,E,F*	**0**
2	*A,C,D*	**0**
3	*B,C,E*	**0**
4	*A,F*	**1**
5	*A,C,D,F*	**1**
6	*A,B,D,F*	**1**

Rules

A, F >> toxic
A,D, F >> toxic

CPAR (Classification based on Predictive Association Rules)

$$\text{Gain} = |P^*| \left(\log\frac{|P^*|}{|P^*| + |N^*|} - \log\frac{|P|}{|P| + |N|}\right)$$

[decay factor of 2/3 , similarity ratio of 1 : 0.99, minimum gain 2.5]

Workflow for controlled emerging patterns

Data
Bioactivity/substructure
Versus toxicity labels
(Tox21,toxalerts, GHS labels
for acute toxicity

Identify the model design matrix
(applying monotonicity constraints to improve interpretability)

	Bioactivity	substructure
Toxic rules	+	+
Non toxic rules	-	+

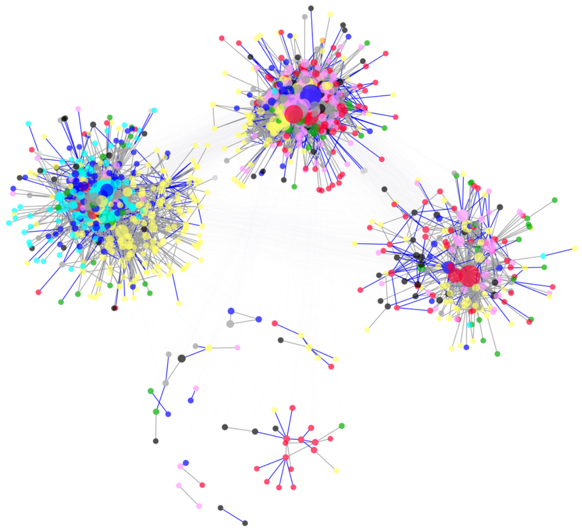
Generate emerging patterns
for toxic and non toxic compounds

{Bioactivity **A** + substructure **S**} >> **Acute toxicity**
{Bioactivity **C** + bioactivity **E**} >> **Acute toxicity**
{Substructure **D** + substructure **J**} >> **Acute toxicity**

Transaction lists representation

Entry	***Feature set***	***Label***
1	*B,C,E,F*	**0**
2	*A,C,D*	**0**
3	*B,C,E*	**0**
4	*A,F*	**1**
5	*A,C,D,F*	**1**
6	*A,B,D,F*	**1**

Analysis



1. Synergy
Mining synergy interactions between features
2. Networks
Clusters, adjacency

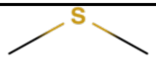
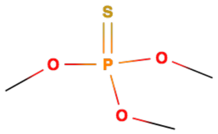
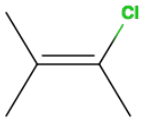
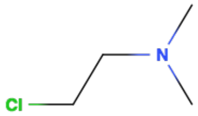
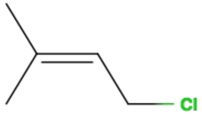
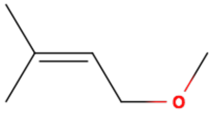
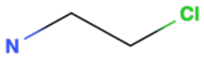
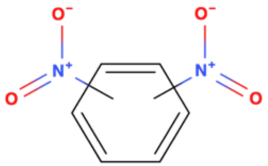
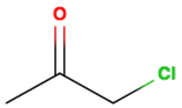
Thousands of rules were generated with an average accuracy (confidence) above 80%

	*N# of all rules**	*Single condition rules*	*Multiple condition rules*	*Accuracy***	*Compound coverage per rule*	*N# of conditions per rule◊*
Toxic rules	9165 (7381)	1267 (566)	7898 (6815)	0.85 ± 0.083	26.3 ± 12.8	2.6 ± 0.83
Non-toxic rules	4613 (3866)	410 (155)	4203 (3711)	0.82 ± 0.082	34.1 ± 20.8	3.3 ± 1.3

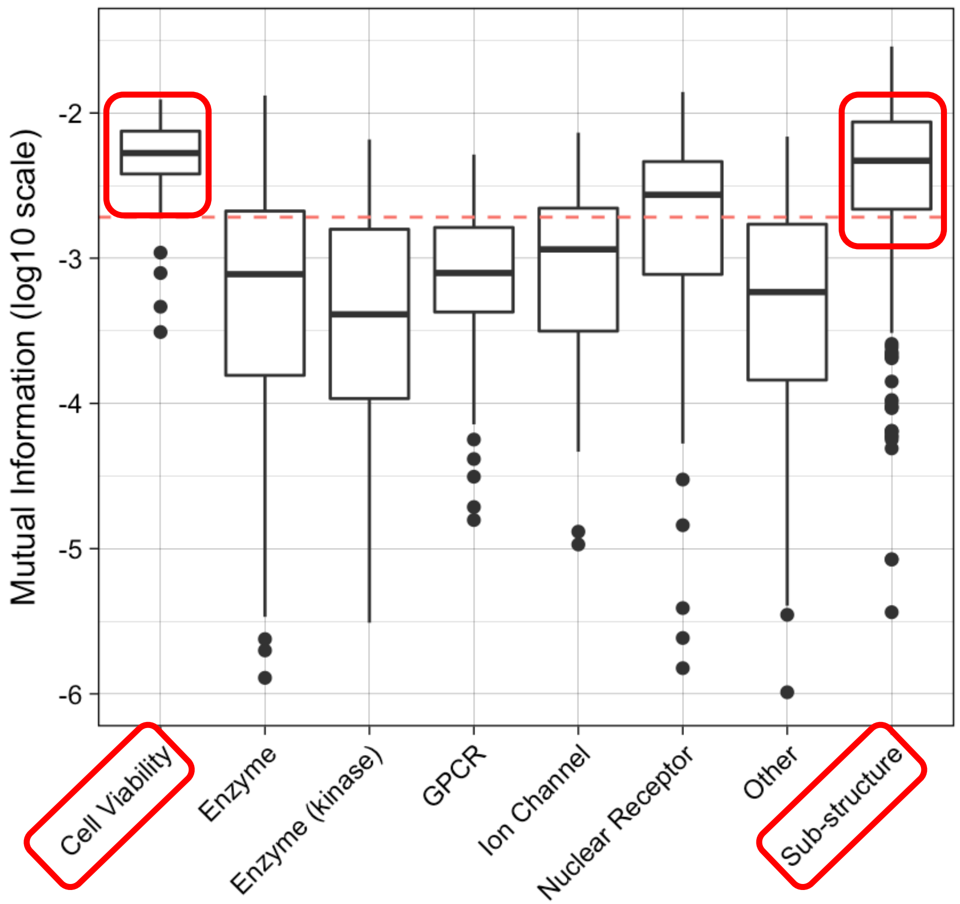
* Number of unique rules between parentheses
** Values represent mean and standard deviation
◊ Number of conditions in rules excluding single condition rules

Unique single condition rules represent less than 10% of all unique rules

Frequent single condition rules

Features	*Chemical Structure*	*NMI*
Cytotoxicity (>60%)	-	-2.04
Sulfenic acid derivatives		-1.66
Organophosphorothionate esters, Thiophosphoric acid esters	 	-1.76
Vinyl chlorides		-2.02
Haloethyl amines (N-mustard)		-2.03
Allylic halides and alkoxides	 	-2.07
β-Haloamines		-2.22
Dinitroarenes		-2.35
Monohalogen substituted ketones		-2.50
Nitroso compounds		-2.50

Cytotoxicity and toxicophores showed the strongest univariate associations

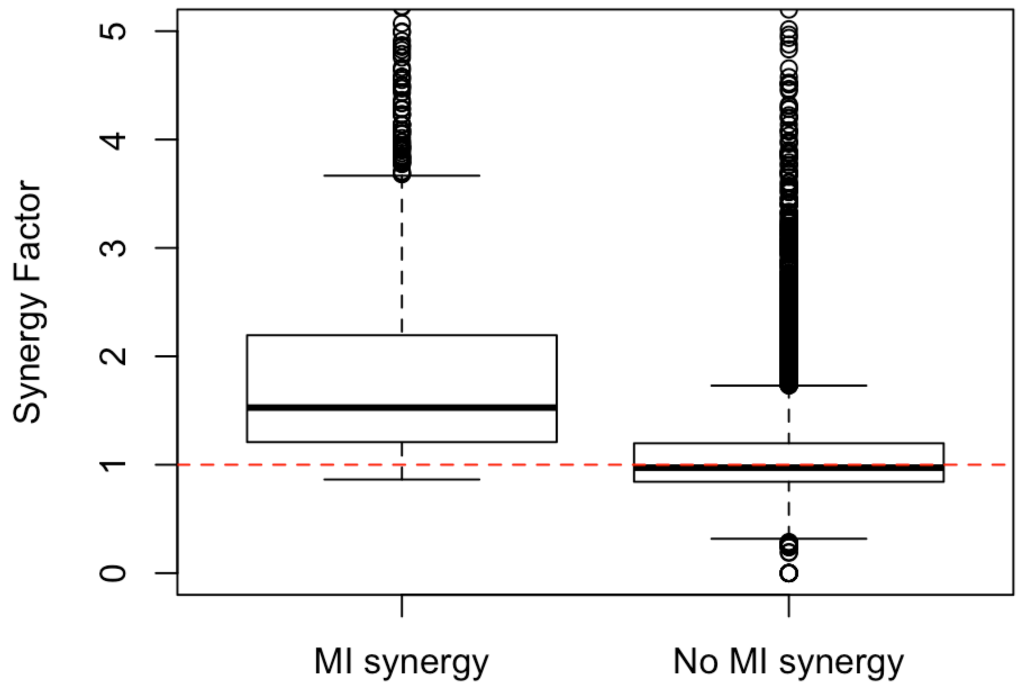


Synergy analysis to understand polypharmacology

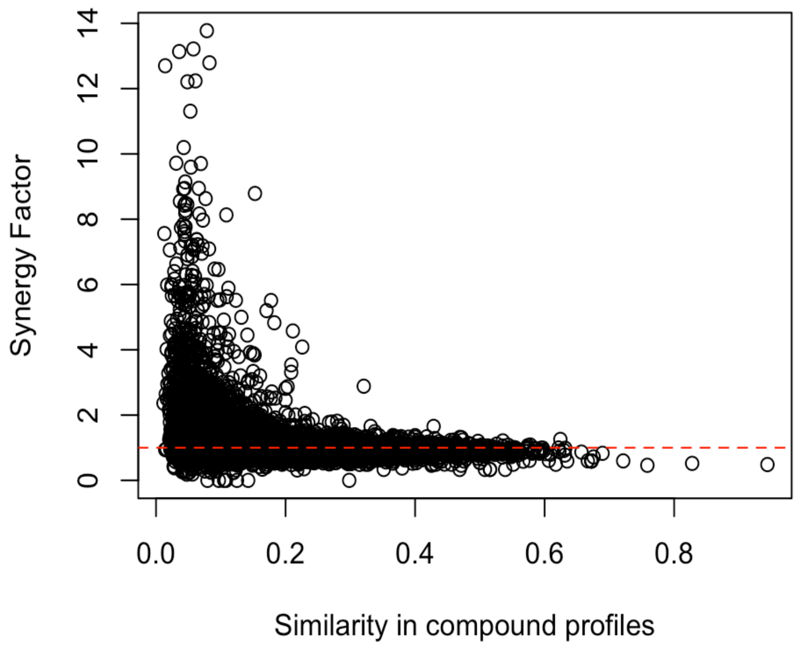
- **1) Synergy using mutual informtion**
- For c= {a,b} pair in Rule i against toxicity label Z
- **Synergy = $MI(c_i,Z) - [MI(a_i,Z) + MI(b_i,Z)]$**
- **Improvement = $MI(c_i,Z) - \max[MI(a_i,Z) , MI(b_i,Z)]$**

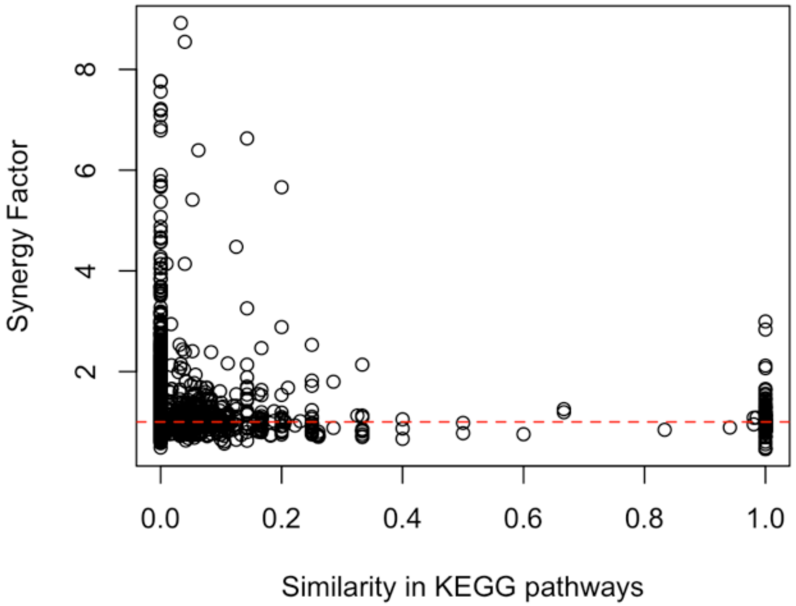
- **2) Synergy Factor from odd ratios (equivalent to interaction weight in regression equation)**

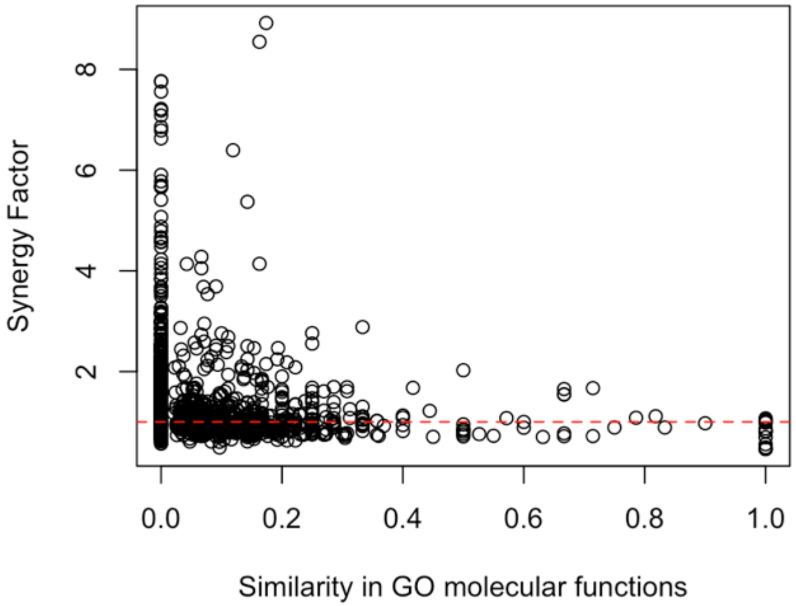
- **Synergy Factor = $OR_{12} / (OR_1 \times OR_2)$**



Synergistic pairs are dissimilar in their chemical profiles and biological function







Known key events in acute toxicity show synergy with disruption of TR, VDR and AhR

Key Event	Frequently associated Key Events	Odd Ratio comb	Synergy factor	N# toxic compounds	Interpretation from literature
Androgen antagonism **Estrogen antagonism**	Six-member ring heterocycles	6.5 2.9	**3.8** 1.5	37 47	Heterocycles steroids have enhanced activity and can produce neurotoxicity and convulsions
Glutamate receptor	**<u>TR antagonism</u>** Cell viability Disruption of mitomembrane potential	3.7 3.9 2.4	**2.2** 2.0 1.5	46 37 68	Thyroid hormone activates glutamic neuronal reuptake. Mitochondrial toxicity potential toxicity of glutamate disruptors.
GABA receptor	ARE agonist **<u>TR antagonism</u>** Disruption of mitomembrane potential HIF2 **<u>VDR</u>**	21.7 22.8 23.8 5.8 2.4	**6.8** **5.4** **6.1** **2.5** 1.4	21 21 22 23 62	TR and ROS control GABA reuptake. Vitamin D3 via VDR regulate GABA expression.

Known key events in acute toxicity show synergy with disruption of TR, VDR and AhR

Cyp2C19	CAR antagonism **VDR antagonism** **AhR activation**	1.4 1.5 1.5	1.0 1.1 0.9	327 308 326	CAR, VDR and AhR regulate the expression of Cytochrome P enzymes.
AChE	Derivatives of carbamates Phophstidyl inositol 5 phosphate kinase **VDR antagonism** **AhR activation** Troponin T cardiac NLRP3	17.4 6.4 1.8 1.8 1.6 1.7	**4.9** **2.6** 1.2 1.1 1.0 1.0	17 37 164 141 243 233	Depletion of PIP2 mediated the inhibition of ACh K^+ ion channels via PI5P4K inhibition. Cholinergic toxidrome involve Ca ion dysregulation and inflammation. Interference with calcium sensitization of troponin and inflammatory responses of NLRP3 are associated with cardiovascular effects.
Nitric Oxide Synthase (NOS)	Retinal dehydrogenase **VDR** Alkyl halides	2.1 2.0 3.2	1.3 1.2 1.2	99 160 61	VDR and retinal dehydrogenase activities can induce NOS expression.

Analysis of rule networks can reveal interesting patterns

Rule input

Network output

Entry	Feature set	Label
1	B,C,E,F	0
2	A,C,D	0
3	B,C,E	0
4	A,F	1
5	A,C,D,F	1
6	A,B,D,F	1

List representation

Entry	A	B	C	D	E	F	Label
1	0	1	1	0	1	1	0
2	1	0	1	1	0	0	0
3	0	1	1	0	1	0	0
4	1	0	0	0	0	1	1
5	1	0	1	1	0	1	1
6	1	1	0	1	0	1	1

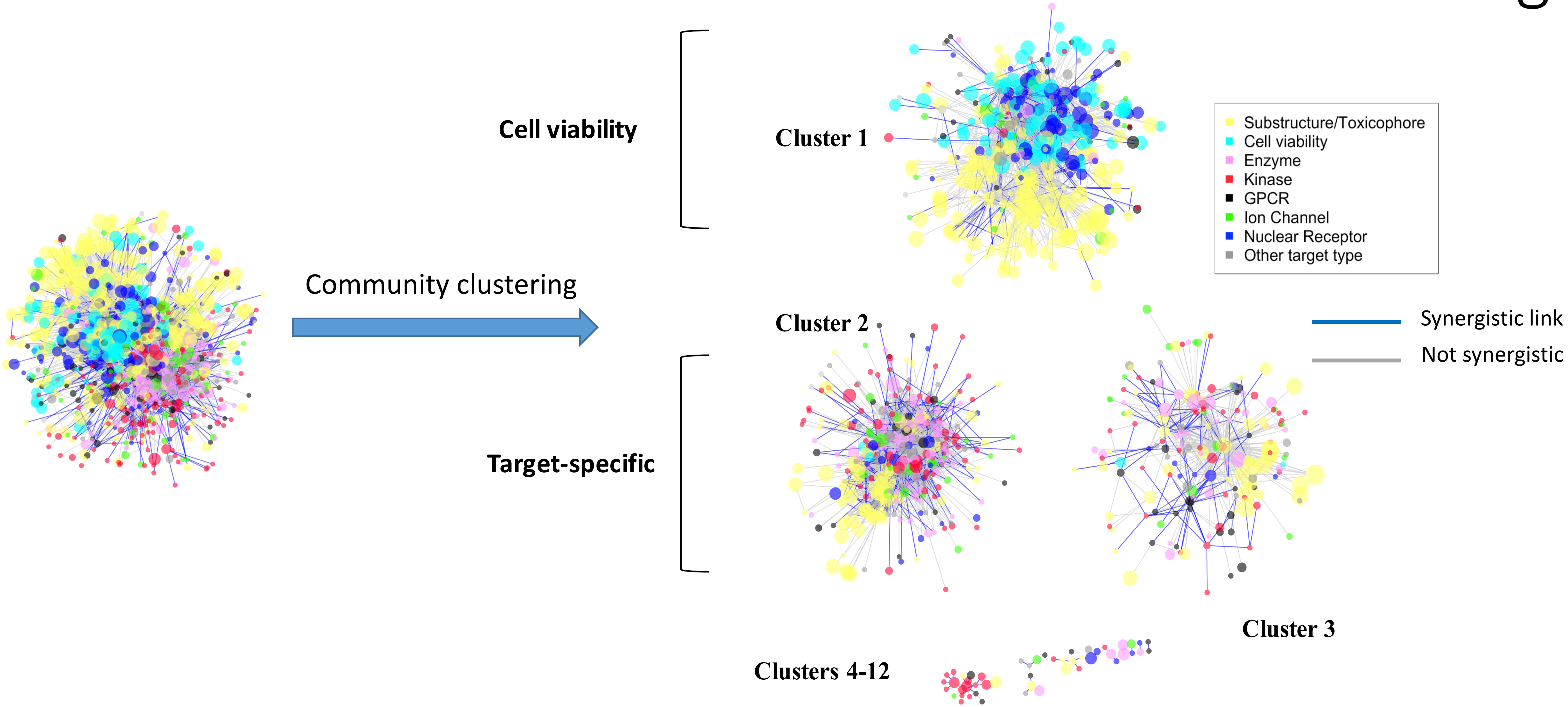
Binary representation

	A	B	C	D	E	F
A	-	1	1	2	0	3
B	1	-	0	1	0	1
C	1	0	-	1	0	1
D	2	1	1	-	0	2
E	0	0	0	0	-	0
F	3	1	1	2	0	-

Adjacency representation of toxic class

Rule network
Nodes are features and edges are adjacency in rules

Rule networks show mechanisms-based clustering

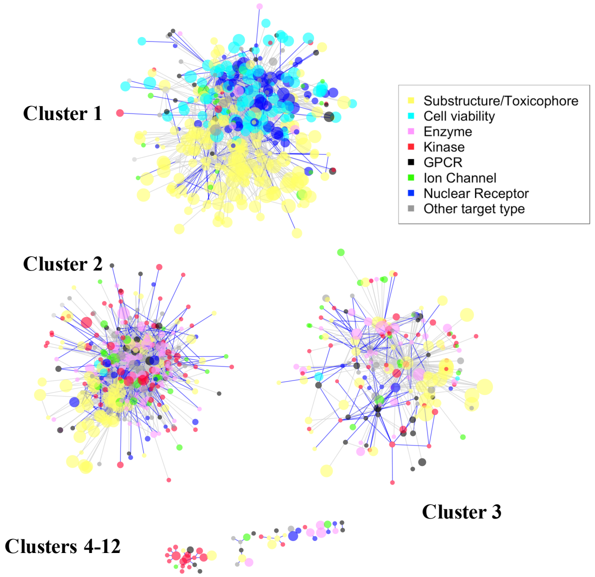


Specific and non-specific pathway perturbations form **independent clusters**.
Therefore, analysis of key events acute toxicity should consider both

Known key events in acute toxicity were central in rule networks

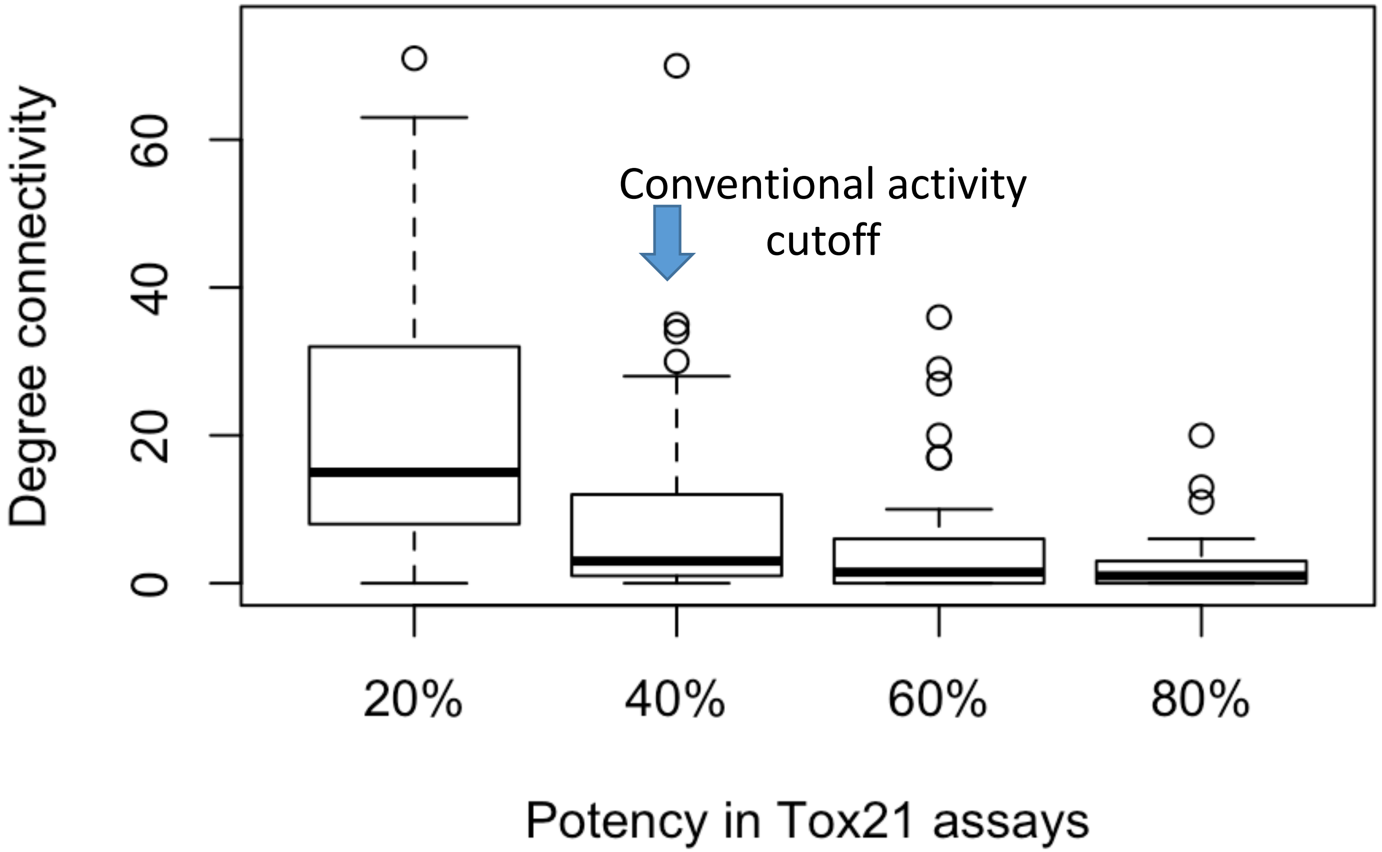
	Structural alerts		Bioactivity features	
	Top degree	*Top NMI*	*Top degree*	*Top NMI*
Custer 1	- six membered heterocyclic compounds - C=O - C=N - halogen derivatives - nitrile - saturated heterocycles - nitrogen linked to saturated carbon chain - α,β-unsaturated bond linked to oxygen atom (Michael rection acceptor)	- oxygen- linked to aliphatic carbon chain (variable length) - nitrogen linked to saturated carbon chain (variable length)	- **TR (antagonist)** - **NFE2** - CAR (antagonist) - Glutamate receptor (ion channel) - Disruptors of mitochondrial memberane potential - SIR2 - Cell viability	- ARE (agonist) - CAR (antagonist) - ER (antagonist) - AR (antagonist) - Cell viability
Custer 2	- **Halogens** - Aromatic amines - N - Amines - Oxygen group (O,S,SE)	- Halogenated alkyls and allyls - N and S mustard - O	- **AhR** - **Troponin T cardiac** - **VDR(antagonists)** - **NOS** - **AMPK** - AChE - MAP kinase kinase	- DAO - Neuronal Ach - Tyrosine kinase TYRO3 - AChE - Serine threonine protein kinase - NOS
Custer 3	- **Pnictogen (N group)** - Carboxylic acid derivatives - Tertiary amine - P or S - Derivatives of urethane (carbamates)	- Thiophsophoric acid derivatives - P - N-substituted anilines - Pnictogens (N-group) - Benzyl amine	- **Ephrine type A receptor** - Ca calmodulin protein kinase - Cyclic phosphodiesterase - AhR (activator) - AChE - PIPK	- Cyclic phosphodiesterase - Carboxic acid ester hydrolase

NMI: normalized mutual information



Nuclear receptor Disruption on their own may not explain how acute toxicity was triggered. Because they are related to chronic effects.

Exploring polypharmacology is important especially at low potencies

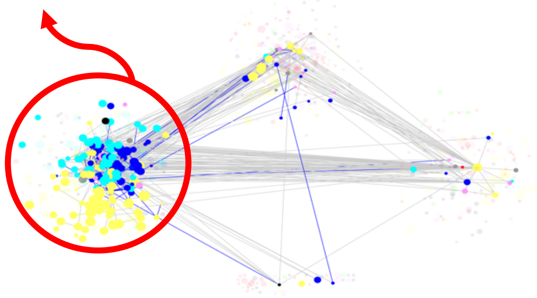


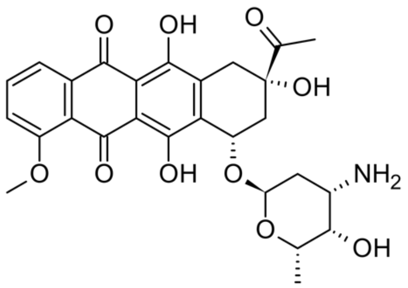
Degree connectivity is associated with how frequent the feature is used in the rules. There is an inverse relationship between potency in Tox21 assay and connectivity in the network

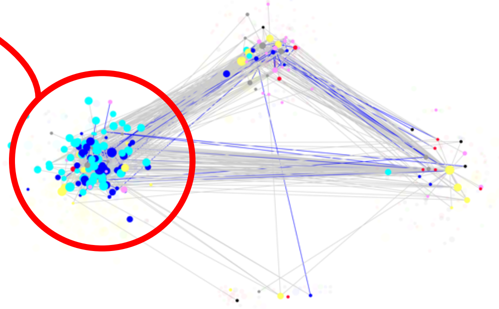
Visualizing compound-specific rule networks help to navigate polypharmacology and assess risk

Toxic compounds

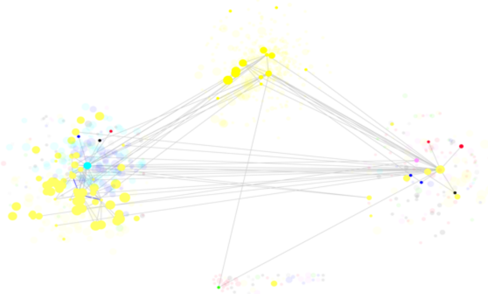
Overactivation of cytotoxicity cluster

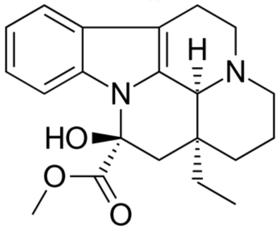


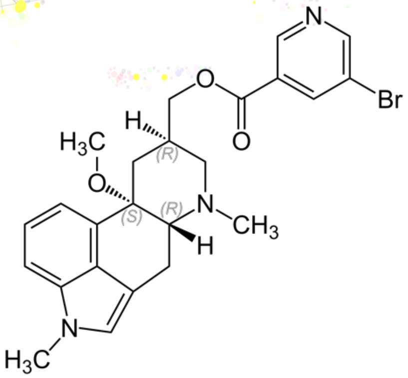


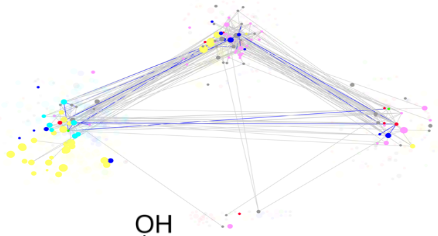


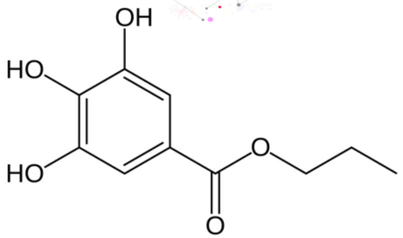
Non- toxic compounds





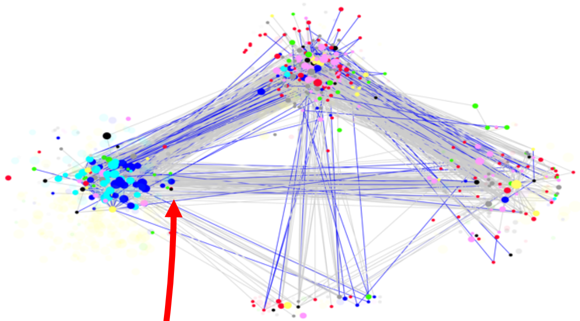


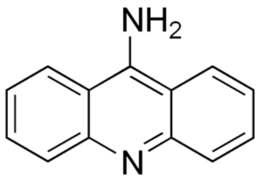


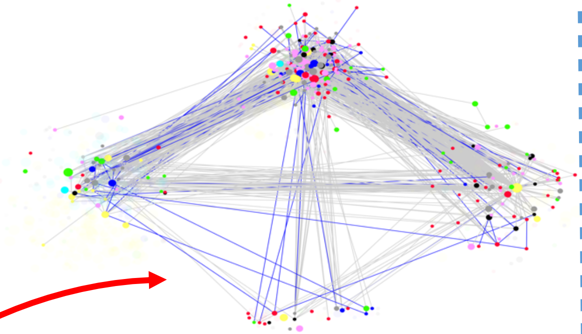


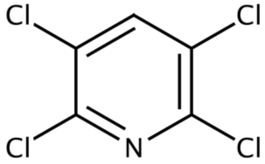
Weak activation of network clusters

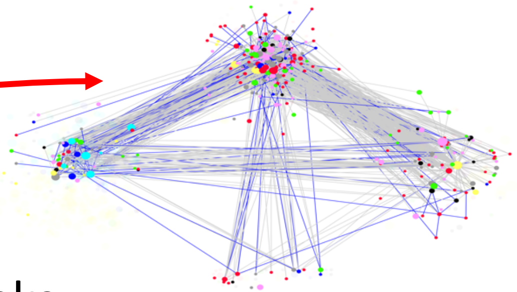
Toxic compounds



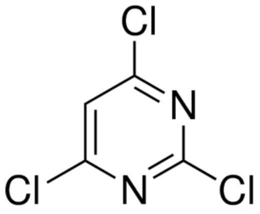




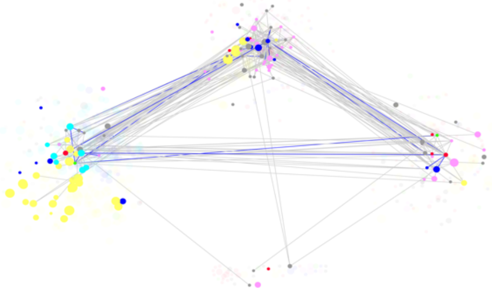


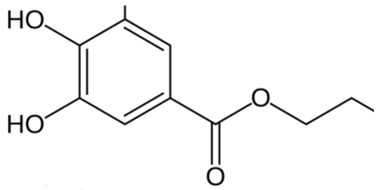


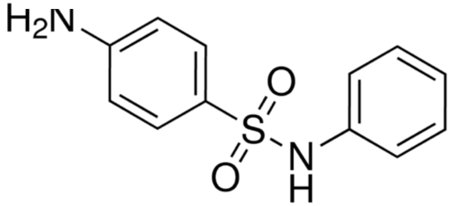
Higher density of synergistic links (synergy density greater than 15% has 3.5 higher odds to observe acute toxicity)

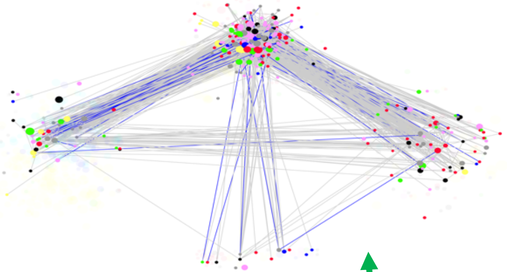


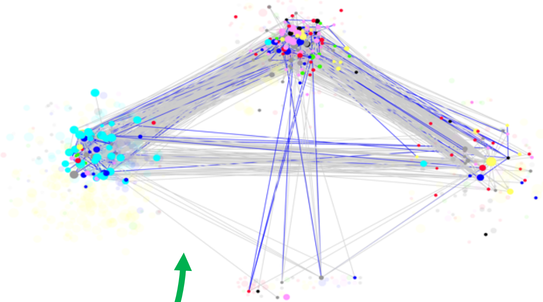
Non- toxic compounds

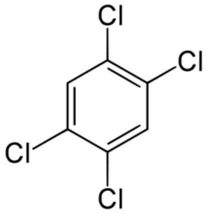


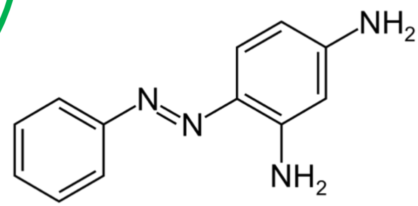












Promiscuous compounds of relatively lower density of synergistic links (compared to toxic compounds of similar substructures)

Conclusions

- **Hepatotoxicity** cannot very well be captured by single assay endpoints, but better by a **combination of bioactivities in relevant assays**, with the likelihood of hepatotoxicity increasing with **assay promiscuity**
- *In vitro-in vivo* associations improved by incorporating physicochemical properties, such as number of rotatable bonds, especially for the potent toxicity levels
- In order to capture **acute toxic**ity using *in vitro* methods, **polypharmacology** should be considered, especially **at weak potencies** which can be overlooked using conventional safety margin methods
- **Synergistic polypharmacology** is common between known key events and the disruption of relevant nuclear receptors (**TR, VDR and AhR**)
- Understanding significant polypharmacology can be used to guide cost and time effective iterative screening protocols for toxicity assessment

Acknowledgments

University of Cambridge

Dr Andreas Bender

Dr Fredrik Svensson

Kathryn Giblin

Dr Dezso Modos

Ines Smit

Dr Avid Afzal

Azedine Zoufir

Dr Natalia Aniceto

Dr Lewis Mervin

University of Liverpool
Dr Frans Coenen

OCHEM platform
Dr. Igor V. Tetko

Lhasa ltd
Dr Lilia Fisk
Dr Jonathan Vessey

EPA National Centre for Computational Toxicology
Dr Russell Thomas

Islamic Development Bank
Cambridge Trust Fund

Thank you