UNDERSTANDING CONDITIONAL MODES OF ACTIONS IN CHEMICAL-INDUCED TOXICITY USING RULE MODELS

University of Cambridge

Samar Yousif Mohamed Mahmoud

Murray Edwards College

This dissertation is submitted for the degree of Doctor of Philosophy

January 2019

PREFACE

This thesis is submitted in fulfilment of the requirements for a Doctor of Philosophy in Chemistry. I declare that this dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

This dissertation does not exceed the word limit of 60,000 words.

Samar Mahmoud

January 2019

UNDERSTANDING CONDITIONAL MODES OF ACTIONS IN CHEMICAL-INDUCED TOXICITY USING RULE MODELS

SUMMARY

It is estimated that 115 million animals are used in experimental testing each year. Hence, shifting efforts toward alternative methods for toxicity assessment is essential. However, slow regulatory acceptance of new approaches is governed by knowledge gaps in toxicity modes of action. In this thesis, I describe these challenges and the use of *in vitro* screening as an alternative of animal testing. I also discuss common data-based methods to derive hypotheses about toxicity modes of actions, and the associated limitations in capturing multiple biological perturbations.

I applied novel data-based workflows, using rule models, to prioritize *in vitro* assays predictive of toxicity as well as to detect significant polypharmacology profiles. I explain how constraints were applied to rule-based models to inform meaningful mechanistic interpretation for two toxicity endpoints: rat hepatotoxicity and acute toxicity.

I compared assays selected, by rules, for predicting hepatotoxicity with endpoints used in *in vitro* models from commercial sources. An overlap was observed including cytochrome activity, mitochondrial toxicity and immunological responses. However, nuclear receptor activity, identified in rules, is not currently covered in commercial setups. I also demonstrate that endocrine disruption endpoints extrapolate better into *in vivo* toxicity when a set of specific conditions are met, such as physicochemical properties associated with good bioavailability. Next, I examined synergistic interactions between conditions in rules describing acute toxicity. I gained novel insights into how specific stressors potentiate the perturbation by known key events, such as acetylcholinesterase inhibition and neuro-signalling disruption. I show that examining polypharmacology profiles is particularly important at low bioactive potencies. Further, the overall predictive performance of rules describing acute toxicity was tested against a benchmark Random Forest model in a conformal prediction framework. Irrespective to the data type used in the training, the models were prone to bias over compounds promiscuity, by which high promiscuous compounds were more likely to be predicted as toxic.

Overall, the studies conducted in this thesis provide novel insights into molecular mechanisms of toxicity, namely hepatotoxicity and acute toxicity, and with regards to chemical properties and polypharmacology. This knowledge can be used to improve the utility and design of alternative methods for toxicity, and hence, accelerate the regulatory acceptance.

Samar Mahmoud

ACKNOWLEDGEMENT

ABBREVIATIONS

AC50	Concentration At Half Maximum Activity
AChE	Acetylcholinesterase
AhR	Aryl Hydrocarbon Receptor
ALDH1	Retinal Dehydrogenase-1
AMPK	AMP-Activated Protein Kinase
AO	Adverse Outcome
AOP	Adverse Outcome Pathway
AOP-KB	AOP-Knowledge Base
AR	Androgen Receptor
ARE	Antioxidant Response Elemnt
BSK	Bioseek
C/EBPB	CCAAT/Enhancer Binding Protein B
CaMK	Calmodulin Kinase
CAR	Constitutive Androstane Receptor
CAT	Catalase Gene
CCR	Correct Classification Rate
Cmax	Maximum Plasma Concentration
CMC NITE	Chemical Management Center Of Japan National Institute Of Technology And Evaluation
Conf	Rule Confidence
CPAR	Classification Based On Predictive Association Rules
CREB	Cyclic AMP Response Element Binding Protein
CTB	Cytotoxic Burst
CTD	Comparative Toxicogenomics Database
CXCL10	Chemokine Ligand 10
CXCl9	Chemokine Ligand 10
CYP	Cytochrome P Enzyme
D1	Horizontal Axis Of MDS Plot Using Circular Fingerprint Descriptors
D2	Vertical Axis Of MDS Plot Using Circular Fingerprint Descriptors
DAO	Diamine Oxidase
DAT	Dopamine Transporter
DILI	Drug Induced Liver Injury
EC	EU Regulation
ECG	Electrocardiogram
ECHA	European Chemicals Agency
EP	Emerging Patterns
EPA OPPT	EPA Office Of Pollution Prevention And Toxins
EPHA2	Ephrin Receptor
ER	Estrogen Receptor

ERE	Estrogen Receptor Response Element
FAERS	Fda Adverse Event Reporting System
FDA	Food And Drug Administration
FOIL	First Order Inductive Learner
FP	Chemical Fingerprint
FXR	Farnesoid X Receptor
GABA	γ−αμινοβυτψριχ αχιδ
GHS	Reduced Glutathione
GHS	Globally Harmonized System Of Classification And Labelling Of Chemicals
GO	Gene Ontology
GPCR	G Protein-Coupled Receptors
HCIS	Australian Hazardous Chemical Information System
HCS	High Contenet Screening
hERG	Potassium Voltage-Gated Channel Subfamily H Member 2
HIF	Hypoxia Response Element
HSP	Heat Shock Protein
HTS	High Throughput Screening
IATA	Integrated Approaches In Testing And Assessment
IC50	Half Maximum Inhibitory Concentration
IG	Information Gain
IL-12	Interleukin-12
IVIVE	Invitro In Vivo Extrapolation
JNK	C-Jun N-Terminal Kinase
KE	Key Event
KER	Key Event Relationship
LD50/LC50	Lethal Dose/Concentration Of 50% Of Population
LEL	Lowest Effective Level
LSEC	Liver Sinusoidal Endothelial Cells
LUCS-KDD	Liverpool University Computer Science – Knowledge Discovery In DATAS
LXR	Liver X Receptors
MAPK	Mitogen-Activated Protein Kinase
MDL	Minimum Description Length
MDS	Multidimensional Scaling Analysis
MI	Mutual Information
MIE	Molecular Initiating Event
MOA	Mode Of Action
MoA QSAR	Mode Of Action QSAR
MPT	Mitochondrial Permeability Transition
NCM	Non-Conformity Measure
Ner2	Erythroid-Derived 2
NMI	Normalized Mutual Information

NO	Nitric Oxide
NOAEL	No Observed Adverse Effect Level
NOS	Nitric Oxide Synthase
NR3C1/GR	Glucocorticoid Receptor
NRU	Neutral Red Uptake
NVS	Novascreen
OCHEM	Online Chemical Modeling Environment
OECD	Organization For Economic Co-Operation And Development
OR	Odd Ratio
PBPM	A Physiologically Based Pharmacokinetic Model
PCA	Principal Component Analysis
PGE2	Prostaglandin E2
PI5P4K	Phosphatidylinositol-5-Phosphate-4-Kinase
PIP2	Phosphatidylinositol-4,5-Diphosphate
PIPK	Phosphatidylinositol Phosphate Kinase
PKC	Protein Kinase C
PPAR	Peroxisome Proliferator-Activated Receptors
PPB	Plasma Protein Binding
PPI	Protein-Protein Interactions
PRM	Predictive Rule Mining
PRP	Proportional Reporting Ratio
PTGER2	Prostaglandin E Receptor
QSAR	Quantitative Structure Activity Relationship
RAR	Retinoic Acid Receptor
REACH	Registration, Evaluation, Authorisation, And Restriction Of Chemicals
RF	Random Forest
ROS	Reactive Oxygen Species
RXR	Retinoid X Receptor
SIR2	Sirtulin-2
SLC6A3	Sodium-Dependent Dopamine Transporter
Supp	Rule Support
TCDD	2,3,7,8-Tetrachlorodibenzo-*P*-Dioxin
TF	Tissue Factor
TNR	True Negative Rate
Tox21	Toxicity Testing In The 21st Century
ToxRefDB	Toxicity Reference Database
TPR	True Positive Rate
TR	Thyroid Hormone Receptor
TSPO	Translocator Protein
US EPA	United States Environmental Protection Agency
VDR	Vitamin D Receptor
VDRE	Vitamin D Receptor Response Element

TABLE OF CONTENT

INTRODUCTION

According the Cruelty Free International, 115 million animals are used in experimental testing each year, where the USA, Japan and China top the list.[1] A worldwide movement against animal testing resulted in number of regulatory changes including the animal ban in the EU on cosmetic products since 2013.[2] Yet, the adoption of animal free drug testing is still far from common practice, as the number of animals tested each year is not declining.[1] The demand for banning animal tests does not only arise from ethical pressures but also the lack of efficiency and reliability of animal models as representatives of effects in human. Research have shown low concordance between outcomes from preclinical animal experiments and human effects.[3,4] This partly explains why 90% of drug candidates which were found promising in animal tests fail in clinical trials.[5] Rectifying the current situation requires broad-scale actions from the regulatory and scientific communities to define alternative measures, as well as from the general public to raise awareness.

In the first part of this chapter, we address challenges of animal testing and discuss alternative toxicity testing methods, giving great attention to *in vitro* methods. In part 2, we explain how data can be used to generate hypotheses about toxicity modes of action, as well as how to predict toxicity with confidence using conformal models.

PART 1: THE USE OF *IN VITRO* SCREENING IN TOXICOLOGY

1.1 Current status in toxicity testing and the need to reduce, refine and replace animal testing

In order to fulfil the requirements by regulatory agencies for compounds produced or used at a large scale, stakeholders are obligated to provide information about substance properties, exposure and risk management measures.[6] As defined by the United States Environmental Protection Agency (EPA), risk assessment is the characterization of "the nature and magnitude of health risks to humans and ecological receptors (e.g., birds, fish, wildlife) from chemical contaminants and other stressors, that may be present in the environment".[7] Risk assessment and toxicity profiling are key steps in the regulation and legalization of chemicals. Safety profiling is also a vital part in the drug discovery process, which runs as early as stages of hit identification and optimization. Suboptimal safety profiles contributes to the high attrition rates of drug candidates reaching the market, due to unwanted adverse effects.[5]

Chemical safety requirements in both pharmaceutical and environmental industries demand extensive repeated-dose animal testing for a wide range of endpoints. Consequently, a considerable of animals battery is consumed to assess acute and chronic toxicity effects.[8] For example, reproductive, developmental and carcinogenic endpoints require long-term animal exposure to chemicals on two or more species.[9,10] Moreover, toxicity profiling of compounds involves conducting pharmacological safety tests to survey all organs for any adverse effect using various toxicity measures; such as lethal doses (LD50) and no observed adverse effect level (NOAEL).[11] The timeline of animal-based safety assessment during the drug discovery process is presented in Figure (1-1). The reader can refer to relevant reports on the protocols of *in vivo* testing for drugs[12] and environmental chemicals.[13]

Animal-based testing is inefficient and raises considerable economical and ethical concerns, resulting in global pressures.[14] Large numbers of potentially hazardous chemicals are not yet tested because current traditional methods are economically unfeasible. The estimated cost to legalize 68,000 compounds (in 2009) under the REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) regulations is $9.5 billion using 54 million

vertebrate animals, which is difficult to handle. Additional limitations are the lack of mechanistic information on the toxicological modes of action[15] as well as the difficulty in extrapolating results from animal models to estimate effects in human. From the pharmaceutical industry's perspective, the vast majority of potential drug candidates are still failing during late stages of clinical trials due to insufficient safety evidence, despite passing preclinical safety testing.[5] Therefore, actionable steps are to be conducted in order to fill the gaps in traditional methods.

IND   NDA

Drug Discovery	***Preclinical Studies***	***Human clinical trials***
• Exploratory non-GLP single and repeated dose studies in 2 species • Safety Pharmacology • Hepatotoxicity • Embryotoxicity	• GLP and non-GLP 28 days repeated dose in 2 species • GLP safety pharmacology • GLP Genotoxicity	• GLP chronic toxicity for 3-12 months in 2 species • GLP reproductive toxicity • GLP carcinogenicity for 24 months in 2 species

Figure (1-1) The timeline of animal-based safety assessment during drug discovery and development.[16] GLP: good laboratory practice, IND: investigational new drug, NDA: new drug application.

The first step towards reducing the reliance on *in vivo* testing methods is to reduce the number of animals used. The 3Rs principle, which stands for Replacement, Reduction and Refinement of the use of animals in testing and research, was first introduced by Russell and Burch in 1959.[17] Since then, efforts to apply the 3Rs principles in the pharmaceutical sector resulted in reducing the number of animals used for testing through improving study design and coordinating projects.[18] The most common and accepted animal testing guidelines, which is described by OECD documentation, has already entitled a significant reduction in the number of animal used to assess systemic toxicity.[19]

The second step, to facilitate the aspirations of the 3Rs, is to develop and validate non-animal testing methods, which can be used to integrate with or replace *in vivo* approaches.[20] The most common alternative methods are *in vitro* screening and *in silico* models, as will be discussed subsequently in this chapter. Several *in vitro* tests have already been validated and accepted in toxicity profiling as exemplified in Table (1-1). The set of validated assays represent endpoints

that can directly translate into *in vivo* effects. However, this may not be the case for complex endpoints such as toxicities in the liver and the nervous system, where multiple pathways and mechanisms are involved. In order to address this, a broad scale *in vitro* approach is adopted, as will be reviewed next.

Table (1-1) Examples of accepted *in vitro* assays to predict *in vivo* toxicity.

Endpoint	*In vitro assay examples*	*OECD document number*	*Adopted*
Genotoxicity[21]	Bacterial reverse mutation assay	471	1983
	In vitro mammalian chromosomal aberration test	473	1983
	In vitro mammalian cell gene mutation test	476	1984
	In vitro mammalian cell micronucleus test	487	2010
	In vitro gene mutation assays using the TK locus	490	2015
Acute toxicity[22]			
- Skin corrosion	*In Vitro* Skin Corrosion: Human Skin Model Test	431	2004
- Phototoxicity	*In Vitro* 3T3 NRU phototoxicity test	432	2004
Endocrine disruption[23,24]	ER binding assay	TG 493	2015
	Steroidogenesis	TG 456	2011

1.2 Large scale *In vitro* methods as a potential alternative for animal testing

Driven by the need for an in-depth understanding of the molecular basis of toxicity events, and parallel to the calls for reducing animal testing, several technologies have been proposed and studied to characterize and predict adverse effects.[25] The vision of toxicology in the twenty-first century is motivated by shifting the standard and predefined battery of toxicology tests into hypothesis-driven methodologies which are adapted to the specific characteristics and use of chemicals.[26] With the wealth of *in vitro* screening methods that have been proposed for

toxicity profiling,[27,28] our focus in this chapter will be on *in vitro* approaches aimed for broad biological space screening.

The special attention towards large-scale screening methods has arisen due to the unprecedented advances in biotechnology, genetics and engineering which have created an enriched environment for the rapid screening of chemical compounds for their gene expression profiles and biological activities via high-throughput screening (HTS) technologies.[14,29,30] Within the scope of drug discovery, HTS has not only allowed for the identification of hits and lead optimization, but also facilitated the success of some drug candidates in reaching the market.[31,32,33] Following this success, several projects such as the Tox21[15] (Toxicity testing in the 21st century) and ToxCast[16] programs have been launched to evaluate toxicity via HTS techniques and consequently generate large *in vitro* data.

Typically, HTS involves the screening of chemical entities in high-quality assays that capture biological functions at various degrees of complexity ranging from interaction with pure biological macromolecules, such as proteins and DNA, up to whole cellular systems and tissues. HTS is performed in an efficient automated manner using 96, 384, 1536 or 3456 microtiter wells with the aid of high-performance detection techniques and followed by data processing and chemical profiling.[30] Assays can be categorized as target-based, or phenotypic-based according to readout complexity (Figure (1-2)). The measured phenotype can be an alteration in a biomarker or cellular components, such as changes in cell morphology and physiology. In contrast, target-based assays measure the direct or indirect interactions of compounds with specific predefined targets. The hypothesis behind the utilization of HTS in toxicity assessment is that the chemical-induced toxicity is initiated by an interaction between the chemical and biological targets, by which one or multiple pathways are perturbed.[34]

Several repositories were created to store measurements from *in vitro* screening, such as PubChem,[35] ChEMBL,[36] DrugBank[37] and ChemBank.[38] We will focus on two examples where broad scale *in vitro* screening is conducted for toxicity assessment purposes. The first example explains the use of secondary pharmacology profiling for drug development projects. Then, considerable attention will be given to the US EPA ToxCast initiative, which forms the backbone of data used for the analysis in this thesis.
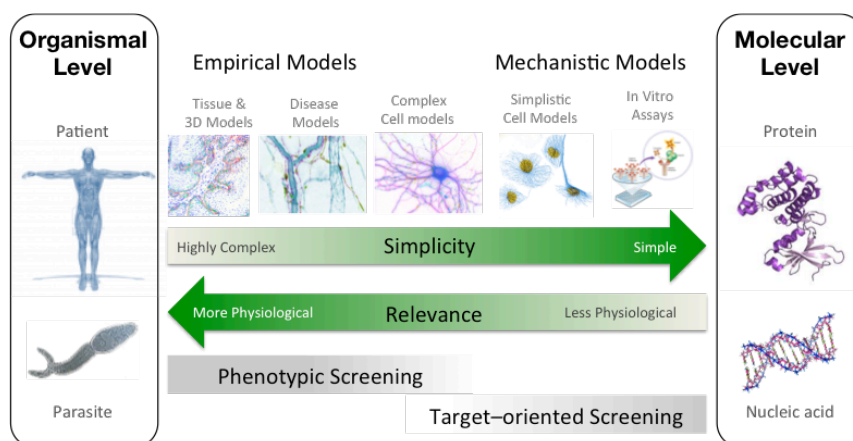
Figure (1-2) Target-focused screening versus phenotypic screening. Target-focused assays are simpler and explain specific molecular functions. On the other hand, phenotypic assays involve at least a cellular biological system and can also extend to tissues of various cell types. Adapted from http://www.sulsa.ac.uk/research-facilities/uk-npsc/phenotypic-screening.

1.2.1 Secondary pharmacological screening

During the drug development process, secondary pharmacology screening, often termed off-targets profiling, is conducted to identify possible bioactivities against targets other than the intended therapeutic target.[39,40] It is reported that around 75% of adverse drug reactions can be attributed to the pharmacological profile of drugs, and hence, these unwanted effects can be predicted early on.[41] Secondary pharmacological profiling involves screening against a range of targets types including enzymes, nuclear receptors, ion channels and transporters. A list of targets commonly used by four major pharmaceutical companies have been reviewed by Bowes *et al*.[41]

Data from secondary pharmacology screening can be used to assess the promiscuity of compounds.[42] Promiscuity is defined as the percentage at which a compound show activity at a given concentration (for example 1μM) against multiple pharmacological targets.[43] Despite controversial reports, high promiscuity has commonly been associated with a higher likelihood of adverse reactions.[44] Some marketed drugs show high promiscuity, yet tolerable effects.[45] Hence, disregarding promiscuous compounds in drug discovery projects may not be the best approach to address the risks of potential clinical effects.[46] In Chapter 6, we will address the confounding effect of promiscuity on the classification of toxic compounds.

Bowes *et at* argued that the associations between hitting particular targets can have higher relevance to human effects in comparison to animal models. Hence, these profiles provide

mechanistic insights about side effects observed in human. The secondary pharmacology profiling data can also be used to prioritize potential drug candidates as well as to inform subsequent tests. For example, if a hit target is identified, further experiments should be conducted to determine the selectivity across the target family and the potency in dose response assays. Also, target activity can guide the design and selection of *in vivo* studies relevant to the detected activity.

One of the key off targets is potassium voltage-gated channel subfamily H member 2 (hERG), which is responsible for cardiovascular effects upon inhibition through the prolongation of QT interval of the electrocardiogram.[40] Due to the high impact of inhibiting hERG, a regulatory requirement in drug submissions in first in human trials is including results of screening against this ion channel.[47] One way to improve risk assessment using off-target activity is estimating safety margins by calculating the ratio of *in vitro* off-target potency (IC_{50}) to the maximum plasma concentration C_{max}.[48] Several studies have shown how incorporating C_{max} can improve the predictivity of clinical effects (see section 1.4.1). Yet, the limitation of this approach is overlooking the consequences of polypharmacology profiles of low potencies. The later will be addressed in Chapter 5 for chemicals inducing acute toxicity.

1.2.2 ToxCast *in vitro* screening

ToxCast is a chemical prioritization research program launched by the EPA in 2007 in collaboration with the National Toxicity program and the National Institute of Health Chemical Genomic Centre.[15,49,50] One of the EPA's goals is to effectively manage and regulate environmental chemicals that are likely to cause harm to humans. However, such regulations require the assessment of large numbers of chemical entities. Inspired by the vision of the National Research Council for measuring perturbations in toxicity pathways as signatures to predict adverse effects *in vivo,*[15,51] the EPA's ToxCast program has adopted high throughput screening as a tool to assist in toxicity profiling by incorporating mechanistic *in vitro* assays[50]. Moreover, the advances in toxicogenomics, transcriptomics and proteomics have made a great contribution in the determination of toxicological pathways and targets responsible for inducing diseases *in vivo*[25]. Hence, the toxicity biomarkers identified by transcriptomics have been utilized as assay endpoints, in conjunction with conducting a large scale screening approach for building the ToxCast database.

The wealth of ToxCast assays ranges from cell-free biochemical to physicochemical and phenotypic endpoints. These are contributed by several platform sources in collaboration with

the EPA.[49] An overview of the core screening platform within the ToxCast project is given in Table (1-2), describing the various technological screening systems. This shows the diversity of assays in the ToxCast assay pool, and hence, the potential of that data to derive novel mechanistic information associated with *in vivo* toxicity. The platforms providing the largest number of assay measurements are Apredica, Attagene, Bioseek and NovaScreen[52], in addition to Tox21 using NCGC technology[53] (see Table (1-2) for assay details).

The ToxCast chemical library was designed by selecting compounds using specific activity and physicochemical criteria. First, the environmental compounds that have the potential of *in vivo* toxicity were selected (*e.g.* as pesticides, insecticides in addition to drugs).[50] In addition, the ToxCast chemical library shows variable scaffold structures in order to assure as much chemical and bioactivity diversity as possible.[49] Figure (1-3) represents the number of compounds screened in different phases of the ToxCast project. The Tox21 phase cover measurements for over 8000 compounds against a subset of endpoints including cell viability and nuclear receptor disruption.[53] Overall, the number compounds with the broadest assays screens is around 1061.[52]

Table (1-2): Description of HTS screens in phase I of the ToxCast project, adapted from Kavlock *et al.*, *Chem. Res. Toxicol.,* 2012.[49]

Technology platform source	Description
ACEA	Real time cell electronic sensing (RT-CES) of growth of A549 cells
Apredica	Cellular high content screening (HCS) evaluating cellular markers such as stress pathways, mitochondrial involvement, cell cycle, cell loss, mitotic arrest, and the cytoskeleton in HepG2 cells
Attagene	Multiplexed transcription factor profiling in HepG2 cells
Bioreliance	Gene mutation and DNA and chromosomal damage assays
Bioseek	ELISA based readouts of interactions of co-cultures of primary human cells
CellzDirect	qNPA on select genes relevant to xenobiotic metabolism in primary human hepatocytes
Gentronix	GreenScreen genetic toxicity assay using GADD45a GFP in TK6 cells
NCGC	qHTS profiling of nuclear receptor function in agonist and antagonist mode by reporter genes using a variety of cell types
NovaScreen	Biochemical profiling, largely using human proteins, of receptor binding, enzyme assays, GPCRs, and ion channels
Odyssey-Thera	Protein complementation assays for a wide variety of intracellular signalling networks.
Vela Sciences	High content multiparameter assays providing quantitative digital imaging of cultured cells as well as information on a variety of cellular proteins/structures/function
Zebrafish	Zebrafish embryonic development assay
Mouse embryonic stem cells	Mouse embryonic stem cell cytotoxicity and differentiation
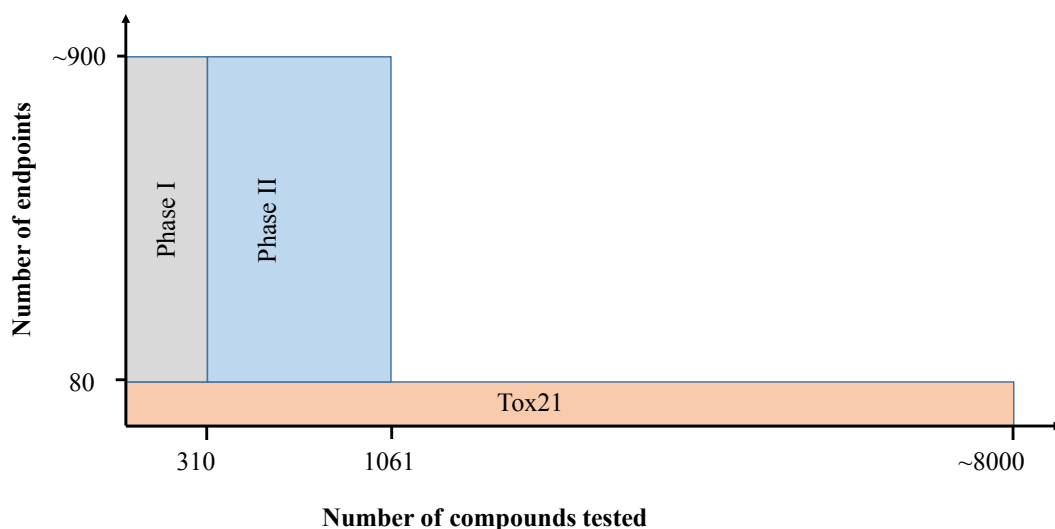
Figure (1-3) Number of compounds tested and assay endpoints screened in the key ToxCast and Tox21 phases.

1.3 ToxCast for Integrated Approaches in Testing and Assessment (IATA)

The new toxicology approach, commonly referred to as the Integrated Approaches in Testing and Assessment (IATA), is mainly intended to reduce, refine and replace animal use for toxicity evaluation (3R measures).[26] These approaches are used to prioritize testing, identify and characterize hazard, as well as to assess risk according to exposure measures. The aim through IATA is to provide a mechanistic oriented system for an efficient and human/environmental relevant safety evaluation.[54] The IATA approach is designed to tailor the specific chemical and its exposure level. The components of IATA can be categorized into i) mode of action and adverse outcome pathways, ii) *in vitro in vivo* extrapolation (IVIVE) and iii) data-based methods[55] (see Figure (1-4). These are performed by integrating two or more of the following; physicochemical properties, *in vitro* assays, mode of action analysis (MOA), human epidemiology data, animal test data, kinetic models and *in silico* models.[56] These components, however, are not discrete. For example, data methods can be used to derive hypotheses about modes of action and also to predict kinetic parameters in IVIVE.

In vitro methods form a key pillar in the IATA paradigm. In the following sections, each component will be explained followed by examples of how ToxCast assays has been implemented.

Integrated Approaches for Testing and Assessment (IATA)

Mode Of Action (MOA) / Adverse Outcome Pathway (AOP)	In Vitro - In Vivo Extrapolation (IVIVE)	Data storage/ Data mining
Molecular/Cellular Changes [Initiating Events]	*Physicochemical properties / In vitro assays*	*Knowledge/studies accumulated in 40 years*
↓ downstream	↓ predict	↓ process
Adverse effects of regulatory relevance	**Metabolism / Kinetics**	**Curated databases / Develop predictive models**

Figure (1-4) Integrated Approaches in Testing and Assessment (IATA) which consists of three main components, namely Mode of Action/ Adverse Outcome Pathway (MOA/AOP), *in vitro in vivo* extrapolation (IVIVE), and data mining.

1.3.1   Mode of Action/ Adverse Outcome Pathway

1.3.1.1 Overview of MOA/AOP framework

The mode of toxicity can be defined as a set of physiological signs that characterize adverse responses in the biological system.[57] It is associated with identifying the perturbation in the biochemical pathway as well as the resulted physiological changes produced by that perturbation. A more recent term for the concept, which is the adverse outcome pathway (AOP), was first described by Ankley *et al.*[58] The AOP can be defined as the conceptual framework which reflects the existing knowledge linking the initial triggers at the molecular level and the adverse outcome at the organizational level[58,59] (see Figure (1-5) and Table (1-1)). The AOP can be viewed as a series of sequential events reflecting a gradual increase in the complexity of the biological levels. These key events are linked via causal, mechanistic or inferential connections termed as key event relationships (KERs).[59,60] The downstream adverse outcomes commonly describe effects which are relevant to regulatory testing requirements. [61]

Molecular ➤ Cellular ➤ Tissue ➤ Organ ➤ Individual

MIE → KE → KE → KE → AO
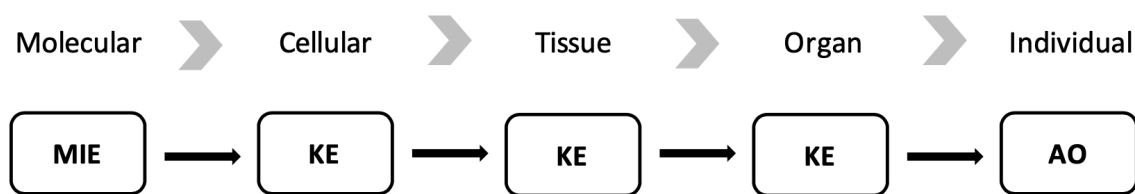
MIE: Molecular Initiating Event

KE: Key Event

AO: Adverse Outcome

Figure (1-5) Adverse outcome pathway (AOP) framework.

Table (1-3) Components of AOP.

Key Event (KE)	• Represent the nodes in the AOP • Describes measurable endpoints in the biological state
Molecular initiating event (MIE)	• Specific type of key events and the starting node in the AOP • Describes the interaction between chemical and biological targets, chemical reactivity, such as the formation of a hapten conjugate
Key Event Relationship (KER)	• Directional relationship between adjacent key events • Describes the weight of evidence which supports the functional or structural relationship between upstream and downstream key events
Adverse Outcome (AO)	• Key event at the biological organization level • Describes an apical endpoint usually of a regulatory relevance or part of guideline test

High throughput and high content screening methods combined with computational modelling have provided useful tools for predicting adverse reactions and identifying measurable biomarkers. However, in order to incorporate this knowledge in the regulatory decision making, the mechanistic framework has to be validated and verified.[62] The AOP not only provides a means for mechanistic verification and approval but also outline the current knowledge on how initiating events progress into adverse outcomes. Another advantage of the AOP system is reducing the complexity of the biological signalling pathways involved in the toxicological events.[59] This is achieved by identifying measurable markers in the pathway in form of key events (KE). Hence, each key event in the pathway captures an upstream effect

which induces the relevant downstream effect if sufficient magnitudes and durations of perturbations were reached.

In an attempt to promote the utility of the AOP framework as a predictive tool in risk assessment, the Organization for Economic Co-operation and Development (OECD) has lead initiatives towards guiding and promoting AOP development. The OECD formulated a set of guidelines for the development and evaluation of AOPs in 2013[63] followed by a revised version in 2017.[54] Additionally, the OECD constructed an assembly of the accumulated knowledge in AOPs, namely the AOP-knowledge base (AOP-KB) which was made publicly available in 2014.[60] The AOP-wiki module of the AOP-KB represents a documented source of key events and associated adverse outcomes. The OECD has also organized workshops to encourage the collaboration and harmonization in the AOP community across the academic, governmental and private sectors to accelerate the acceptance of AOP in regulatory toxicology.[61] Next, we will explore examples to derive molecular initiating events and early key events in the AOP captured as chemical substructures, *i.e.* structural alerts, and biological activities from *in vitro* assays.

1.3.1.2 Structural alerts to understand initiating events

Structural alerts are molecular substructure patterns that are associated with adverse reactions,[64,65] also known as toxicophores. These molecular features are popular in elementary evaluations of the possible adverse effects because of their simplicity and interpretability.[66] The application of structural alerts can be in a form of a rule, which is described as; if alert X present in a molecule then it will have Y activity. The use of these alerts can be used for a specific *in vitro* assay activity[67] or a complex adverse effect such hepatotoxicity.[68] Structural alerts can be derived from expert knowledge and literature as well as from data-derived methods.[69] However, the boundaries between these approaches are not fixed, and rather a spectrum where expert knowledge can be integrated to refine and curate substructures detected from large datasets.[70]

Structural alerts resources include Derek[71], Toxtree[72] and ToxAlerts[73], which are derived from curating literature. Data-based methods used to prioritize these alerts vary in complexity and enrichment methods. For example, with univariate methods, alerts are represented as independent substructural features for a given endpoint, such as the case of calculating the binomial distribution in a dichotomous fashion.[74] Other methods such as emerging patterns

rather identify a set of one or more substructural features as a unit of enrichment.[67,75] In such case, the rule involves one or more substructural conditions. These methods will be further discussed in part 2 of this section.

The success of structural alerts is dependent on the type of endpoint used. Hence, in cases where the effect originates from a specific pharmacophore or chemical reactivity, *i.e.* electrophilicity, structural alerts are successful in flagging active compounds. This is particularly evident for effects such as mutagenicity[75] and carcinogenicity[67,76]. Yet, the use of structural alert as a sole predictor can come with limitations. There is a probability to incorrectly label flagged compounds as toxic, encountering for false positives.[66] Also, the use of structural alerts on their own is not enough to capture modes of actions in toxic compounds, and hence the integrating bioactivity data can be highly beneficial.[77]

1.3.1.3 Using ToxCast to derive important key events

The Toxcast *in vitro* project provides a wealth of bioactivity data for hundreds of compounds, for many of which *in vivo* toxicological data are available. These bioactivity measurements represent key events at the molecular and cellular levels which can be used to elucidate statistical links with adverse outcomes,[61] and hence, understand the associated modes of action. The Toxicity Reference Database[78] (ToxRefDB) has commonly been used as a source of animal-based *in vivo* toxicity endpoints generated in compliance with standard guidelines. For example, Martin *et al*. evaluated the effects of 309 environmental chemicals on the gene regulatory network by targeting nuclear receptors and transcription factor response elements in 73 assays.[79] The authors were able to detect 133 significant univariate associations between ToxCast *in vitro* assays and 77 rodent *in vivo* endpoints and rabbits from the ToxRefDB. In this study, significant associations were reported for Peroxisome proliferator-activated receptors; PPARα and PPARγ, with liver tumors in rats, whereas, estrogen receptor (ERα) and its response element (ERE) had associations with reproductive effects such as decreased fertility. Additionally, Hu *et al*. utilized Comparative Toxicogenomics Database to explore the annotated toxicities and adverse reaction in humans induced by ToxCast chemical set.[80] The annotated adverse effects and toxicities were grouped into four major toxicological categories; cardio-, hepato-, neuro- and renal-toxicities. The authors identified significant associations, by calculating the proportional reporting ratio (PRR), between the annotated toxicities and some genes mapped by ToxCast *in vitro* library. The analysis has shown that assays for modulating androgen receptors (AR) and glucocorticoid receptors (NR3C1) appeared repeatedly as

significant predictors of the four toxicological categories. Svensson *et al* used ToxCast assays to prioritize key events in structural cardiotoxicity (morphological damage of cardiac cells).[81] In this study, ToxCast compounds were mapped with drugs reported in the FDA Adverse Event Reporting System (FAERS). This resulted in identifying 22 adverse events-assays associations, involving estrogen receptors, PPARα/γ and androgen receptors. Additionally, the authors developed two adverse outcome pathways, namely, downregulating tissue factor (TF) and modulating translocation protein (18 kDa) TSPO as key events for heart failure and mitral valve incompetence, respectively. Overall, the data-based enrichment of important *in vitro* activities has shown that nuclear receptor disruption, such as ER, AR and PPARs, frequently and significantly correlated with a broad range of *in vivo* toxicities.

However, the insights gained from *in vitro* assays are limited to the biological space of the input data. Some of the chemical-gene interactions, which are important in toxicological events, were missing from the ToxCast assay library. For example, the catalase gene (CAT), which encodes for a vital antioxidant enzyme is missing from the ToxCast *in vitro* set. [80] The retinoic acid signalling pathway is involved in development toxicity,[82] for which several endpoints have been screened in the ToxCast project, such as retinoic acid receptor (RAR) and retinoid X receptor (RXR). However, retinal dehydrogenase is missing in the data which represent an early key event in the pathological pathway.[83]

 It is important when analysing *in vitro* data of broad biological space, to understand the confounders attributed by chemical and biological properties. *In vitro* bioactivity measurements rarely act as independent variables. An observed and reported confounder is the cytotoxic-based assay promiscuity of ToxCast activity measurements.[84] Chemicals can show cytotoxic effects at same concentration ranges where target specific assays are activated. Judson *et al* reported that cytotoxic compounds (up to 100 μM), activated, in average, 12% of assays in comparison to only 1.3% of assays activated by non-cytotoxic compounds.[85] This phenomenon is called the "cytotoxic burst" (CTB) at which intracellular machinery is activated due to unspecific effects such as cell stress and disruption of proteins and membranes. Also, the authors reported that *in vivo* toxicity can be explained by either i) specific target activities below cytotoxic concentrations or via ii) unspecific cytotoxicity and cell stress pathways. ToxCast assays that are frequently activated below CTB concentration were investigated.[84] 50 unique targets were prioritized which occupied the top 90th percentile of assays with the greatest percentage of hits below CTB. A fraction of these was proposed as novel targets for

the development of AOP, such as sodium-dependent dopamine transporter (SLC6A3), prostaglandin E receptor (PTGER2) and chemokine ligand 10 (CXCL10).[84] Additionally, target-specific assays were utilized to map genes and biological pathways. For example, Judson *et al*. mapped 315 unique genes from 467 assays so as to identify the magnitude of perturbing biological pathways by chemicals and its relation to toxicity.[15] The results revealed an association between the number of pathways perturbed by a chemical and the minimum bioactivity concentration in a set of 15 cytotoxic assays. There was also an association between the concentration that caused pathway perturbation *in vitro* and the lowest effective dose on rat prenatal developmental toxicity.[15] Another pattern is the inter-assay correlations of relevant biological functions. As will be shown in Chapter 3, assays for endocrine disruption (such as ER, AR and GR) broadly correlate with phenotypic and cytotoxic assays. This in part may explain why endocrine disruption assays frequently top *in vitro* readouts that are prioritized in data-based approaches as shown in previous examples. We will also report that assays for endocrine disruption are key predictive endpoints for hepatotoxicity (in Chapter 4) and acute toxicity (Chapter 5).

Overall, linking *in vitro* bioactivities to mapped genes and biological pathways is useful to gain novel mechanistic information about *in vivo* toxicity. However, confounders such as inter-assays correlation, cytotoxicity and promiscuity should be considered when *in vitro* data are interpreted. We will demonstrate in Chapter 4 that potent hepatotoxicants are characterized by specific pathway perturbations in contrast to low potent hepatotoxicants which tend to trigger more often non-specific phenotypic effects. Additionally, in Chapter 6 we report compound promiscuity as a confounder that can bias toxicity prediction in two different models.

Deriving hypotheses about toxicity modes of action using data-based methods to is the backbone of research conducted in this thesis. These approaches will be further explained in part 2 of this chapter. Important *in vitro* profiles of chemicals-inducing hepatotoxicity and acute toxicity will be discussed in Chapter 4 and Chapter 5, respectively.

1.3.2   *In vitro in vivo* extrapolation (IVIVE)

Quantitative *in vitro in vivo* extrapolation (QIVIVE), which is the second component in the IATA, is defined as the process of estimating the human dose which achieves concentrations at the tissue target equivalent to that observed in *in vitro* tests.[86] As the traditional estimation of pharmacokinetic parameters involves *in vivo* exposure tests to measure plasma

concentrations, volume of distribution and clearance rates, QIVIVE provides an alternative approach to model the kinetic behaviour of compounds. The main components in QIVIVE rely on i) *in vitro* screening of biological properties and ii) kinetic parameters of compounds which are together incorporated in a physiologically based pharmacokinetic model (PBPK).[87] The biological characterization includes deriving the characteristics of dose-response effects and identifying the tissue target. Whereas, *in vitro* estimation of kinetic parameters includes measurements of permeability, plasma protein binding and hepatic clearance. The QIVIVE process has also benefitted from the advancement in computational modelling. The use of Quantitative Structure Activity Relationship (QSAR) models was proposed as a third pillar in QIVIVE in order to aid the prediction of bioactivity effects and kinetic parameters.[88] Therefore, combining QSAR with *in vitro* tests and pharmacokinetic modelling has the potential to replace quantitative *in vivo* studies.

QIVIVE can facilitate time- and cost-effective strategy in risk assessment by improving the translatability of in *vitro* assay results into *in vivo* effects.[89] Wetmore *et al* used IVIVE methods to calculate daily human oral doses that are required for steady state blood concentrations equivalent to AC50 or lowest effective level (LEL) of ToxCast assays.[90] Assessing the possibility of observing adverse effects can be improved when the real exposure dose is compared with the estimated dose associated with toxicity. [91]

Despite the significant steps towards replacing animal testing for exposure assessment, limitations still exist in modelling the pharmacokinetic parameters. This is particularly evident in the case of estimating compound clearance, at which models fail short to provide accurate predictions because of the complexity of this parameter.[88] Wambaugh *et al* argued that the assumptions and approximations in common IVIVE methods can result in false estimation of equivalent *in vivo* doses when using *in vitro* derived kinetics.[87] The authors reported that inaccurate estimation (up to 10 fold of underestimation or overestimation) results from excluding or improper incorporation of active transport, secretion and reabsorption in bile, and extra-hepatic metabolism.

In Chapter 4, and in order to avoid assumption about estimating plasma and tissue concentrations, the approach we applied uses physiochemical properties as conditions to extrapolate important *in vitro* activities into *in vivo* hepatotoxicity. We demonstrate how some physicochemical properties can act as simple proxies for exposure, such maximum plasma concentrations, because they reflect properties of good bioavailability.

1.3.3 Applications of data-based methods

Data-based statistical mining on ToxCast can be categorized into i) understanding toxicity modes of action by linking *in vitro* activity to *in vivo* toxicity (see section 1.3.1.3 for examples), ii) identifying chemical categories using bioactivity profiles as well as iii) using assay measurements as toxicity predictors in machine learning models. We report below examples of studies where ToxCast data were used to achieve the latter two goals.

The ToxCast data were explored for their ability to detect compound clusters of similar bioactivity and phenotypic properties. MacDonald *et al.* analysed the chemical-induced changes on Protein-Protein Interactions (PPI) by 107 ToxCast phase I compounds across 49 high-content protein fragment complementation assays.[92] Ward clustering of compounds showed that structurally similar entities clustered together, as well as compounds modulating the same therapeutic target classes. The authors, therefore, reported that known structure activity relationships were reproduced and suggested the mechanism of action of four drugs with unknown biological interactions. Moreover, the bioactivities of 976 compounds (ToxCast phase II) were analysed against 311 biochemical-target binding assays in cell-free environments. The results also demonstrated co-clustering of compounds sharing similar scaffolds as well as related targets.[93] 309 environmental[94] chemicals and later 776 environmental and pharmaceutical compounds[95] were screened against BioAMAP system as part of the phase I and phase II ToxCast program, respectively. The results demonstrated that the BioMAP screening system was able to classify compounds according to major toxic and therapeutic mechanisms upon clustering bioactivity profiles. These studies show that ToxCast assays can be successfully used in grouping chemical classes and providing insights into some mechanistic information about modes of action.

However, several studies have shown the negligible benefit of incorporating *in vitro* measurements to improve the predictive power of machine learning models.[96,97] For example, Abdelaziz *et al.* used chemical descriptors, ToxCast *in vitro* readouts and the corresponding perturbed biological pathways to predict 61 toxicity endpoints via various classifiers.[98] The authors concluded that only 10 endpoints, including rat developmental and hepatic toxicity, actually significantly benefited from using *in vitro* measurements as predictors in multi-domain predictive models[98]. Similarly, Thomas *et al.* applied more than 600 ToxCast *in vitro* assay readings for constructing predictive models against 60 *in vivo* endpoints.[99] It was reported that bioactivities from assays had a limited effect on the overall predictive power, with the

exception of cholinesterase inhibition. Dix *et al*, however, argued that using statistical models without incorporating chemical and biological knowledge will result in poor predictivity from *in vitro* measurements.[100] The authors recommended selecting relevant assay endpoints, applying relevant data aggregation and ensuring a sufficient size of compound data. For example, Lui *et al* reported a good overall accuracy when data of relevant hepatotoxicity endpoints were aggregated, and grouped into hypertrophy, injury and proliferative lesions. [101] It is often seen that building predicting machine learning models from *in vitro* measurements, in general, performs better when chemical properties are incorporated.[102] Yet, standard predictive models, especially non-linear models, provide little insight into the complex conditional associations between input data and toxicity outcomes.[103]

1.4 Case studies: *in vivo* endpoints of regulatory importance
1.4.1 Hepatotoxicity

Hepatotoxicity is an adverse effect manifested in liver injury induced by drugs or chemicals. Drug-induced liver injury (DILI) resulted in the withdrawal of several drugs postmarketing, such as troglitazone in 2000 and sitaxentan in 2010. Between 2005 and 2010, hepatotoxicity attributed to 14% and 12% of the termination of AstraZeneca's drug discovery projects due to unsatisfactory safety in preclinical and clinical studies, respectively.[104] Liver injury can be induced by the parent compound or its reactive metabolites in a predicted dose-response fashion, as in paracetamol, or unpredictable (idiosyncratic) such as the case of phenytoin.[105] One way drugs can initiate liver injury is via chemical reactivity, by forming covalent bonds with intracellular components, *eg* proteins, or by depleting reduced glutathione (GHS) content at which cells are incapable of averting oxidative stress. As a result of these chemical reactions, cytotoxic immune responses and mitochondrial toxicity can be induced. Additionally, injury can result from specific interactions that perturb biological targets and signalling pathways, such as inhibiting bile acid pump, *eg* BSEP, resulting in blocking bile flow, *i.e.* cholestasis.[106] Russmann *et al* have reviewed a three-step model for mechanisms in DILI, which involves one or more of the three initial injuries, namely, cell stress, mitochondrial impairment and induction of immune reactions.[107] These can further lead to inducing cell death pathways and severe mitochondrial dysfunction. Given the regenerative and detoxification capabilities of the liver,[108] capturing mechanistic pathways in chemical-induced hepatotoxicity is challenging.

In order to detect hepatotoxicity via *in vitro* models, relevant mechanistic endpoints are prioritized. However, this is not a straightforward task, primarily because hepatotoxicity

involves complex pathological pathways. Based on expert knowledge and known mechanisms of hepatotoxicity, several *in vitro* models have been developed, such as cytotoxicity, bile salts pump inhibition, mitochondrial impairment[109], Cytochrome P activity,[110] and covalent binding.[111] Alternative models consider immunological activity through changes in cytokines profiles.[112] Another *in vitro* technique used to screen for hepatotoxic compounds is high content screening (HCS), which utilizes multiple cellular measurements, such as changes in nuclear size, cell count, mitochondrial mass and cell membrane integrity, as biological responses.[113–115] Considering this diversity in biological mechanisms leading to hepatotoxicity, it is not surprising that *in vitro* assays which adopt a fraction of this bioactivity space would suffer from relatively low detection rates, with sensitivity ranging from 40 to 60%.[95,113–115] It has been hence recommended that *in vitro* models for hepatotoxicity should involve a broad range of bioassay endpoints that cover wider biological perturbation points and cellular phenotypes, in order to increase overall sensitivity.[114,116] However, which endpoints to consider, and how to combine their readouts, is not clear.

The second important parameter for an efficient *in vitro – in vivo* translatability is exposure, *i.e.* to what extent the compound actually reaches the site where it exerts its action[86], which is related to physicochemical properties.[117,118] This parameter poses considerable difficulties in predictive toxicology modelling, mostly because this information is difficult to obtain for large sets of compounds. As a proxy for exposure, parameters such as maximum plasma concentrations (C_{max}) and administered dose levels have shown improvements in the prediction of compounds *in vivo* toxicity. [109,114,115] For example, compounds had significant odds ratio for liver injury when levels of maximum plasma concentrations, C_{max}, are greater than $1.1 \mu M$ combined with a set of three bioactivities, namely cytotoxicity with an IC_{50} below $100 \mu M$, bile transport inhibition with an IC_{50} below $30 \mu M$, and mitochondrial impairment assays IC_{50} below $25 \mu M$.[109] It has also been recommended, as a rule of thumb, to have 100-fold separation between the concentration at which compounds are toxic in *in vitro* HCS assays and C_{max} value *in vivo*.[114,115] Additionally, some studies used total daily dose as an estimate for hepatotoxic liability combined with lipophilicity levels of compounds or *in vitro* bioactivity. For example, "the rule of two" states that the drug is likely to be hepatotoxic (with an odds ratio of about 3.9) if the lipophilicity expressed as logP is greater than 3 and the daily dose is higher than 100mg.[119,120] Dose was also combined with *in vitro* covalent binding to generate a zone classification system which has shown a clear separation of drugs causing idiosyncratic toxicity.[121] Another study has also shown that the likelihood for observing hepatotoxicity is

significantly higher when the administrated daily dose is higher than 100mg and the drug i) forms active GHS adduct, ii) has 5-fold IC_{50} decrease in Cytochrome P450 metabolism-dependent inhibition, or iii) binds covalently to proteins at levels higher than 200 pmol eq/mg protein.[111] These studies demonstrate how incorporating daily dose or C_{max} can be powerful in improving the predictivity of hepatotoxicity. Obtaining these measurements from *in vivo* experiments, however, is generally not feasible at early stages of drug development, or for profiling large number of compounds.

In conclusion, early assessment of hepatotoxicity can be challenging for two factors. Firstly, how to capture the minimum biological space in assays for maximum compound coverage, and secondly, how to account for exposure when *in vivo* parameters are unavailable. We attempted to address these challenges in Chapter 4 using ToxCast *in vitro* data and physicochemical properties.

1.4.2   Acute toxicity

Acute toxicity is a key adverse outcome required in registration and legalization of chemical compounds used in industry.[19] According to the R.7.4 documentation by European Chemicals Agency (ECHA)[122], it is defined as "health adverse effects following short-term exposure" upon oral or dermal administration as well as the inhaled route. These effects include irritability, changes in organ weight, sensitization or mortality. The conventional method for assessing systemic acute toxicity is via measuring doses in oral and dermal exposure, or concentrations in inhalation route, at which mortality occurs for half of the tested animals, namely (LD_{50}) and (LC_{50}), respectively.[6] The most common and accepted animal testing guidelines, which are described by OECD documentation, have already allowed for a significant reduction in the number of animals used to assess systemic toxicity.[19] In 1981, 30 animals per chemical were needed to derive a LD_{50} for the oral acute toxicity and 20 animals for dermal systemic toxicity.[19] The current protocol was fewer than 10 animals for both oral and dermal toxicities according to adjusted OECD guidelines.[123,124]

In order to reduce and replace traditional animal testing for acute toxicity in concordance with the 3Rs principles, alternative *in vitro* methods have been proposed. The current state of art for alternative approaches include *in vitro* assay for basal cytotoxicity using neutral red uptake (NRU) to identify starting doses in animal testing.[125] Another method uses embryo-larval fish models, which has shown high correlations with acute systemic toxicity in rodents.[126] Recent efforts have shifted into target-based bioactivities and chemical features to understand the underlying molecular mechanisms. Some of the key mechanisms identified for acute toxicity

include the interference with neurotransmission, ion channels, energy haemostasis, antioxidant pathways and cellular integrity.[6] The description of toxic effects through adverse outcome pathway (AOP) frameworks provides, in conjunction with computational models, a realistic application of mechanistic data in risk assessment. Several AOPs were proposed around these key events (KE) to describe acute mortaltity[127–129], which are published in the AOP database. These include the modulation of ionotropic GABA (gamma-Aminobutyric acid)[129] and ionotropic glutamate receptors[128], and the inhibition of acetylcholinesterase.[130]

The key element to fully utilize the AOP framework is to understand how to implement mechanistic information at the molecular level to capture an endpoint with complex and diverse pathological pathways as in acute toxicity. Current efforts focus on understanding the complications of each key event separately. However, a more plausible approach is to investigate the toxicity triggered by multiple key events, which is more realistic given the polypharmacology nature of bioactive compounds. Experimental validation of multiple target activities can be laborious. For example, knockout mice are used to inactivate an existing gene by altering its sequence.[131] When a knockout mouse of a specific gene is exposed to bioactive compounds, phenotypic effects can be compared with a control group which helps to form hypotheses about the effects of combining multiple activities. A study reported that aryl hydrocarbon receptor (AhR) activity, as a single variable, correlates with *in vivo* acute toxicity.[132] AhR knockout mice have shown a higher expression of α1D-adrenergic receptors and an increase in the maximal effect in aorta contraction by α-receptors agonists; noradrenaline and phenylephrine.[133] This indicates that combining AhR disruption with α-receptors modulators have greater implications as compared to single bioactivities. While this approach is very informative to understand the effects of multiple targets, it is not feasible to explore all possible interactions that can lead to observable toxicological effects.

In alignment with this concept, the AOP networks integrate multiple intersecting AOPs in order to explain and anticipate complex toxicological pathways.[134,135] AOP networks allow for two key objectives to be achieved. First, to anticipate the actual critical path in which molecular initiating events (MIE) eventually result in the observed outcome. Second, to understand the possible effects when multiple MIE/KE are combined for which their AOPs have downstream intersections, leading to additive, synergistic or antagonistic interactions.[136] The current challenge to fully utilize the AOP network framework is the incomplete information about KE/MIE linked by key event relationships (KER) resulting in either known or unexpected

phenotypic outcomes.[135–138] Therefore, it likely that the current state of knowledge will not capture all possible risks in cases where multiple stressors exist in the occasion of bioactive mixtures or compounds' polypharmacology.[138]

Thus data-based approaches, which consider statistical interaction of variables, can offer an alternative way to investigate significant polypharmacology as will be proposed in Part 2 of this chapter, and investigated in Chapters 4 and 5.

1.5 Filling knowledge gaps in toxicity modes of action can accelerate regulatory acceptance of new methods

Despite the rapid advancement in new testing technologies, the acceptance of these methods by regulatory bodies is still slow. In the US, regulatory acceptance and utilization of non-animal methods vary in formal legislative requirements.[6] For example, the guidance on nonclinical safety studies for the conduct of clinical trial recommends obtaining data from animal studies including lethality tests, as described in the ICH 2009.[139] Whereas, the EPA Office of Pollution Prevention and Toxins (EPA OPPT) would accept and recommend non-animal test such as QSAR models and *in vitro* tests.[19] In the EU, the Cosmetics Directive has implemented regulations in 2009, which imposed an animal testing ban on cosmetics ingredients.[2] Given the large numbers of chemicals yet un-tested in traditional animal studies, REACH has also recommended implementing non-testing methods including QSAR, read-across and physicochemical methods.

Although regulatory bodies in environmental industries have recommended the use of alternative methods, the approval of these methods for drug regulation to substitute animal testing is yet hindered by a number of factors.[140] For example, the lack of clarity in applying weight of evidence, as well as concerns about robustness and reproducibility, are some of the obstacles slowing regulatory acceptance.[140] The purpose of alternative methods is not to reproduce outcomes of animal studies, but to provide insights into possible risks relevant to human.[55] Given the large number of potential mechanistically-based models, effort should be streamlined for international harmonization. One way to accelerate this process is by using data to derive novel hypotheses about modes of action and use these insights to rationalize toxicity assessment. We will next review some data-based approaches used to derive mechanistic hypotheses about *in vivo* toxicity, then propose rule models as a potential method to address

challenges in capturing important mechanism of toxicity as well as significant polypharmacology profiles.

PART 2: DATA-BASED METHODS TO UNDERSTAND TOXICITY MODES OF ACTION USING *IN VITRO* DATA

This section will address data-based methods used to generate hypotheses about toxicity modes of action. While generating hypotheses from data is a broad topic, our focus will be on how to prioritize molecular and cellular key events in the toxicological pathway using large scale *in vitro* data. Examples of studies reporting important key events from data were given in section 1.3.1.3. In this section, we will first review some the most common methods to find significant associations between *in vitro* activity and *in vivo* toxicity. Then we will explore the potential of rule models to prioritize key events, as well as to understand conditional associations, such as significant polypharmacology profiles and important physicochemical conditions. Finally, we will address how to classify toxic compounds with confidence using the conformal prediction framework.

2.1 Common data-based methods to prioritize key events

In order to generate hypotheses about modes of action, interpretable feature space and algorithms are to be used. With regards to data, common sources include *in vitro* bioactivity measurements such as assay readouts and gene expression, as well as substructures and physicochemical properties. Some of the popular methods for exploring key events, including univariate association and regression methods, are reviewed below.

2.1.1 Univariate associations

Data-based endpoint prioritization has until this stage, been commonly performed by analyzing significant univariate associations and correlations between *in vitro* readouts and *in vivo* observations. These associations are frequently performed in a dichotomous manner to link a binary activity call in an assay, *i.e.* active or non-active, against a two-class toxicity label, *i.e.* toxic or non-toxic. The counts of these flags can be summarized in a contingency table, as presented in Figure (2-1). When the activity in assays is captured as continuous variables, such as IC_{50} (concentration at half maximum inhibition), a discretization step is performed by setting an arbitrary value as a threshold for activity.

		Toxic	Non-toxic
Assay activity	active	a	b
	inactive	c	d

Figure (2-1) 2×2 contingency table used to calculate dichotomous associations between *in vitro* activity and *in vivo* toxicity.

Examples of algorithms commonly used to mine univariate associations are summarized in Table (2-1). Prioritization of key events in this way is useful to identify predictive *in vitro* assays as well as enriched structural alerts. This process can also be performed during exploratory analysis as a primary step to select important features prior to advanced statistical modelling. The limitations of these methods, however, arise when important key events may not be predictive on their own and need to be accompanied by other conditions such as specific chemical space, permeability requirements or multiple biological perturbations. As a result, key events of toxicity importance may not top the rankings and hence, excluded or overlooked.

The analysis in this chapter attempts to overcome this limitation by using rule models. Rules will be used to mine conditional associations using chemical and biological properties, as will be explained in section 2.2 and demonstrated in the subsequent chapters.

2.1.2   Multiple linear regression

Linear regression describes the relationship between a response variable and predictor variable using linear parameters. [141] The simplest regression can be described as follows:

$$y = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_n x_n \quad [1]$$

where y is the numeric response, w_0 is the intercept, $(w_1, w_2, \ldots, w_n)$ are the coefficients for x_1, x_2, ..., x_n variables. The advantage of the linear regression is interpretability given that the absolute numeric values of the coefficients reflects the importance of the corresponding variables.

Table (2-1) Examples of common algorithms to estimate univariate associations for key event prioritization

Univariate method	Equation	Comment
Relative Risk/Positive Reporting Ratio	$$\mathrm{RR/PRR} = \dfrac{a}{a+b} \Big/ \dfrac{c}{c+d}$$	a,b,c,d correspond to values in Figure (2-1). This measure was used to prioritize ToxCast assays against multiple *in vivo* and human toxicity endpoints.[79,80]
Odd Ratio	$$\mathrm{OR} = \dfrac{a}{b} \Big/ \dfrac{c}{d}$$	a,b,c,d correspond to values in Figure (2-1).
Fisher test	$$p = \dfrac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{a!\,b!\,c!\,d!\,N!}$$	a,b,c,d correspond to values in Figure (2-1), N is the sum of a, b, c and d. P value for significance of association (probability of observing values as random). A threshold of 0.05 or lower is usually used to describe significance. Fisher test was used to identify ToxCast assay associated with tumors in rat.[15]
Mutual information	$$\mathrm{I}(X;Y) = \sum_{x \in X}\sum_{y \in Y} P(x,y) \log \dfrac{P(x,y)}{P(x)P(y)}$$	Mutual information between two variables X and Y of discrete values, where x∈X and y∈Y. mutual information was used to prioritize ToxCast assays associated with structural cardiotoxicity in human.[81]
Binomial distribution	$$P = \sum_{i=m'}^{n'} ((n'!)/(i!\,(n'\text{-}i)!)) \times (m/n)^{i} \times (1\text{-}(m/n))^{(n'\text{-}i)}$$	P value is derived from the probability density function of a binomial distribution. This measure is used to identify substructures that are enriched in active or toxic compounds.[74,142] n is the total number of compounds, m is count of active/toxic compounds, n' is counts of compounds with substructure and m' is count of active compounds with the substructure.

In cases where the response is dichotomous, *i.e.* a label accounting for 2 classes, such as toxic and non-toxic, the response (y) can represent the probability P of observing one class variable. In order to fit the probability estimates between 0 and 1, a sigmoidal function (σ) is used as:

$$P = \sigma(y) = \frac{1}{1+exp^{-y}} \quad [2]$$

As linear regression can capture multiple features at a time, this model has been applied to assess drug toxicity by combining *in vitro* activity and exposure. For example, assay measurements for covalent binding in hepatocytes and daily dose were used to predict idiosyncratic drug reactions.[121] Using regression analysis, the authors reported a three zone system that classifies multiple safety categories.

Additionally, linear regression equation can also capture feature interactions when the variables used are mutually dependent.[143] For a response y described by variables x and z with an interaction, the regression equation can be represented as follows:

$$y = w_0 + w_1 x + w_2 z + w_3 (x \times z) \quad [3]$$

where w_1, w_2, w_3 are the coefficients and $(x \times z)$ is the interaction variable. Lazic *et al* used hERG potassium channel activity from *in vitro* measurements as well as clinical C_{max} (maximum plasma concentrations) to derive probabilities of observing QT elongation on an electrocardiogram.[144] The authors used a linear model and considered an interaction between hERG activity and C_{max}. Additionally, the authors controlled the direction of association between hERG activity and QT elongation assuming a positive monotonic association between these variables.[144] Combining hERG activity and C_{max} has shown improvement in capturing cardiotoxicity in comparison to using hERG *in vitro* activity alone.

Although linear models help understand the contribution of multiple variables towards the target endpoint, they are sensitive to noise and may not perform well using highly correlated variables. A prior pre-processing step is to be conducted, most importantly, to capture the most important and causal predictor variables.[145] As shown in the example above, when few variables of direct association with the endpoint are used, the linear regression model can be very informative. However, these models may not be able to capture toxicity endpoints characterized by highly diverse underlying mechanisms.[101] High dimension feature space can be captured by nonlinear models such as Random Forest[146], support vector machines[147], and deep neural networks[148], which have shown high performance, but at the cost of

interpretability. In order to derive interpretable associations between large scale *in vitro* activity and toxicity, we used rule models. Rules can capture both conditional associations and allow for the description of toxic compounds that exert diverse modes of action. In the next section, we will explain generic rule models and algorithms. Additionally, we will propose modifications so as to overcome limitations of conventional rule models when capturing toxicity data.

2.2 Rule models to understand polypharmacology of toxic compounds

Rules can be defined as sets of feature conditions that are commonly associated with another variable and presented in if-then statements.[149] Hence, unlike univariate associations, rules find a set of features, when are all present (using an AND operator), the association with the response label is stronger. Here, we will define different rule types, and algorithms used to generate these rules. Then we will explore the limitations of conventional rule methods and how to overcome these limitations so as to extract meaningful *in vitro-in vivo* associations as well as to identify significant polypharmacology profiles.

Rules are meant to be simple representation of data. They can run in a supervised or unsupervised manner. The notion, association rules, run in an unsupervised fashion where one or more features are associated frequently with another feature, *i.e.* finding the frequent itemset.[150] Whereas, classification rules use a predefined response label to find discriminating condition sets in a supervised fashion. Table (2-2) compares classification rules with association rules.

Table (2-2) Comparison between association rules and classification rules.

	Association rules[‡]	*Classification rules*
Mining approach	Unsupervised	Supervised
Most common searching score	Frequency	Entropy or Gini index
Data representation	Transactions binary matrices	Matrices of categorical or continuous variables (rules) Transactions/binary matrices (emerging patterns)
Left hand side (LHS)	Combination of feature items	Combination of feature values (rules) or feature items (emerging patterns)
Right hand side (RHS)	One feature	A class label
Algorithm example	Apriori	C4.5, CPAR

2.2.1 Classification rules versus emerging patterns

Emerging patterns are a form of contrast data mining techniques which identify feature sets frequent in one class of the data, but not the others.[151] Rules and emerging patterns are describing the same concept. However, during the construction of conventional rules models, a discretization process is usually involved especially when continuous features are used. In contrast, emerging patterns take as input features described in true/false or categorical values. Hence, data are discretized beforehand and prior the generation of emerging patterns. For a set of binary variables, the common input structure for classification rules is a binary matrix as the left representation in Figure (2-2), whereas, emerging patterns handle transaction representation of data as an input (right in Figure (2-2)).

Given a dataset of X samples where $X_i \in X$ described by feature set Y, where Y= {A, B, C, …, Z} and class C, where C = {1, 2}. A rule R can be represented as follows:

[‡] Associations rules can be used for classification by pruning all rules not describing the class label as the RHS variable. An example for this implementation is the CBA[383] algorithm (Classification Based on Association).

```
IF predictor A > 0.1 AND predictor B < 0.5   THEN   class C=1
```

The example above describes a rule for data samples that are described by continuous variables. Each rule condition in this case is represented by two segments, which are the feature and a corresponding split. For example, the first condition in the rule entitled the predictor A and split specification at values greater than 0.1.

Binary matrix representation									Transaction lists representation		
Entry	*A*	*B*	*C*	*D*	*E*	*F*	*Label*		*Entry*	*Feature set*	*Label*
1	0	1	1	0	1	1	0		1	B,C,E,F	0
2	1	0	1	1	0	0	0		2	A,C,D	0
3	0	1	1	0	1	0	0		3	B,C,E	0
4	1	0	0	0	0	1	1		4	A,F	1
5	1	0	1	1	0	1	1		5	A,C,D,F	1
6	1	1	0	1	0	1	1		6	A,B,D,F	1

Figure (2-2) Pseudo dataset of binary features and class labels represented as a conventional binary matrix (left) commonly used to derive classification rules, and transactions lists (right) which are used to generate emerging patterns.

Emerging patterns aim to find combinations of discriminating features in a supervised classification problem.[149] For each entry in the data (a compound in the training), the features are described as 's', where s $\subset$ S, and label c $\subseteq$ C, where S are all feature set and C are all labels. An itemset is a subset of features which are present in one or more entries. If an itemset of features is present frequently among entries in the data set, such itemset is called a frequent pattern.

Measures to describe the performance of rules and emerging patterns are the same. The frequency of a pattern (*i)* or a rule R is described by the support ($Supp_i$), which is the count or fraction of entries in the dataset that support this patterns. In other words, the support is the count of entries where a pattern or a rule is fulfilled. The minimum support level, for an itemset to be frequent, is set by the user. Another key measure for the performance is the confidence (Conf) which describes how accurately a pattern or a rule can predict the class label of fulfilling entries. Confidence is presented in *Equation* [4] for a rule *i* describing label c:

$$\text{Conf}(i \Rightarrow c) = \text{Supp}(i \cap c) / \text{Supp}(i) \qquad [4]$$

Where $i \subseteq S$ and $c \subseteq C$, S is the group of all feature set, and c is a class label in all labels C.

2.2.2 Algorithms to generate discriminating rules

Many algorithms were developed to generate rules. Here we will focus on two algorithms, which are C50 for classification rules and CPAR for emerging patterns. These algorithms were used in Chapter 4 and Chapter 5, respectively. For comparison purposes, we applied a Random Forest model in Chapter 6 to evaluate the predictive performance of emerging patterns. Hence, the Random Forest algorithm was also reviewed below.

2.2.2.1 C5.0

C5.0 algorithm was developed by M. Kuhn[152] based on the C4.5 algorithm introduced by Quinlon in 1992.[153] Rules, in the C4.5 algorithm, are constructed by collapsing an unpruned decision tree. The decision tree, in a classification problem, consists of nested if-then statements of predictors to describe a class label. A simple tree can be represented as follows:

```
IF predictor A > 0.1 then
|      if predictor B < 0.5    THEN    class C=1
else class C=2
```

This tree can be converted into the following rules:

```
IF predictor A > 0.1 AND predictor B < 0.5    THEN    class C=1
IF predictor A <= 0.1    THEN    class C=2
```

As each rule represents a unique path in the tree (a route to a terminal node), the number of conditions in a rule is equal to the number of nodes in that path.

The first step in constructing a tree is finding the feature set and split points that achieve maximum discrimination between classes (highest purity). To find the best split, searching measures quantify the purity in order to rank the generated splits. C4.5 uses the information statistic, entropy, which is originated from the information theory[154]. First, the information statistic is calculated before and after the split using *Equation* [5].

$$info = -[\, p\, \log_2 p + (1-p)\log_2(1-p)\,] \qquad [5]$$

Where *info* is the information statistic, p is the probability of the first class and $(1-p)$ is the probability of the second class.

To estimate the improvement in purity, the difference in the information statistic before and after the split is calculated and represents the entropy, *Equation* [6]. The higher the entropy, the higher the ability of the condition to discriminate between the classes.

$$information\ gain\ (entropy) = info\ (before\ split) - info(after\ split) \qquad [6]$$

The splits are gradually added, in a nested tree, to eventually create homogenous partitioning of data classes.

The second step is collapsing the tree into rules. Then, rules are pruned using a pessimistic pruning approach to remove conditions with the least improvement in rule accuracy. The error is estimated for the unpruned rule (baseline error) as well as for the rule except for one condition. The condition with the highest error rate is removed iteratively, given that the error rate is greater than the baseline error. This is repeated until the maximum error rate of any conditions is greater than the baseline error or until all conditions are removed, *i.e.* the rule is removed. The remaining rules are then prioritized to select the minimum representative set. The minimum description length (MDL) metric selects the simplest rule combination using search methods such as simulated annealing. Finally, the rules are ordered by accuracy. C5.0 algorithm adds on to the C4.5 features such as boosting and unequal cost.[141,152] Also in C5.0, rules are not ranked. Instead, predictions are made according to a voting system which is weighted by confidence.

2.2.2.2 CPAR (Classification based on Predictive Association Rules)

CPAR is an emerging pattern algorithm developed by Yin and Han,[155] extended from the PRM (Predictive Rule Mining) and FOIL (First Order Inductive Learner) algorithms. FOIL[156], which was developed by Quinlan, uses gain (*Equation* [7]) as the searching measure to build rules. The best feature in gain is added, one by one, until the maximum rule size is reached.

$$\text{Gain} = |P^*| \left(\log \frac{|P^*|}{|P^*| + |N^*|} - \log \frac{|P|}{|P| + |N|}\right) \qquad [7]$$

Where $|P^*|$ and $|N^*|$ are the counts of positive and negative instances after adding the feature condition to the rule, whereas, $|P|$ and $|N|$ are the counts of positive and negative instances before adding the feature.

A key feature in the FOIL algorithm is that example entries are removed from the training data as soon as a fulfilling predictive rule is generated. PRM (Predictive Rule Mining) extends from FOIL by applying a decay factor to the used examples instead of removing these instances. CPAR adds up to these features by considering multiple options in the gain-based feature selection, which results in generating multiple rules simultaneously. For example, if the top gain feature co-occurred with other features with 0.99 gain similarity, multiple rules are

produced considering these options. Therefore, CPAR can be seen to combine the properties of the exhaustive association classification and the conventional greedy rule-based algorithms.

2.2.2.3 Random Forest

Random Forest as described by Breiman (2001)[157], is an ensemble of m number of unpruned tree models, each producing an independent prediction. Using m number of trees, m number of predictions are then aggregated for each instance, *i.e.* via number of votes in classification or average prediction in regression. The Random Forest is based on the CART[158] algorithm, originally developed by Breiman *et al* in 1984, which uses Gini impurity as the search metric for the best split. Gini index which was introduced by Breiman *et al*[158] is presented in *Equation* [8].

$$gini\ index = \ p1\ (1-p1) + \ p2\ (1-p2) \qquad [8]$$

where p_1 and p_2 are the probabilities of class 1 and class 2, respectively. In two class data, the Gini index is equivalent to $2p_1p_2$. At equal distributions of the two labels, *i.e.* $p_1=p_2$, the Gini index value is maximum and equal to 0.5. Therefore, as the purity increases the value of the Gini index decreases.

The random component in the Random Forest algorithm is introduced by training trees on a random subset of predictors/features P, referred to as m_{try} using random bootstrap samples. This randomness prevents the low variance in predictions as a result of highly correlated trees, *i.e.* trees of similar structures. The recommended value for m_{try} is $1/3$ P for regression and $\sqrt{P}$ for classification, where P is the number of features.[157] The higher the number of trees, the lower the error rate of the ensemble model. However, as a rule of thumb, a recommended number of trees is between 500[159] and 1000.[141]

2.2.3 Limitations of conventional rule models

Rules are useful to understand patterns within data and can be highly interpretable. They find conditions at which high purity of a label class is achieved. In other words, rules with sufficient discriminating power can be useful to extract the combination of related features which together can translate better into the outcome property.

Despite the interpretability of rule models, they suffer from two key limitations that may hinder their applicability for extracting meaningful information from toxicity data. Firstly, conventional rule algorithms do not imply any constraints on the direction of the association between activity *in vitro* and toxicity *in vivo*. The general assumption in the relationship between the activity in an *in vitro* assay and the *in vivo* effect has often been a positive direction. In other words, *in vivo* toxicity is related to activity against a target or perturbation of a pathway rather than inactivity. This assumption is applied when *in vitro – in vivo* associations are interpreted or univariate correlations are derived.[80,101] Also, this directional association is fundamental in developing AOPs from a series of key events (see Figure (1-5)).

In this thesis, constraints were applied in rule models to preserve a positive direction for the association between *in vitro* activity and *in vivo* toxicity. Hence, the presence of toxicity is linked to the presence, (but not absence) of *in vitro* activity. Whereas, the absence of toxicity is related to the absence of *in vitro* activity. These constraints were applied in two ways, i) pruning rule space generated by C50 algorithm in Chapter 4, and ii) pruning the data space prior to running CPAR algorithm in Chapter 5.

Secondly, when a high dimensional feature space is used against a toxicity label of complex mechanistic pathways, the number of generated rules can be large and difficult to interpret. This was handled in two ways: i) selecting the most representative rules which have the maximum assay diversity and maximum compound coverage, and ii) combining rule models with network analysis to visualize and extract important patterns. The first approach was used in Chapter 4 for hepatotoxicity and the second was used in Chapter 5 for acute toxicity.

2.3 Using Conformal prediction framework to assess uncertainty in toxicity classification

The main challenges for accepting alternative methods in regulatory toxicology are the low interpretability of common statistical models, as well as the uncertainty level of predictions made.

One of the most common data-based approaches in predictive toxicity is QSAR models. QSAR is a mathematical model that associates molecular representations of compounds with an important chemical property or biological endpoint.[160] The development of efficient molecular representations, such as 2D molecular fingerprints by Klopman, has also assisted in the

growing utilization of QSAR models.[161] The reader can refer to published articles for detailed accounts of molecular representations and machine learning algorithms used for QSAR.[160,162,163]

In order to improve model interpretability, the QSAR framework in toxicity prediction has evolved to incorporate knowledge in biological pathways and perturbation points.[100] The so-called MoA QSAR (Mode of Action QSAR) combines biological assay results with structural features as predictive descriptors.[160] For example, readouts from ToxCast *in vitro* assays were used with chemical descriptors to predict hepatotoxicity in rats.[101] It is argued that associating *in vivo* phenotypic effects of compounds from their biological information alone can be limiting.[146] In fact, utilizing chemical information in addition to bioactivity data can be critical in deriving meaningful mechanistic interpretations.[146]

A key step in the generation of QSAR models is defining the applicability domain, which assesses the chemical space where reliable predictions can be made.[164] In concordance with this principle, the OECD has determined a set of specific criteria for the use of QSAR in toxicity prediction in regulatory applications.[165] These criteria include defining the applicability domain, and the ability to provide mechanistic interpretation.

 The applicability domain aims to assess the similarity of the test points with those used in the training the model.[166] Yet, a more practical assessment of the model performance should address the uncertainty or confidence of each prediction. In this context, the conformal prediction framework has been proposed as a technique to allow predictive models to only generate valid predictions within a given level of confidence.[167] This framework has a great potential in the field cheminformatics[167] and drug discovery[168] where a robust measure of uncertainty is required for decision making. The level of confidence is set by the user and it controls when a model can assign a prediction to a new instance. Svensson *et al* have applied conformal prediction using a Random Forest model to predict the compound cytotoxicity in different cell lines.[169] The conformal models achieved 80% average accuracy and 87% average coverage at a confidence level of 80% despite the significantly unbalanced cytotoxicity labels. Moreover, Ji *et al* have developed a web server, eMolTox, to predict 174 *in vitro* and *in vivo* toxicity endpoints for compounds under a conformal prediction setting.[170] Hence, evaluating the reliability of predictions made by models is very insightful in toxicity assessment.

In order to have an unbiased evaluation for the confidence of the conformal model, a reference dataset which is independent to the training set is used, *i.e.* the calibration set. Figure (2-3)

represents a basic workflow for conducting a conformal predictor using Random Forest. The training data are split into proper training and calibration sets. The proper training set is used to develop the underlying model, whereas, the calibration set is used as a reference to rank the test samples.
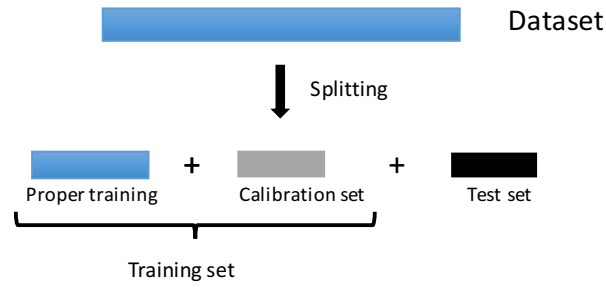
The conformal predictor uses a non-conformity measure (NCM) to make predictions and estimate the confidence. This numeric measure correlates with the likelihood of the class label in question. In other words, the NCM represents a way to describe the similarity of new instances in comparison to the training set. Some of the commonly used NCM include the class probabilities from machine learning models such as Random Forest[169] as well as similarity measures used in applicability domain evaluation.[171] With respect to the calibration set, test instances are ranked. The level of accepted rank is determined by the confidence level set by the user. For example, a confidence level of 80% means that the models will only give a valid class label for the new instances if its NCM rank falls within the top 80% of the calibration data. On the other hand, if the NCM of new instance does not fall within that region, the model fails to assign a valid label, and hence, it is considered outside the confidence level of the conformal model.

When the data set is significantly unbalanced, it is preferred that the rankings are performed for each class label independently as in the Mondrian conformal prediction, which was introduced by Vovk *et al.*[172] In a binary classification problem, conformal predictor models produce four types of outcome for each instance. For example, given arbitrary class labels of A and B, the possible outcomes are either 'A', 'B', 'Both' or 'None'. In terms of a conformal predictor, a correct classification is achieved when the correct label is assigned, hence 'A' or 'B' in addition to 'Both'. The latter label represents cases where the instance satisfies the model's requirements of both classes for a given confidence level. Whereas, the 'None' label means that the model cannot assign a confident classification for an instance because it does not satisfy the model characteristics for either class.

The performance of conformal models is evaluated by the validity and efficiency. Validity represents the accuracy of the conformal model, *i.e.* the rate of correct classification including the 'Both' label. Whereas, the efficiency represents the percentage of single class predictions, *i.e.* the ability of the models to discriminate between classes with confidence.

In Chapter 6, the conformal prediction framework was applied to evaluate the uncertainty of predictions made by rule models in comparison to a benchmark Random Forest model.

A) Data splitting

Dataset

Splitting

Proper training + Calibration set + Test set

Training set

B) Training model (Random Forest)

Proper training

Learning → Predicting*

Trees predicting toxicity

Trees predicting absence of toxicity
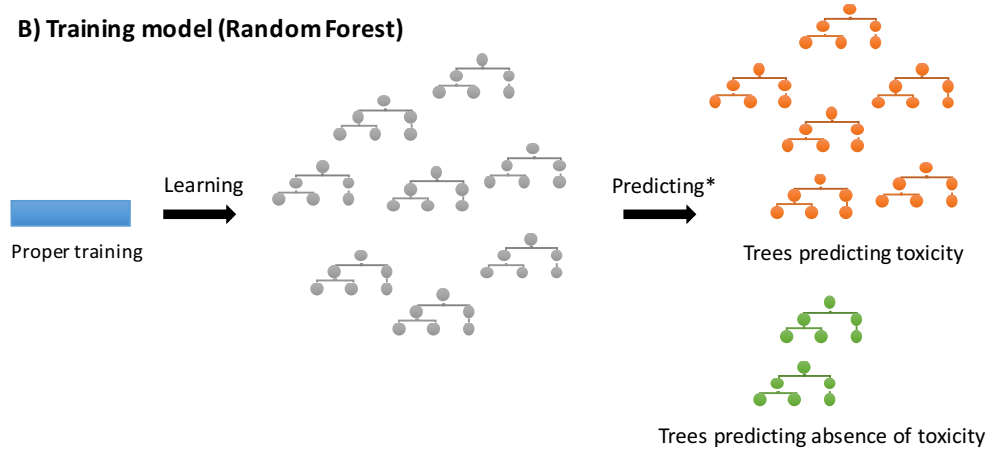
C) Conformal Predictor

Probability from RF (P)

α_{toxic} = (P) = N# of trees predicting toxicity

$\beta_{non\text{-}toxic} = 1 - \alpha_{toxic}$ = N# of trees predicting absence of toxicity

α β

α_1 β_1

α_i β_i

α_n β_m

Toxic Non-toxic

Calibration set
Test
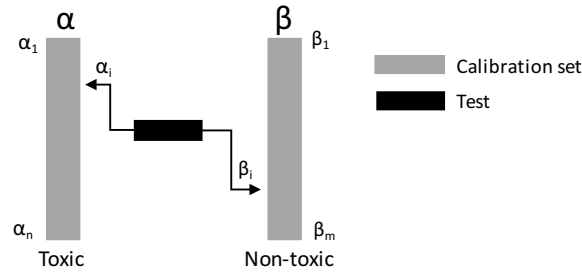
* Predicting Calibration and test sets

Figure (2-3) Mondrian conformal prediction using Random Forest as an example for the learning algorithm. The first step is to split the data into training and test set, where the training data is further split into proper training and calibration sets (step A). The proper training is used to build the machine learning model (Random Forest in step B), whereas the calibration set is used as an unbiased reference for the test set. A key component in the conformal model is the non-conformity score, which is a continuous variable that measures the similarity of the predicted instance to the calibration samples with regards to each class label. In the case of Random Forest, the probabilities are used as the nonconformity score (step C). Two rankings for a test point are obtained for the two class labels (α_i, β_i). Based on a predefined level of confidence and the two obtained rankings, a prediction is made or rejected.

2.4 The application of data-based methods in this thesis

The aim of this thesis was to improve our understanding of chemical-induced toxicity by extracting meaningful knowledge from data. A better understanding of the mechanistic basis of toxicity allows to utilize *in vitro* methods more efficiently, and eventually accelerates the regulatory acceptance of the alternative methods. The conventional data-based approach in deriving mechanistic information on toxicity is conducted from interpreting univariate associations between molecular bioactivities and *in vivo* observations. The novelty, in this thesis, extends this approach by extracting important interacting stressors, which together have greater associations with *in vivo* toxicity, namely hepatotoxicity and acute toxicity. These interacting stressors may represent significant polypharmacology profiles or combinations of biological activities and chemical properties.

While *in vitro – in vivo* associations are often directional and involve multiple conditions, there is a shortage in studies that consider these constraints in a multivariate non-linear level. The analysis here is first to explore these associations using i) rule models with constraints over *in vitro* readouts, and ii) in combination with chemical properties in order to capture chemical reactivity and kinetics. This approach allows to inform meaningful conditional associations with regards to predictive bioactivities and chemical properties.

First, the ToxCast dataset will be explored in Chapter 3, as it represents the backbone of *in vitro* data used throughout the thesis. Here we explained the key patterns observed in target-based and phenotypic assays, and how these patterns relate to the cytotoxic burst phenomenon reported in literature,[85] and mechanisms of hepatotoxicity and acute toxicity, explored in Chapter 4 in Chapter 5, respectively. In Chapter 4, we will prioritize assays predictive of hepatotoxicity, as this adverse effect significantly contribute to failures of drug discovery projects. Current *in vitro* models have low detection rates due to incomplete biological coverage.[95,113–115] Hence, we will compare assays prioritized by rules with endpoints used in four commercial *in vitro* setups for hepatotoxicity. Additionally, we will explore how physicochemical properties used in rules can act as proxies for exposure and can help improve the translatability of *in vitro-in vivo* associations. Next, in Chapter 5, we will investigate modes of action in compounds which induce acute toxicity, given that the *in vitro* mechanistic screening for this remains immature. Significant polypharmacology profiles of toxic compounds will be explored by combining rules and network analysis. Next, to examine the predictivity of rules generated in Chapter 5, we will compare the performance of rules with a

benchmark Random Forest model in Chapter 6, using a Mondrian conformal prediction framework.

3 EXPLORING CORRELATION PATTERNS OF SPECIFIC AND UNSPECIFIC PATHWAY PERTURBATIONS IN TOXCAST ASSAYS

3.1 Introduction

In order to properly utilise bioactivity data in toxicity assessment, it is important to understanding *in vitro*-related biases and confounders. For example, an observed phenomenon, "cytotoxic burst" (CTB) is reported in which cytotoxic compounds show high promiscuity by activating a broad range of target specific assays.[84] While 1.3% of assays are activated by non-cytotoxic compounds, 12% of assays are activated by cytotoxic compounds.[85] This indicate that observing cytotoxicity at a given concentration can bias the measurements of target-specific assays, and hence any interpretation from these can be misleading. A list of ToxCast assays that are frequently activated below CTB concentration were investigated by Fay *et al*.[84]

It is argued that toxicity *in vivo* can be explained by either i) specific target activities below cytotoxic concentrations or via ii) unspecific cytotoxicity and cell stress pathways. In order to examine the associations between specific and non-specific *in vitro* effects, we analyse the variance and similarity within the ToxCast assay library. We group the assays into target and phenotypic based according to the intended target type. The target focussed assays were designed to capture major biological pathways that have known relevance to toxicological events, either by direct chemical-protein binding or indirect effects on up- or downstream proteins. Phenotypic assays screen a broad-spectrum of cellular changes, with focus on cytotoxicity, proliferation and the cell cycle as well as other intracellular changes related to organelles and proteome production.

We will show that cytotoxicity assays have shown broad correlation with target specific readouts, which align with the CTB phenomenon. Also, we report that assays predictive of toxicity show specific patterns in the bioactivity space.

3.2   Materials and Methods

3.2.1   Dataset

The ToxCast *in vitro* data[173] were used to extract bioactivity measurements for 673 compounds against 821 assays in concentrations at half maximum activity (AC_{50}). The matrix was curated by removing chemical mixtures and inorganic constituents. The overall percentage of complete AC_{50} readouts in the matrix was 9%. Therefore, in order to increase data density, a subsets was retrieved by keeping assays with at least 20% complete AC_{50} measurements. An overall of 109 assays were selected for this study. Then assays were categorised into two parts: target-based and phenotypic-based assays, according to the "type of intended target" in the ToxCast assay annotation file. If the intended target is a protein, the assay is classified as target-focussed, otherwise, it was considered to be a phenotypic assay. The target-based and phenotypic-based assay matrices contained 56 and 53 assays, respectively. The AC_{50} readout values ranged from as low as fractions of nM up to just below 1 mM concentrations (Appendix A, Figure (A-1) and (A-2)). Detailed information on some targets is given in Appendix A, Table (A-1).

3.2.2   Analysis of variance using principal component analysis (PCA)

Principal component analysis converts original dataset variables into their uncorrelated linear combinations, which are called principal components.[141] To test the diversity of the assays matrices, a PCA was performed. First, the data were pre-processed via scaling and centring using the caret[174] and e1071[175] packages in R[176] environment (version 3.3.3). The data were sclaed by converting measurements into Z-scores (subtracting the mean from the values, then dividing by the standard deviation). Then the normalized values were centred via Box-Cox transformation statistic. Next a PCA was then conducted using the FactoMineR[177] R package. The percentage of variance explained by each principal component for both target and phenotypic assays were visualised.

3.2.3   Cluster analysis

Clustering is a multivariate technique that assembles apparently unrelated objects into fewer, similar or homogenous groups called clusters.[5,179] Hierarchical clustering was performed on Spearman coefficients derived similarity matrices using the complete linkage method. Spearman correlation is a non-parametric rank method [180], derived from the following

equation: $\rho = 1 - \frac{6\sum d^2}{n(n^2-1)}$ [9]

The results were visualized in a heatmap plot. Clustering and visualization were conducted using the function 'heatmap.2' in the (gplots)[181] package under R programming environment in default settings.

3.3   Results and Discussion

3.3.1   PCA shows high diversity among phenotypic assays in ToxCast data

In order to examine the variance in ToxCast assays, we performed PCA analysis on 109 assays as well as on subsets of 56 and 53 target and phenotypic assays, respectively. The percentage of variance was plotted for assays in Figure (3-1). The PCA on the overall set had shown that 70% of the variance was described by 33 components, meaning that the ToxCast assays are diverse and highly dimensional. The Figure also show that variance is higher among phenotypic assays in comparison to target-based assays. The first principal component of phenotypic and target assays explains around 15% and 25% of the total variance, respectively. Therefore, we can conclude that phenotypic assays carry higher overall variance and can provide novel insights on toxicity which may not be captured by target assays. Recent efforts were directed into understanding specific target and pathways perturbation.[58] However, it is argued that *in vivo* toxicity may also result from non-effects such cytotoxicity and cell stress.[84]
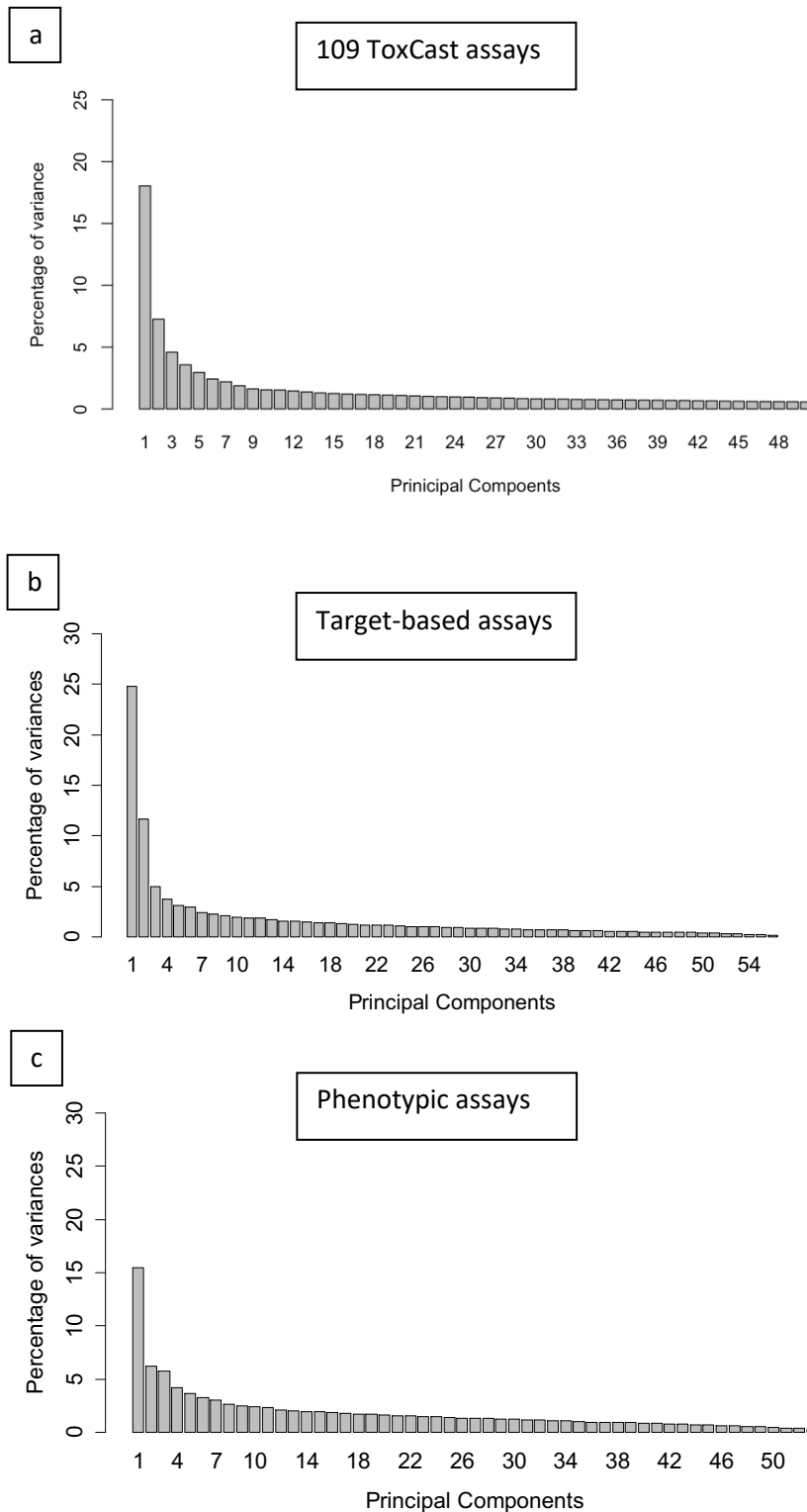
Figure (3-1) Percentage of variance captured by principal components with regards to: a) 109 ToxCast assays, b) 56 target-based assays, c) 53 phenotypic assays. In (a) the accumulated variance of 70% require 33 principal components. Capturing variance in a large number of principal components indicates the low variance described by principal components and, hence, high diversity within the ToxCast data. The first two principal components in target-based assays explain around 25% and 12% variance respectively, whereas the first and second components of phenotypic assays explain around 15% and 6% of variance, respectively. Therefore, phenotypic measurements are more diverse than readouts from target-based assays.

3.3.2 Assays relevant to toxicity show specific patterns in cluster analysis

Given the challenges in analysing *in vitro* data for toxicity assessment due to the confounding effect of cytotoxic burst, we were motivated to examine the assay space of ToxCast with regards to correlation patterns between target and phenotypic assays. A similarity matrix between the two classes was created by calculating the Spearman correlation coefficient between each target and phenotypic assay pair. The count of mutual readouts for each assay pair are presented in Figure (3-2). The number of complete pairwise measurements between target and phenotypic assays ranged from 25 up to 244, with a mean of 90 common data points. This means that there were enough mutual data points for each assay pair to generate meaningful and informative similarity values.
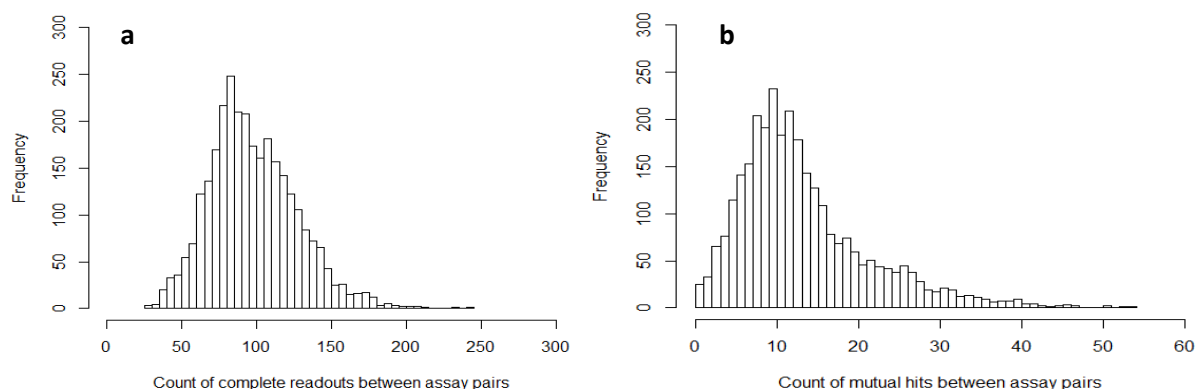
Figure (3-2) a) The frequency of complete pairwise measurements between target- and phenotypic-based assays. b) The frequency of mutual hits counts in assay pairs where both possess AC_{50} activity $\leq 10\mu M$. Figures show that there were sufficient mutual readouts to generate informative similarity scores.

Unsupervised hierarchical clustering was performed on the Spearman-derived similarity matrix and visualized as a heatmap in Figure (3-3). Assays of similar activity profiles, co-cluster. The horizontal axis in Figure (3-3) illustrates the co-clustering of similar target-target assays, whereas, the vertical axis illustrates co-clustering of phenotypic assays. We found that assays correlation values ranged from -0.44 to 0.71. The average Spearman correlation coefficient was 0.2, whereas only 25% of the assay pairs had shown magnitudes above 0.34. These results align with findings from PCA and indicate assay diversity.
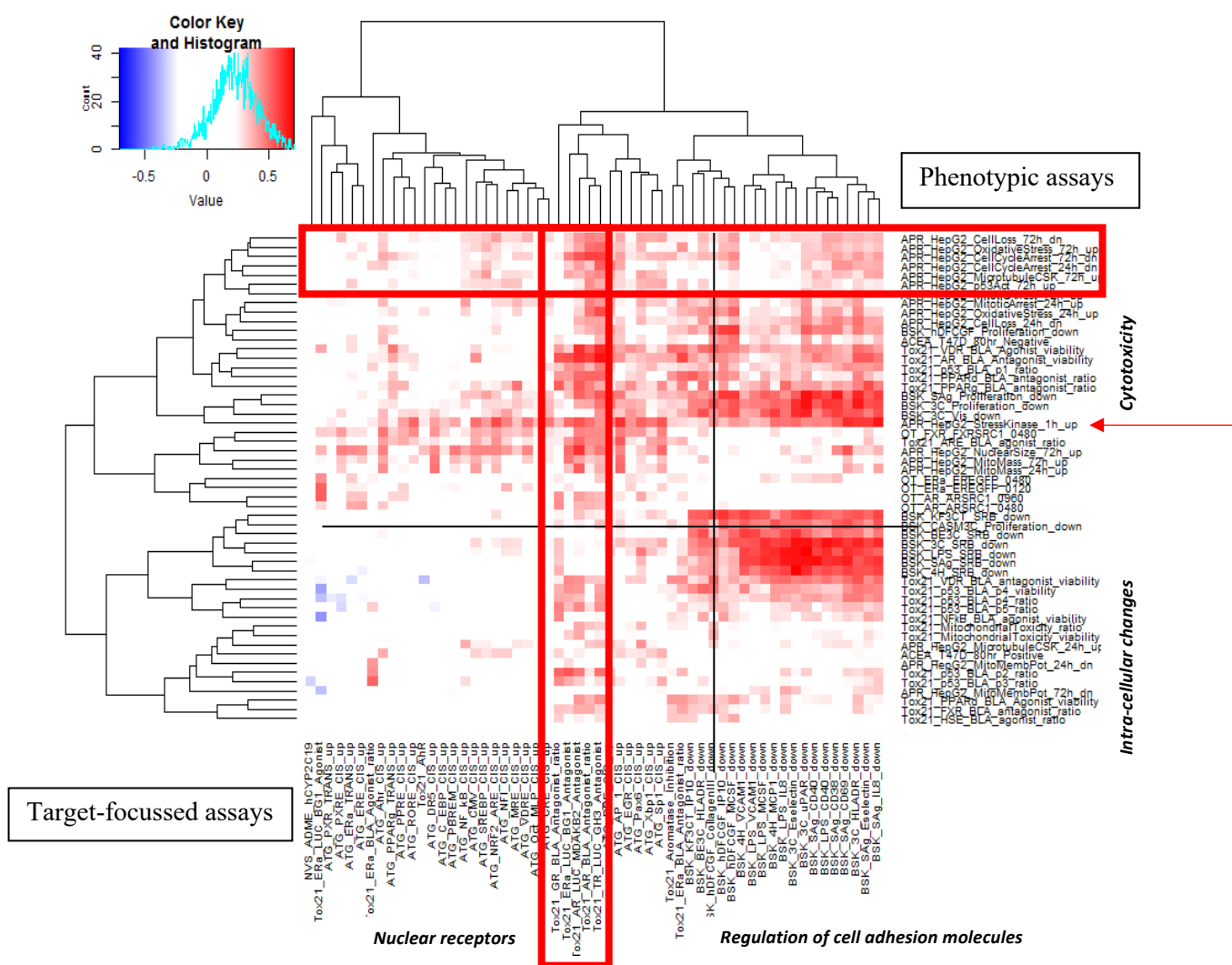
Figure (3-3) Heatmap plot of the hierarchical clustering of target- and phenotypic-based ToxCast assays along the horizontal and vertical axes, respectively. The colours in the heatmap plot were set to visualize correlation magnitudes that are under -0.3 in blue or over 0.3 in red. Clustering divides the matrix into four major sectors describing the chief biological classes of assays. Assays for cell loss and cell cycle arrest correlate broadly with target specific assays used in this analysis, aligning with the cytotoxic burst phenomenon. Tox21 nuclear receptor assays have wide correlations with the majority of phenotypic assays. ATG nuclear receptor assays, however, have weak correlations with most phenotypic assays. The arrow point to stress kinase assay which broadly correlates with targets assays in the dataset.

The biological-dependent clustering was seen for assay families as well as individual assays. Phenotypic assays clustered roughly into two major groups; assays for cytotoxicity and assays for changes in intracellular content. For example, the first phenotypic cluster involved in screening for cell loss, oxidative stress, cell cycle arrest, mitotic arrest as well as increases in nuclear size and mitochondrial mass on various cell lines. The second sub-cluster included assays involved toxic effects on microtubules and mitochondria, as well as changes in the total protein content. Target-based assays also clustered into two groups; the first related to DNA

and nuclear receptors binding, while the second mainly consisted of assays that screen changes in cell adhesion proteins. Details on some of these targets can be found in Appendix A, Table (A-1).

In general, the heatmap in Figure (3-3) shows three key bioactivity patterns: i) broad-scale correlations in bioactivity space of assays relevant to toxicity, ii) co-clustering of assays of relevant biological pathways, and iii) co-clustering of assays conducted by the same platform.

First, the broad-scale correlation can be seen for three groups, namely, cytotoxicity (cell loss and cell cycle arrest), endocrine disruption and the stress kinase assay. We can observe from the heatmap in Figure (3-3) that phenotypic assays describing cell loss and cell cycle arrest have shown broad correlations with the set of target assays used in this study. This observation agrees with the cytotoxic burst phenomenon reported in literature,[85] which is defined as broad intracellular machinery activation due to unspecific effects.

Additionally, assays for endocrine disruption, listed below, correlated broadly with phenotypic measurements at an overall magnitude of 0.32 up to a maximum of 0.67 Spearman coefficient.

- Tox21_GR_BLA_Antagonist_ratio (glucocorticoid receptor)
- Tox21_ERa_LUC_BG1_Antagonist (estrogen receptor)
- Tox21_AR_LUC_MDAKB2_Antagonist (androgen receptor)
- Tox21_AR_BLA_Antagonist_ratio (androgen receptor)
- Tox21_TR_LUC_GH3_Antagonist (thyroid hormone receptor)

These nuclear receptors control a variety of metabolic and developmental functions by sensing steroids and thyroid hormone.[182–185] *In vitro* screening for endocrine disruption is classically used to predict reproductive and developmental toxicities.[186] However, endocrine disruption is also associated with a broad range of chronic effects including cardiovascular diseases, neurodevelopment disorders, carcinogenicity and cognitive functions.[187] In a data-derived approach, Hu *et al* investigated the associations between ToxCast *in vitro* measurements and human effects from Comparative Toxicogenomics Database (CTD).[80] The authors reported significant associations between assays for endocrine disruption and four major toxicity categories in human, namely, hepato-, cardio-, renal- and neuro-toxicities. Our analysis of mechanisms in hepatotoxicity in Chapter 4 and acute toxicity in Chapter 5 have reproduced these patterns. As will discussed in Chapter 4, endocrine disruption represented a key independent cluster of hepatotoxic compounds, which included activities against estrogen, androgen and glucocorticoid receptors. In Chapter 5, disruption of thyroid hormone receptor

had shown significant synergistic links with key events in chemical-induced acute toxicity. Despite the broad correlation of these assays in the bioactivity space, seen here, as well as in the disease space, reported in literature,[80,187] their utility as predictors for organ-specific effects is limited.

Another assay that had shown a unique pattern was the stress kinase assay on Hep G2 cell lines (APR_HepG2_StressKinase_1h_up). This assay had a broad-spectrum correlation with target-based assays, at an average of 0.41 Spearman coefficient up to a maximum of 0.66. Additionally, as will be described in Chapter 4, the stress kinase assay has the strongest correlation among the pool of ToxCast assays with *in vivo* hepatotoxicity. Upon environmental stimuli, such as exposure to toxins, stress-activated protein kinases are activated and consequently phosphorylate a range of transcription factors.[188] Eventually, a variety of gene expression events are initiated in response to the external stimuli. This explains why the stress kinase up-regulation assay broadly correlated with the majority of target-based assays, including nuclear receptors. Overall, assays that show broad correlations in the bioactivity space seem to act as key predictors for *in vivo* toxicity.

Besides global correlations, the second observation is the co-clustering of assays of relevant cellular pathways co-clustered. For example, phenotypic assays associated with changes in organelle conformation, such as APR_HepG2_NuclearSize_72h_up, APR_HepG2_MitoMass_72h_up and APR_HepG2_MitoMass_24h_up, correlated with a large number of DNA and nuclear receptor binding assays. Such observation can be explained by the fact that nucleus size is proportional to the compactness of chromosomes.[189] Similar behaviour is likely to occur in mitochondria during the Mitochondrial DNA expression for energy production.[190] Also, changes in cell adhesion molecules correlated with a broad range of cytotoxicity assays, oxidative stress screens and reduction of cellular protein content screens. Cell adhesion molecules are responsible for attaching cells together to maintain tissues integrity, cellular functions and anti-inflammatory responses.[191] Inhibition of cell adhesion molecules may be associated with cytotoxicity via the disruption of cellular integrity and communication. BSK assays that screen for chemical-induced changes in the regulation of cell adhesion molecules, were associated with renal toxicity.[80] Our results demonstrated the co-clustering of CIS and TRANS regulatory elements assays of the same transcription factor. This is evident by the alignment of ATG_PXR_TRANS_up next to the ATG_PXRE_CIS_up, ATG_ERa_TRANS_up and ATG_ERE_CIS_up as well as ATG_PPARg_TRANS_up and ATG_PPRE_CIS_up. CIS and TRANS acting elements are DNA sequences that regulate the

transcription of a specific gene. However, transcription factor response elements, CIS, regulate gene expression by acting as binding sites for transcription factors.[192] TRANS regulatory elements are DNA sequences that encode the transcription factors.[192] The co-clustering of CIS and TRANS assays is a pattern previously in the literature.[79] Therefore, the similarity between relevant assays validates the bioactivity content of ToxCast measurments.

The third key observation in Figure (3-3) implies that assays supplied by the same platform cluster together. The majority of APR assays clustered together, especially HepG2 screening for cell loss, oxidative stress, cell cycle arrest and mitotic arrest. Both phenotypic- and target-based assays, from the Bioseek platform, co-cluster together as well as mutually correlate at Spearman score of over 0.5. Moreover, Bioseek's target-based assays show moderate, albeit weaker, correlations with APR-HEPG2 assays. It is clear that not only does the BSK platform co-clustered together but also the assays performed on the same type of cell line were adjacent to each other. For example; fibroblasts, mixtures of peripheral blood, mononuclear cells, endothelial cells and endothelial cell lines co-clustered. Given that each platform performs specific assay setups with regards to cell lines, solvents and detection techniques, it is possible to achieve similar profiles within the same platform. Another factor is the relatively limited space of modes of action and cellular pathways in each platform. This is exemplified by BSK cell adhesion molecules assays and DNA and nuclear receptor binding assays by ATG. Overall, the similarity within platforms can be attributed to the conserved biological space as well as the bias from the assay setup.

3.4   Conclusions

The similarity and variance of the ToxCast bioactivity space were investigated. The diversity and biological patterns in the ToxCast *in vitro* measurements indicate its utility to explore key events in adverse outcome pathways and to understand toxicity modes of action. The higher variance among phenotypic assays, compared to target assays, demonstrate that these assays capture complex bioactivity profiles. Hence, phenotypic measurements can be useful predictors of toxicity key events which may not be captured by specific target interactions.

The results also agree with the cytotoxic burst phenomenon. Assays for cell loss and cell cycle arrest correlated broadly with target specific assays used in this analysis. We also observed that assays for endocrine disruption show unique patterns in the bioactivity space as evident by broad correlations with phenotypic readouts. The utility of these assays to predict organ-

specific toxicities is limited, as will be discussed in Chapter 4 regarding commercial *in vitro* setups for hepatotoxicity. It is argued that the manifestation of endocrine disruption into adverse effects is influenced by a number of genetic and environmental factors.[187] The endocrine disruption-cytotoxicity cluster, seen here, will also be observed among assays predictive of hepatotoxicity (Chapter 4) and acute toxicity (Chapter 5). In the following chapters, we demonstrate that *in vitro* activities, including endocrine disruption, can translate better into *in vivo* effects if specific chemical and bioactivity conditions are considered. We suggest that initiators of toxicity at the molecular level occur as a result of multiple chemical conditions as well as compounds' polypharmacology rather than single initiating events.

4 UNDERSTANDING CONDITIONAL ASSOCIATIONS BETWEEN TOXCAST *IN VITRO* READOUTS AND HEPATOTOXICITY USING RULE-BASED METHODS

4.1 Introduction

Hepatotoxicity is a complex adverse effect which is associated with diverse biological perturbations, at molecular and cellular levels. Several in *in vitro* endpoints were proposed to capture the key perturbation points.[109–112] It has been reported that using a subset of the relevant bioactivity endpoints would result in relatively low detection rates, with sensitivity ranging from 40 to 60%.[95,113–115]

In general, for an efficient translatability of *in vitro* outcomes in to *in vivo* effects, two key factors are to be considered. First, to design and select the assay set up and endpoints relevant to the toxicological effect in question.[15,100] Second, to understand and incorporate relevant conditions such as multiple bioactivity stressors,[193] as well as exposure and bioavailability.[86,194114,116] With regards to which endpoints to consider, and how to combine their readouts, is by no means clear. Parameters such as maximum plasma concentrations (C_{max}) and administered dose levels are commonly used to improve translatability of *in vitro* activity in to *in vivo* hepatotoxicity.[109,114,115] Yet, obtaining these measurements from *in vivo* experiments is generally not feasible at early stages of drug development, or for profiling large number of compounds.

A possible alternative for exposure measures are physicochemical properties, which are associated with pharmacokinetic parameters[117,118], and can be used as proxies to exposure.[195,196] Yet, the mechanistic understanding of how these properties influence the concordance between *in vitro* measurements and *in vivo* effects, to our knowledge, has not been investigated. Therefore, understanding and prioritizing biological effects of hepatotoxic

compounds with respect to chemical properties should involve a holistic approach which combine background knowledge as well as data driven methods.

The current study is, to the best of our knowledge, the first to apply rule-based models on ToxCast *in vitro* measurements and ToxRefDB *in vivo* hepatotoxicity data with the incorporation of physicochemical properties. Rule based classifiers describe each class label by a combination of conditions using the input property set. As such, rules can facilitate prioritizing significant assays, as well as the interpretation and analysis of multivariate associations between *in vitro* activity and *in vivo* toxicity. However, conventional rule models do not consider directional associations which may occur between input features and outcome class, leading to associations that are either spurious or difficult to interpret. Therefore, we modify rules according to two key assumptions: 1.) *Positive bioactivity* in an assay (and not absence of an activity) potentially contributes to hepatotoxicity; and 2.) *Multiple conditions* influence *in vitro–in vivo* associations, which means for an assay to extrapolate well into *in vivo* outcome, number of other conditions have to be met. These conditions can be a combination of bioactivities and/or physicochemical properties (related to exposure). With those two key assumptions in mind, we manually modify the rules for hepatotoxicity to enhance interpretability and biological relevance. The framework we describe here can also generally be used to optimize *in vitro* models for toxicity by selecting significant assay combinations, as well as identifying relevant physicochemical conditions.

4.2 Materials and Methods

The steps followed to generate and prioritize rules for hepatotoxic compounds are summarized in Figure (4-1).
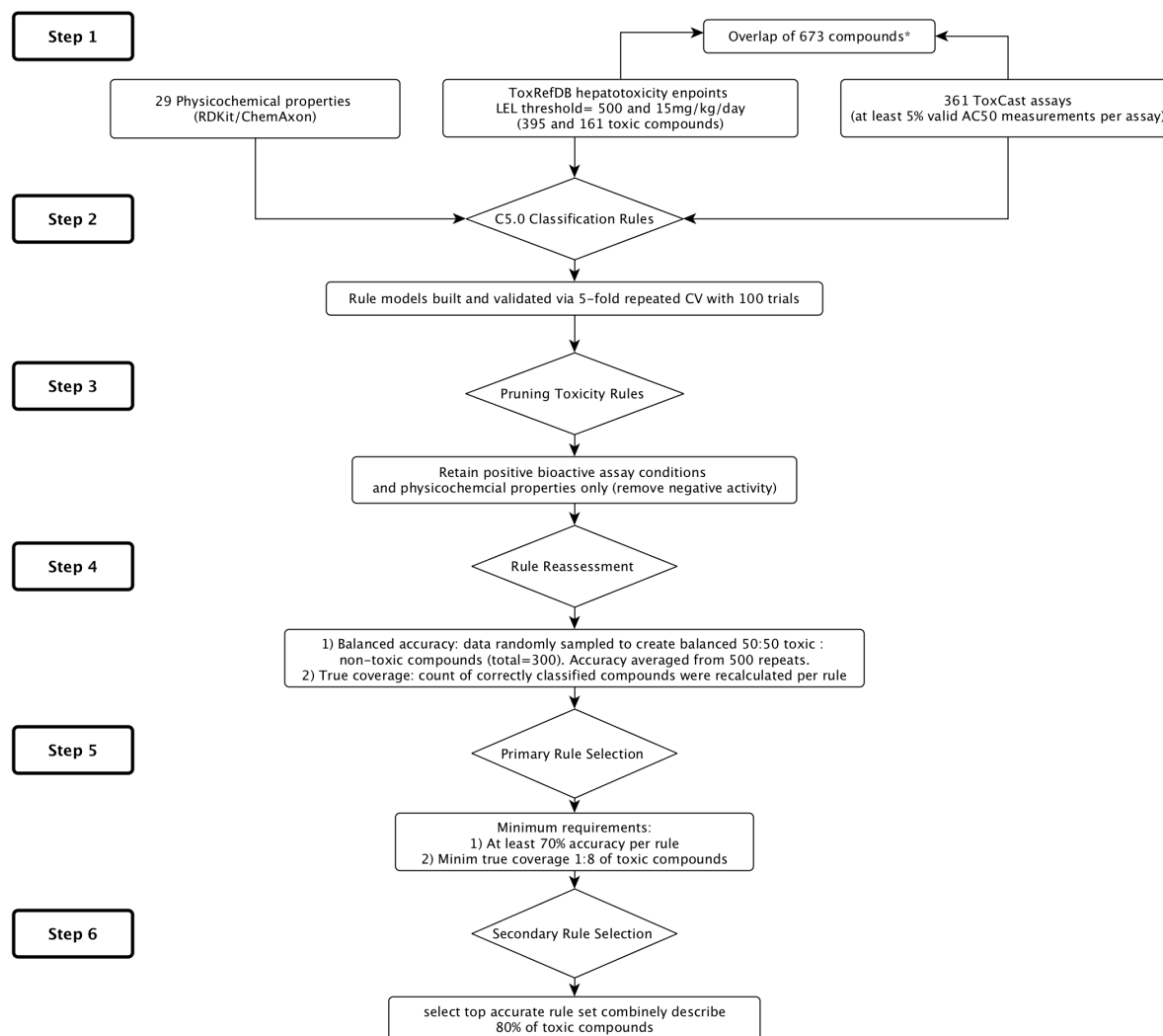
Figure (4-1) Dataset and workflow for extracting biologically relevant rules for hepatotoxic compounds. Firstly, ToxCast bioactivity measurements and an interpretable set of physicochemical properties were used as descriptors against hepatotoxicity labels. Rat hepatotoxicity endpoints, from the ToxRefDB, were converted into binary labels by setting two maximum exposure thresholds for lowest effective level (LEL), 500mg/kg/day and 15mg/kg/day (step 1). The dataset was used to generate rule-based classifiers *via* the C5.0 algorithm (step 2). At each threshold level, rules capturing toxicity were pruned by removing inactive assay statements (step 3). Next, in step 4 the modified rules were reassessed in terms of balanced accuracy and number of correctly classified compounds (true coverage). Prioritization of biologically relevant rules was conducted in multiple steps. The primary selection in step 5 involved performance measures, keeping rules that exerted at least 70% accuracy and median value of coverage (50 at 500mg/kg/day and 20 at 15mg/kg/day). The secondary selection (step 6) was performed based on overall compound coverage, by prioritizing the combination of most accurate rules that describe 80% of toxic compound set. The final set of prioritized rules were analyzed in terms of contributing bioactivity and physicochemical conditions.

4.2.1 Data collection
   (Figure (4-1), step 1)

Hepatotoxicity endpoints. Rodent hepatotoxicity measures were extracted from the Toxicity Reference Database (ToxRefDB, version toxrefdb_v1 released on October 2014)[78] . Histopathological endpoints (Appendix B, Table (B-1)) from rat studies in liver which were observed for chronic, sub chronic, multigenerational and prenatal development were used in this analysis, recorded as lowest effective level (LEL) in mg/kg/day for 882 compounds. These measurements were converted into a binary format by applying two toxicity thresholds of 500mg/kg/day and 15mg/kg/day which were subsequently analyzed separately. This classification was adapted from Martin *et al*[197] considering the highest and lowest quantile bins for toxic effects, corresponding to 15mg/kg/day or less, and 500mg/kg/day or less.

In vitro **measurements.** Assay bioactivity data were extracted from the ToxCast database[173], version December 2014, for 1,057 compounds which represents phase I and II. *In vitro* measurements in ToxCast are recorded as the concentrations at which half-maximum activity is reached (AC_{50}), generated from dose-response curves for more than 800 assays in units of $\log_{10}$ μM concentration. Assays are annotated by their "intended target type", in the ToxCast assays summary file[173] into protein, cellular, pathway, DNA, RNA or unspecified. We used this annotation to describe the assays as target-based or phenotypic; if the intended target type was described as protein, it was considered target-based, otherwise, for high-dimensional readouts, assays are annotated as phenotypic.

Compound set. The compound set used in this study represent the intersection between compounds in ToxRefDB and ToxCast, resulting in 673 compounds in total matched by ToxCast compound ID. Applying the toxicity threshold of 500mg/kg/day resulted in 395 toxic compounds and 278 non-toxic compounds; whereas, at the threshold of 15mg/kg/day 162 compounds were annotated as toxic and 511 as non-toxic (Step 1 in Figure (4-1)).

Dataset curation. Assay endpoints with empty fields (missing values) were considered inactive and an arbitrary (very large) AC_{50} value of 10^6 μM was assigned, adapted from Lui *et al*.[101] In order to select a data matrix that was as complete as possible (to reduce the bias of subsequent analyses) only assays which had valid AC_{50} measurements for at least 5% of the

compounds were selected. This step resulted in 361 assays which are listed in Appendix B, Table (B-2).

Structural preprocessing and calculation of physicochemical properties. Compounds were standardized using ChemAxon standardizer[198] (version 15.12.14.0) using the parameters cleaning 2D, mesomerisation, neutralization, tautomerization (generating the most stable tautomer) and removal of fragments. Physicochemical properties were generated using RDKit[199] via KNIME[200] and the Calculator Plugins[201] in Instant JChem (version 15.12.14.0)[202]. A subset of 29 physicochemical properties were used in the current analysis which are listed in Appendix B, Table (B-3).

4.2.2 Constructing rule-based classifiers (Figure (4-1), step 2)

Multivariate associations between assays and hepatotoxicity endpoints were modelled via rule-based machine learning classifiers (Step 2 in Figure (4-1)) as follows. Input variables were 29 physicochemical properties and 361 ToxCast *in vitro* AC_{50} measurements for 673 compounds against two hepatotoxicity labels of 500mg/kg/day and 15mg/kg/day. The rules were generated using the C5.0 algorithm (modified from the C4.5[153] algorithms by Ross Quinlan) as implemented in the C5.0[152] and caret[174] R packages using 5-fold repeated cross-validation with 100 trials and without winnowing. Other parameters are set to default. The models with the highest correct classification rate (CCR) at each trial were retained so as to be used to generate the rules (accuracy distribution in Appendix B, Figure (B-1)). Throughout text, we refer to rules associated with hepatotoxicity as rules predictive or describing toxicity, exchangeably.

4.2.3 Rule modification
(Figure (4-1), step 3)

Each rule derived from the above procedure consists of one or more conditional statements to predict the hepatotoxicity label at a given dose, based on the input variables (ie physicochemical properties and ToxCast readouts; Step 3 in Figure (4-1)). In order to extract *biologically relevant* associations between activity *in vitro* and hepatotoxicity *in vivo*, rules were manually modified to retain interpretable and biologically meaningful patterns. To achieve this, modification was applied by removing conditions of inactivity in assays from

rules describing hepatotoxic compounds. On the other hand, positive activities in assays and physicochemical properties were kept for further analysis as provided by the machine learning method. As conventional rule methods that take continuous variables as input can generate conditional splits contradicting the assumption stated above, manual curation of rules was applied to retain only splits satisfying this assumption.

Positive activity and inactivity in an assay were discriminated according to the direction of the conditional split. Since the compound potency in an assay is inversely related to the concentration at which a certain biological response was obtained (such as an IC50 value or similar), a bioactivity condition was considered active if the split represent a concentration range below a defined cutoff, and *vice versa*.

One example of how rules were modified is provided in Figure (4-2) and illustrated in the following. For example, in the condition Tox21_p53_BLA_p5_viability <= 0.026, the first part
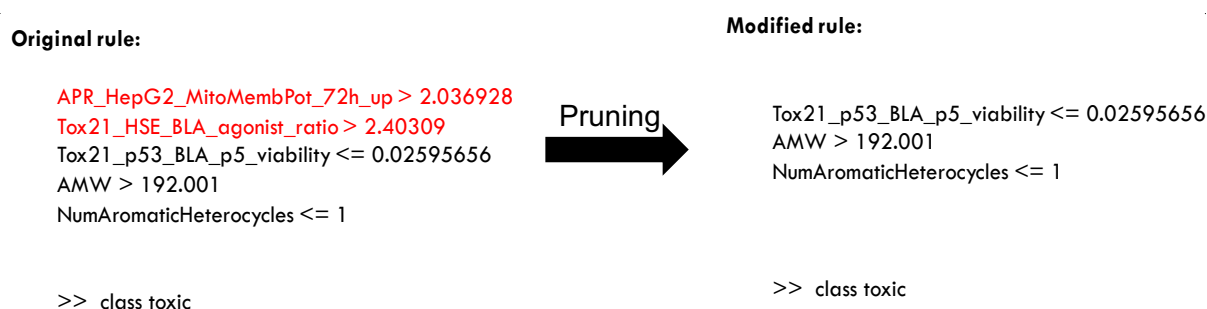
Original rule:

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

Pruning ⟹

Modified rule:

Tox21_p53_BLA_p5_viability <= 0.02595656
AMW > 192.001
NumAromaticHeterocycles <= 1

\>> class toxic

\>> class toxic

Figure (4-2) An example of rule pruning to retain biologically relevant conditions. Left: unmodified rule, describing toxic compounds, as generated by the C5.0 algorithm. This rule consists of five conditions, namely two inactive bioactivities (shown in red), one positive bioactivity and two physicochemical properties. These represent the increase in mitochondrial membrane potential (inactive), heat shock protein agonist (inactive), cytotoxicity assay (active), average molecular weight, and number of aromatic heterocycles (physicochemical properties). The rule was then modified by removing the inactive bioactivities so as to retain only positive bioactivity readout (cytotoxicity) and physicochemical properties (right), which are more meaningful, and biologically more plausible to be associated with toxicity.

describes the viability of a human intestinal cell line, and the second part (<= 0.026) represents the range of AC50 (in $\log_{10}$ μM concentration) to be less than 0.026 (< ~1.06μM) for the condition to be fulfilled. As the statement describes the range below a bioactivity cutoff (split point), this example represents a bioactivity condition with positive activity. On the other hand, the two conditions, i) APR_HepG2_MitoMembPot_72_up > 2.037 and ii) Tox21_HSE_BLA_agonist_ratio > 2, represent respectively the increase of mitochondrial membrane potential and Heat Shock Protein (HSP) agonism (both relative to negative control)

with AC50 values higher than 100μM. The latter conditions were considered inactive, and hence removed from rules capturing toxicity during the pruning process.

4.2.4 Performance assessment and rule prioritization
(Figure (4-1), steps 4 and 5)

In the current study, the performance of rules was assessed at two levels, for individual conditions used in original rule set as well as for rules before and after modification.

Individual rule conditions were first extracted by collecting the unique set of features used by the rules. These were assessed for their association with the hepatotoxicity outcome, at both thresholds, 15mg/kg/day and 500mg/kg/day, by calculating the information gain (entropy), which is described in Chapter 2 under section 2.2.2. The comparison was conducted with respect to condition type, which were categorized into three groups; conditions describing positive assay activity, inactivity in an assay and physicochemical properties.

As the applied modification may change rule performance, the next step was to assess the accuracy (confidence) and coverage of rules before and after modification (Step 4 in Figure (4-1)). Rule confidence, represents the percentage of correctly classified compounds for a given class, *i.e.* here hepatotoxicity. In order to account for the imbalanced distribution of toxicity classes, the balanced accuracy of modified rules was calculated by generating 500 randomly selected balanced data subsets, each composed of 300 data points, and then averaging the accuracy. Rule coverage was calculated for the number of toxic compounds that satisfy the rule conditions (true positives).

As the modification arrived at a rule set that is overall not optimal, a set of prioritization and selection steps were conducted. The prioritized rules were subsequently assumed to capture biologically meaningful information to the maximum possible extent, given the unavoidable limitations of chemical space coverage and bias of the dataset that was available to us.

The rules predictive of toxicity were filtered based on minimum coverage and accuracy (Step 5 in Figure (4-1)). Minimum coverage was set to 50 and 20 compounds per rule at thresholds of 500mg/kg/day and 15mg/kg/day, respectively. The cutoffs for coverage represented the median values after modification, in other words, the best 50% of rules in terms of coverage were selected (see Appendix B, Figure (B-2)). Secondly, an accuracy cutoff of 70% was applied for both rule sets. Another selection step was undertaken to reduce assay redundancy, that is if an assay contributes to multiple rules, then the rule with the highest accuracy that

contains that particular assay was selected. Finally, the minimal rule set which covers 80% of toxic compounds at each toxicity threshold was selected (see Results and Discussion below).

To identify the key bioactivity groups captured in toxic compounds in the dataset, the final rule sets were clustered according to their similarity in compound membership. This was performed by generating a matrix of rules against compounds fulfilling the respective rule conditions, from which a rule similarity matrix (based on shared detected and not detected toxic compounds) was calculated based on the Jaccard index.[203] Hierarchical clustering was applied to the similarity matrix using the Agglomeration method Ward.D2[204] algorithm *via* the 'hclust' function in R (version 3.3.2).[176] Visualizations were generated using 'ggplot2'[205] package in R. The bioactivity assay conditions under each cluster were examined.

4.3 Results and Discussion

4.3.1 Extracting biologically relevant rules for toxic compounds

4.3.1.1 Information content of conditions in hepatotoxicity rule-based classifiers

To have a picture of how original rules were structured, Table (4-1) describes the average frequency of each condition type. Overall, on average, rules capturing toxicity include one positive assay activity and four inactive bioassay conditions at both toxicity levels. Close to one (0.9) physicochemical condition is on average involved in toxic rules at a threshold of 15mg/kg/day, whereas only on average 0.6 physicochemical conditions are included in rules at a threshold of 500mg/kg/day. Hence, although inactive conditions are by themselves not very information-rich, they frequently are contained in automatically derived rules. Given that inactive assay conditions cannot be mechanistically meaningfully linked to toxic events this underlines the need for rule modification as our method of choice (the difficulty of deriving toxicity predictors entirely automatically from ToxCast data have also been discussed, using different methods, before).[99,100]

Table (4-1) Average number of conditions per toxic rule in the original set. 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.

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

In order to evaluate to what extent individual assays might be able to predict hepatotoxicity we firstly examined how much information was gained from each type of conditions in the rules. Given the distribution of data points in the classes of the overall dataset, the maximum possible information gain was 0.79 and 0.98, at thresholds of 15mg/kg/day and 500mg/kg/day, respectively. Figure (4-3) displays the distribution of the information gain (IG), for positive bioactivities, negative bioactivities, and physicochemical conditions in toxic rules at toxicity levels of 15mg/kg/day and 500mg/kg/day, respectively, and we can observe two key trends. Firstly, the maximum observed IG obtained by any split is very low overall, with the maximum IG being only slightly higher than 0.04 and the median IG of positive bioactivity conditions

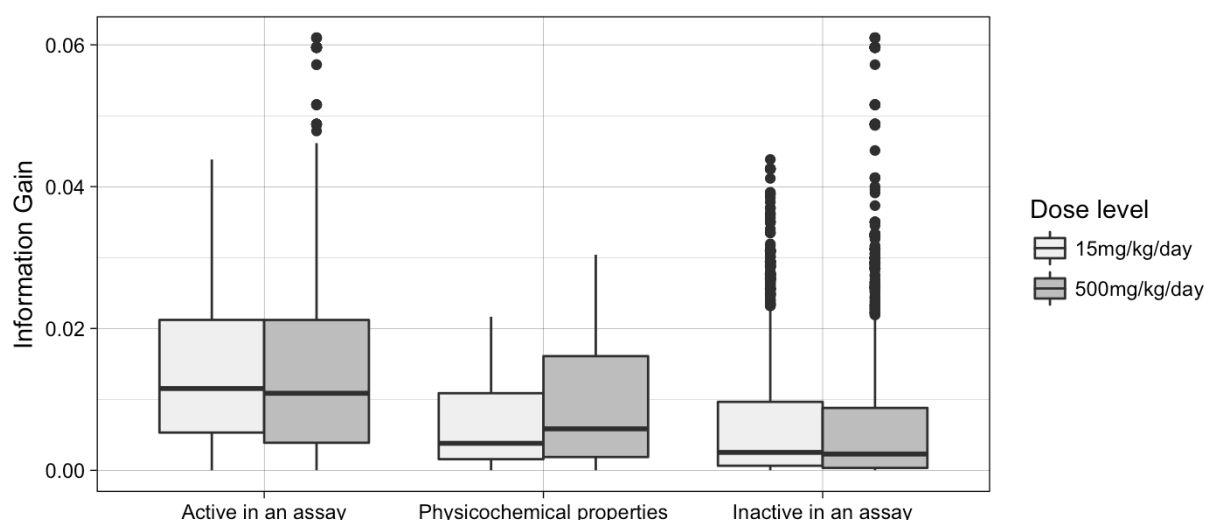being around 0.01. This indicates that single assays on their own have little predictivity for hepatotoxicity.

Figure (4-3) The distributions of the information gain (IG) for each condition type (positive bioactivity, assay inactivity and physicochemical properties), in rules capturing toxicity, at two toxicity thresholds, LEL= 500mg/kg/day and 15 mg/kg/day. IG is a metric used to describe to which extent particular condition (split) improves the homogeneity (purity) of the partitioned data, and which features are hence associated with the classes under consideration. To fully discriminate between all toxic from non-toxic compounds, IG of variable split should be 0.79 and 0.98, for levels of 15mg/kg/day and 500mg/kg/day, respectively. The overall values of IG in the plot are low compared to the maximum possible values. Positive bioactivity conditions had the greatest average IG in comparison to other condition types, whereas, negative bioactivity conditions had the lowest. This means that the predictive power of a single positive activity in classifying toxic compounds is generally larger than single bioassay inactivity and physicochemical property conditions. Yet, single assays on their own are not sufficient to fully predict compounds in the dataset, which aligns with findings in previous studies.[96,97,99]

Secondly, the median IG for physicochemical and negative bioactivity conditions is even lower and does not exceed 0.005. This means that these conditions *on their own* are less predictive for hepatotoxicity than active assay conditions. Examples of assays that scored the highest IG values include the downregulation of CD40 and IP-10 cytokines at the level of 15mg/kg/day and the upregulation of stress kinase and AMP response element binding protein (CREB) at 500mg/kg/day toxicity level (see "prioritizing endpoints for hepatotoxicity detection" below for details).

Although positive bioactivities provide the relatively highest IG overall, their low quantitative values indicate that one condition (or assay) is certainly not sufficient to discriminate all toxic from non-toxic compounds. This observation can be attributed to two factors: Firstly, hepatotoxicity involves diverse and complex mechanisms that cannot usually be captured by single endpoints.[107,193] Secondly, without considering exposure (or at least some

proxies, such as chemical properties),[117] assay readout do not translate readily into *in vivo* outcomes,[99,206] Our conclusion in this work is hence that, in order to improve our ability to predict *in vivo* toxicity, we need to use rules which, on the one hand, involve a combination of assay endpoints to cover wider bioactivity space, and on the second hand incorporate also physicochemical conditions.

4.3.1.2 Rule modification

To make rules capturing toxicity more biologically meaningful, we next modified them by removing inactive assay conditions, and then keeping rules which retain the highest accuracy and coverage (as described in Methods).

The changes in error rates (accuracy) from rule modification are presented in Figure (4-4). modification resulted overall in a deterioration in accuracy of 10% and 20% on average for rules set at thresholds of 15mg/kg/day and 500mg/kg/day, respectively, as shown in Figure (4-4). The change in accuracy is broad, ranging between -20% up to +40% at both toxicity levels. However, the improvement in accuracy (negative values in Figure (4-4)) can be seen for one fifth of rules at the level of 15mg/kg/day and less than a quarter of the rules at level of 500mg/kg/day (Figure (4-4)).
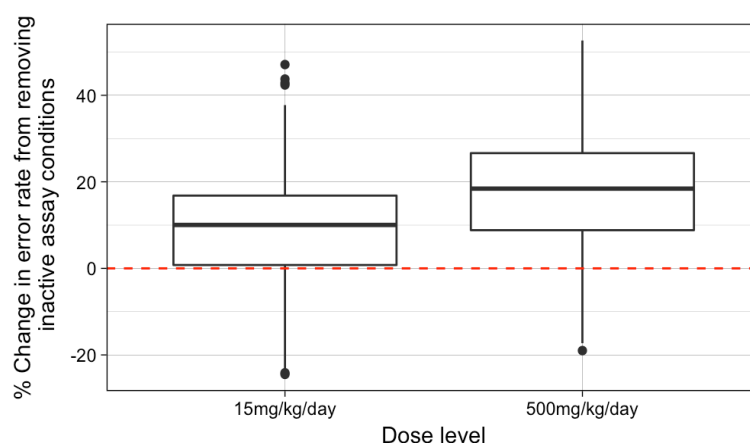
Figure (4-4) The error rate change (accuracy deterioration) in percentage from removing inactive bioactivity conditions from rules. The error rate is calculated by subtracting the accuracy of the rule after modification (by removing inactive bioactivity conditions) from the accuracy of the original rule. The deterioration in rule accuracy after modification is, in general, greater at threshold 500mg/kg/day, at an average of 20%, whereas, the overall drop in accuracy at 15mg/kg/day after modification is 10%. Simplifying the rules by pruning inactive conditions resulted in a variable level of accuracy deterioration. Yet, negative values mean improvement in accuracy which was observed in almost 25% of rules.

4.3.1.3 Prioritizing predictive rule sets (Figure (4-1), step 6)

As some modified rules have shown severe deterioration in accuracy, a selection step was introduced to keep only highly performing modified rules, defined as at least 70% minimum accuracy, in addition to a minimum of 50 and 20 compounds being covered (corresponding to the median in coverage), at the toxicity thresholds 500mg/kg/day and 15mg/kg/day, respectively. Finally, rules were ranked by accuracy, then the minimum rule set able to detect at least 80% of all hepatotoxic compounds was selected (Step 6 in Figure (4-1)).

Figure (4-5) shows the relationship between the overall compound coverage, at both toxicity thresholds, with i) the minimum rule accuracy in the series of ranked rules and ii) the number of unique bioassays used in the rules. Firstly, if higher percentages of toxic compounds are to be detected, then more rules (some of which will have lower accuracy) are needed. For example, to obtain collectively 80% compound coverage at a threshold of 500mg/kg/day requires including 35 rules up to a lower limit of 73% accuracy. On the other hand, in order to achieve equivalent coverage at a threshold of 15mg/kg/day 20 rules with a lower accuracy limit of 81% is sufficient. Hence, potent toxicants can be captured by rules of higher confidence than compounds fall under weaker toxicity levels.
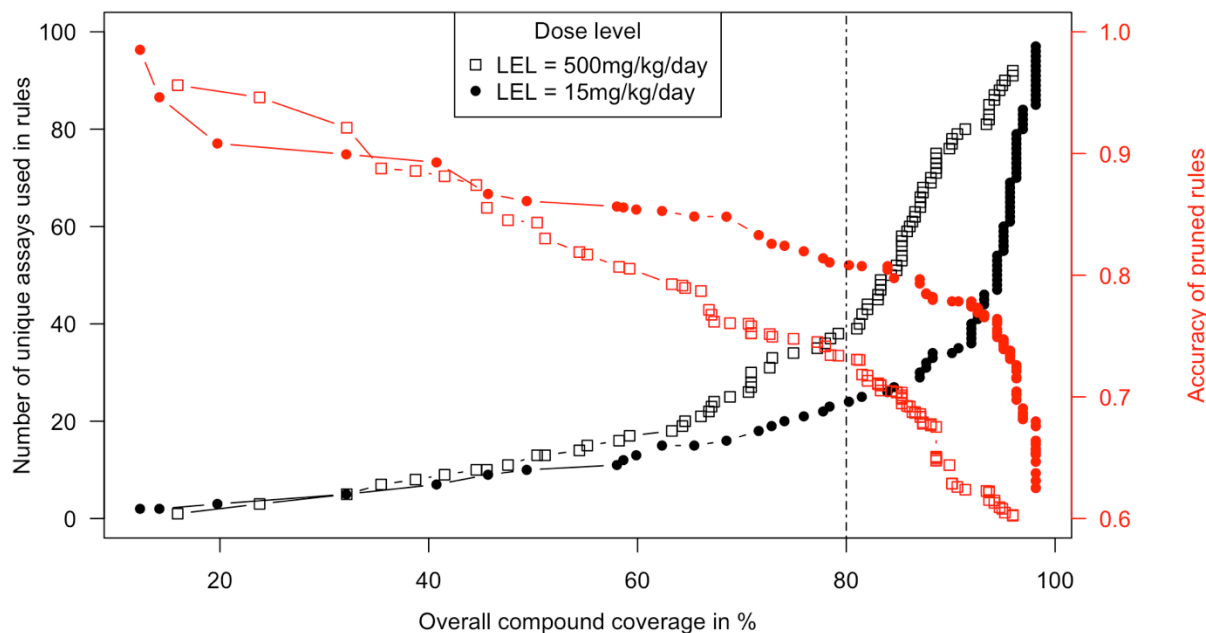
Figure (4-5) Percentage of overall toxic compounds matching rules ranked by accuracy. The x-axis represents the percentage of compound coverage as a function of minimum accuracy in ranked rules (Y-axis, red) and number of unique bioactivity assays used in rule combination (Y-axis, black). The most accurate rule sets sufficient to cover 80% of toxic compounds, at each threshold level, were selected as the minimal rule set to describe hepatotoxicity. Fewer numbers of unique assays were required at a threshold of 15mg/kg/day in comparison to level of 500mg/kg/day to cover 80% of all compounds, namely 24 and 38 assays, respectively.

Secondly, in order to cover larger proportions of compounds, higher numbers of unique assays to be used by the rules (Figure (4-5)). It can be seen that up to 50% of compounds, at both thresholds, can be described in rules using readouts from 11 assays. For 80% compound coverage, 24 and 39 assays are needed in rules, at thresholds 15mg/kg/day and 500mg/kg/day, respectively. Further for 90% coverage, 34 unique assays are needed in rules at 15mg/kg/day, whereas, more than 70 unique assays are required to cover the same proportion of toxicants at 500mg/kg/day threshold.

Hence overall, single assay endpoints, for the dataset employed here, are only sufficient to anticipate in vivo hepatotoxicity for rather few compounds individually. Instead, a combination of bioactivity measurements is required for enhancing the detection rate of hepatotoxic compounds, for which larger numbers of assays (with broader mechanistic coverage) are needed.

4.3.2   Prioritizing endpoints for hepatotoxicity detection
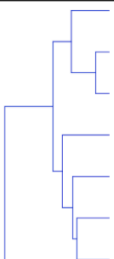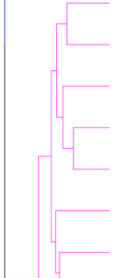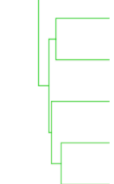
We next interpreted the bioactivity endpoints used in rules predicting hepatotoxicity, as determined from the assays used by the best-performing rules covering 80% of the toxic compounds at both LELs of 15mg/kg/day and 500mg/kg/day.

The diversity of the bioactivity space in the prioritized rules was analyzed by firstly clustering rules according to compound coverage, and subsequently interpreting the biological associations between the selected endpoints and hepatotoxicity. Rules were clustered according to the similarity in toxic compound coverage, which means that rules that satisfy similar compound sets are grouped together. Tables (4-2) and (4-3) present the rule clusters labelled by the major contributing bioactivity described by the assay endpoints used in the rules. At both thresholds of 15mg/kg/day and 500mg/kg/day, the rules clustered into three bioactivity groups, namely those involving i) Cytochrome P enzymes, ii) immune responses, and iii) nuclear receptors and transcription factor elements.

Table (4-2) Description of unique bioactivity assays in best performing rules for toxicity, which cover, combined, 80% of toxic compounds, at threshold of 500mg/kg/day. Rules were clustered according to similarity in compound coverage, which resulted in three major clusters of rules. Each is describing predominantly one class of bioactivity assay, namely, Cytochrome P activity, immunological responses and nuclear receptor activity. The information gain was calculated for individual conditions (split points) and highlighted for values higher than median (0.02), whereas, accuracy represents the rule in which the bioactivity was used. Some rules used two assay conditions at a time, and assays in the table. are linked by the symbol; ⌞.

Bioactivity class	Index	Associated assay	Information gain (split)	Accuracy (rule)	Gene symbol	Function
	A.1	APR_HepG2_MitoMass_24h_up	0.019	0.921	-	Cell morphology
	A.2	ATG_PPARg_TRANS_up	**0.045**	0.874	PPARG	Nuclear receptor
	A.3	⌞ OT_AR_ARSRC1_0480	**0.033**	0.874	AR	Nuclear receptor
	A.4	NVS_ADME_hCYP2C18	**0.028**	0.817	CYP2C18	Cyp
	A.5	NVS_ADME_hCYP2C19	**0.025**	0.752	CYP2C19	Cyp
Activity against Cytochrome P	A.6	⌞ NVS_TR_hDAT	**0.021**	0.752	SLC6A3	Transporter
	A.7	NVS_ADME_rCYP3A1	**0.035**	0.881	Cyp3a23/3a1	Cyp
	A.8	NVS_ADME_rCYP3A2	**0.021**	0.767	Cyp3a2	Cyp
	A.9	NVS_MP_hPBR	0.011	0.734	TSPO	Transporter
	A.10	NVS_NR_hCAR_Antagonist	0.017	0.760	NR1I3	Nuclear receptor
	A.11	OT_FXR_FXRSRC1_0480	0.014	0.748	NR1H4	Nuclear receptor
	A.12	APR_HepG2_CellCycleArrest_72h_dn	**0.022**	0.749	-	Cell cycle
	A.13	⌞ Tox21_FXR_BLA_antagonist_ratio	0.013	0.749	NR1H4	Nuclear receptor
	A.14	BSK_BE3C_uPA_down	0.016	0.791	PLAU	Protease
	A.15	BSK_KF3CT_IP10_down	**0.029**	0.745	CXCL10	Cytokine
Immunological activity	A.16	BSK_KF3CT_MMP9_down	**0.022**	0.762	MMP9	Protease
	A.17	BSK_LPS_CD40_down	0.017	0.752	CD40	Cytokine
	A.18	⌞ BSK_3C_IL8_down	0.014	0.752	CXCL8	Cytokine
	A.19	BSK_LPS_MCP1_down	0.019	0.805	CCL2	Cytokine
	A.20	BSK_SAg_CD40_down	**0.026**	0.772	CD40	Cytokine
	A.21	BSK_SAg_SRB_down	**0.030**	0.807	-	Cell cycle
	A.22	APR_HepG2_MitoMembPot_72h_up	**0.020**	0.819	-	Cell morphology
	A.23	APR_HepG2_MitoMembPot_1h_dn	**0.029**	0.888	-	Cell morphology
	A.24	⌞ Tox21_AR_BLA_Antagonist_ratio	**0.020**	0.888	AR	Nuclear receptor
	A.25	APR_HepG2_NuclearSize_24h_up	0.018	0.761	-	Cell morphology
	A.26	APR_HepG2_OxidativeStress_1h_up	**0.025**	0.789	-	Cell cycle
	A.27	APR_HepG2_StressKinase_1h_up	**0.052**	0.956	-	Cell cycle
Nuclear receptor activity/ phenotypic readouts	A.28	ATG_BRE_CIS_up	**0.021**	0.734	SMAD1	DNA binding
	A.29	ATG_C_EBP_CIS_up	**0.038**	0.843	CEBPB	DNA binding
	A.30	⌞ ATG_HIF1a_CIS_up	0.013	0.843	HIF1A	DNA binding
	A.31	ATG_CRE_CIS_up	**0.048**	0.731	CREB3	DNA binding
	A.32	ATG_FoxA2_CIS_up	0.018	0.741	FOXA2	DNA binding
	A.33	BSK_SAg_PBMCCytotoxicity_up	0.019	0.758	-	Cell cycle
	A.34	Tox21_ERa_LUC_BG1_Agonist	0.009	0.793	ESR1	Nuclear receptor
	A.35	Tox21_GR_BLA_Antagonist_ratio	0.009	0.787	NR3C1	Nuclear receptor
	A.36	Tox21_MitochondrialToxicity_viability	0.018	0.946	-	Cell cycle
	A.37	⌞ ATG_p53_CIS_up	0.014	0.946	TP53	DNA binding

Table (4-3) Description of unique bioactivity assays in best performing rules capturing toxicity, which cover, combined, 80% of toxic compounds, at threshold of 15mg/kg/day. Rules were clustered according to similarity in compound coverage, which resulted in three major clusters of rules. Similar to findings in Table (4-2), the predominant assay class used by each rule cluster are Cytochrome P activity, immunological responses and nuclear receptor activity. The information gain was calculated for individual conditions (split points) and highlighted for values higher than median (0.022), whereas, accuracy represents the rule which the bioactivity was used. Some rules used two assay conditions at a time, and these assays are linked in the table by the symbol; ∟.

Rule cluster/key mechanism	Index	Associated assay	Information gain (split)	Accuracy (rule)	Gene symbol	Function
Activity against Cytochrome P	B.1	BSK_LPS_PGE2_up	0.0215	0.861	PTGER2	GPCR
	B.2	NVS_ADME_hCYP2C19	**0.0287**	0.856	CYP2C19	Cyp
	B.3	NVS_ADME_rCYP2A1	**0.0278**	0.853	Cyp2a1	Cyp
	B.4	∟ OT_AR_ARSRC1_0480	**0.0290**	0.853	AR	Nuclear receptor
	B.5	NVS_ADME_rCYP2C12	**0.0294**	0.908	Cyp2c12	Cyp
	B.6	NVS_ADME_rCYP2C13	**0.0271**	0.854	Cyp2c13	Cyp
	B.7	NVS_ADME_rCYP2C6	**0.0294**	0.985	Cyp2c6v1	Cyp
	B.8	∟ ATG_VDRE_CIS_up	0.0203	0.985	VDR	Nuclear receptor
Immunological activity/Endocrine disruption	B.9	BSK_3C_ICAM1_down	0.0182	0.810	ICAM1	Cell adhesion molecules
	B.10	BSK_4H_MCP1_down	**0.0370**	0.814	CCL2	Cytokine
	B.11	BSK_BE3C_MIG_down	**0.0287**	0.867	CXCL9	Cytokine
	B.12	∟ Tox21_ERa_BLA_Agonist_ratio	**0.0236**	0.867	ESR1	Nuclear receptor
	B.13	BSK_hDFCGF_IP10_down	**0.0402**	0.899	CXCL10	Cytokine
	B.14	∟ Tox21_MitochondrialToxicity_viability	0.0208	0.899	-	Cell cycle
	B.15	BSK_SAg_CD40_down	0.0190	0.808	CD40	Cytokine
	B.16	Tox21_AR_BLA_Antagonist_viability	**0.0251**	0.893	-	Cell cycle
	B.17	∟ Tox21_PPARd_BLA_antagonist_ratio	0.0130	0.893	PPARD	Nuclear receptor
	B.18	Tox21_ERa_BLA_Antagonist_ratio	0.0191	0.856	ESR1	Nuclear receptor
Nuclear receptor activity	B.19	APR_HepG2_MitoMass_72h_up	0.0162	0.833	-	Cell morphology
	B.20	∟ APR_HepG2_NuclearSize_24h_up	0.0181	0.833	-	Cell morphology
	B.21	ATG_LXRb_TRANS_up	0.0141	0.824	NR1H2	Nuclear receptor
	B.22	NVS_TR_hDAT	**0.0243**	0.848	SLC6A3	Transporter
	B.23	OT_ER_ERbERb_0480	0.0098	0.826	ESR2	Nuclear receptor
	B.24	Tox21_FXR_BLA_agonist_ratio	0.0153	0.820	NR1H4	Nuclear receptor

4.3.2.1 Cytochrome P

Activity against Cytochrome P enzymes is one of the key properties of hepatotoxic compounds at both toxicity thresholds (Tables (4-2) and (4-3), cluster 1). There are multiple assays in the rules describing activity against different Cytochrome P isoforms, namely against CYP3A, CYP2C18 and CYP2C19 at a dose of 500mg/kg/day, and against CYP2A, CYP2C6, CYP2C12, CYP2C13 and CYP2C19 at a level of 15mg/kg/day. All Cytochrome P enzymes exerted information gain (IG) values higher than average in the selected assay set (Tables (4-2) and (4-3), cluster 1), *i.e.* higher than 0.02 and 0.022 at levels of 500mg/kg/day and 15mg/kg/day, respectively. The average potency in Cytochrome P enzymes as described by rules is less than 10μM (Appendix B, Figure (B-3)). Activity against Cytochrome P isoforms CYP3A4 and CYP2C19, which are responsible for the majority of drug metabolic reactions[207], can be linked to liver injury via generating toxic metabolites or interfering with the metabolism of co-administered drugs resulting in slow elimination and chemical accumulation.[80,101,207]

4.3.2.2 Immunological responses

Multiple assays contribute to immunological responses at both toxicity cutoffs (Tables (4-2) and (4-3), cluster 2). Immunological responses in hepatotoxicity rules, in general, had shown assay potency requirements lower than Cytochrome P activities, of around 40μM (Appendix B, Figure (B-3)). At a threshold of 500mg/kg/day, assays associated with the downregulation of cytokines CXCL10 and CD40 were frequently selected for rules, and they also had an individual information gain values above the median at values around 0.029, namely the assays "A.15", "A.17" and "A.20". These cytokines are associated with both proinflammatory and regenerative responses depending downstream signaling. [208,209] For example, CD40-mediated activation of IL-12 has a proinflammatory effect.[210,211] On the other hand, it can also activate IL-10 immune response, which primarily participates in regenerating and repairing hepatic cells via anti-inflammatory responses.[208,212] Similarly, the CXCL10 cytokine can reduce liver injury in mice models via upregulating CXCR2.[213] Yet, also blocking CXCL10 has a regenerative effect after liver damage.[209] Hence, hepatotoxicity can be predicted from the perturbing the expression of relevant genes, and can be irrespective to the direction of change. Similarly, immunological endpoints detected at a threshold of 15mg/kg/day involve the downregulation of CXCL10 and CD40, as above, and in addition that of CCL2, captured by the assays "B.13", "B.15" and "B.10", respectively. At this threshold, downregulation of CCL2 had a high associated information gain of 0.037, hinting to its importance for hepatotoxicity

prediction. CCL-2 expression can increase tissue damage via IL-12 signaling[214], but simultaneously has a protective effect by activating the hepatoprotective cytokine IL-10.[214,215] A study has shown an overall downregulation of a variety of cytokines (CCL2, IL-10, IL-12 and IL-16) in primary and liver sinusoidal endothelial (LSEC) cell lines upon the exposure to free fatty acids in contrast to hepatocytes. The authors suggested the anti-inflammatory effect as a response to overcome liver damage.[216] Therefore, assays which detect changes in the expression of immunological cytokines, such as CXCL10 and CCL2, are biologically not entirely understood, but nonetheless they are statistically informative and hence, can be regarded as valuable in hepatotoxicity *in vitro* models.

4.3.2.3 Nuclear receptors and response elements

Activity against nuclear receptors contributes considerably to the overall bioactivity profile of hepatotoxic compounds (cluster 3 in Tables (4-2) and (4-3)). Compared to the above bioactivity groups, lower potencies of approximately 50μM were used in the rules to predict hepatotoxicity (Appendix B, Figure (B-3)). Endocrine disruption is seen at both thresholds, such as activity against estrogen and androgen receptors, which is captured in multiple assays ("A.3", "A24", "A.34", "B.4", "B.12", "B18", "B23"). There are established links between estrogen and glucocorticoid receptors ("A.35") with cholestasis (impairment of bile flow)[217] and steatosis (fatty liver)[218], respectively. In addition, androgen receptor antagonism is associated with a range of hepatotoxic effects with hepatitis as most commonly reported .[219] Rules obtained at both thresholds for hepatotoxicity also share activity against CAR (constitutive androstane receptor), FXR (farnesoid X receptor) and PPAR (peroxisome proliferator-activated receptor) which are described by the assays "A10", "A.11", "B.24", "A.2" and "B.17". Both CAR and FXR play a key role in preventing xenobiotic induced hepatotoxicity by regulating a number of genes including phase I and II metabolizing enzymes and bile acid transporters.[220,221] The activity of FXR was found to be correlated with the degree of protection from chemical-induced liver injury.[221] PPAR-gamma agonist activity, which represented by up regulation in A.2, is known to be involved in chemical-induced liver injury. [222] Whereas, PPAR-delta has a hepatoprotective effects against toxicants[223].

Additionally, there were bioactivities contributing specifically to each toxicity cutoff under the nuclear receptor activity class. For example, at 500mg/kg/day, several transcription factor response elements were involved in cluster 3 in Tables (4-2), such as the regulation of cyclic-AMP response element binding protein (CREB), which provided a relatively high information

gain, at this dose, of 0.048 in "A.31". It has been found that the activation of CREB-binding protein/β-catenin interaction promotes liver fibrosis.[224] Another response element is the upregulation of SMAD ("A.28") , which mediates TGF-β-induced apoptosis and fibrosis via downstream immune responses.[225,226]

At a toxicity threshold of 15mg/kg/day, bioactivities under the nuclear receptors cluster included those against the vitamin D (VDR) and liver X receptors (LXR) in the assays "B.8" and "A.21". Upregulation of the VDR response element (VDRE) is associated with anti-inflammatory properties and xenobiotic metabolism.[227,228] LXR has an anti-inflammatory effect[227] and can reduce chemical-induced toxicity[229]. Thus, the activation of these nuclear receptors can be linked to triggering protective response against xenobiotics.

Overall we can conclude that hepatotoxic compounds are frequently associated with multiple changes in nuclear receptor activity, with different types of nuclear receptors being activated or inhibited at different dose levels, based on dataset and analysis type considered here.

4.3.2.4   Other targets and mechanisms

Furthermore, the prioritized rules involved a number of phenotypic and less specific activities that cannot be attributed to a specific target. These include cytotoxicity, cell cycle arrest, oxidative stress and mitochondrial impairment (Tables (4-2)). There are ten assays describing cell cycle or cell morphology under the toxicity level of 500mg/kg/day, whereas this is only the case for four assays detected by best rules at 15mg/kg/day (Tables (4-2) and (4-3)). Also, these assays had shown significant difference in the overall assay potencies, of around 50μM, between levels of 500mg/kg/day and 15/mg/kg/day. The phenotypic assay overlapping between the two toxicity levels include increase in nuclear size and mitochondrial mass, represented by "A.1", "A.25", "B.19" and "B.20". Mitochondrial dysfunction is one of the key mechanisms of chemical-induced hepatotoxicity.[107,212,230,231] Nuclear size increase is accompanied by nuclear receptor activity, as a result of activating gene expression. Additionally, at threshold of 500mg/kg/day, two assays describe mitochondrial effects with information gain higher than the median (0.02 and 0.029), which are associated to changes in mitochondrial membrane potential, namely "A.22" and "A.23". Chemical toxicants can cause mitochondrial permeability transition (MPT) via the opening of permeability transition pores in the mitochondrial membrane, either directly or indirectly. As a result, mitochondrial depolarization takes place which leads to ATP depletion and reactive oxygen species (ROS) release, followed by mitochondrial membrane rupture and apoptosis or necrosis.[107] Another

effect is the increase in stress kinase expression as a response to stress; "A.27". This assay has amongst the highest information gains of all assays (of 0.052) and the accuracy of its rule is 95%, which indicates that it is highly associated with toxicity at a 500mg/kg/day level. Also, rules at this level involve other cell cycle assays which screen for cytotoxicity, oxidative stress and cell cycle arrest. Hence, we can conclude that less potent hepatotoxic compounds are more likely to have unspecific effects which can be difficult to detect using target-specific assays. Judson *et al* reported a greater likelihood of disrupting target specific pathways when active concentrations fall within the range of eliciting cytotoxicity.[85] What we observe here is an association between potency of toxicity *in vivo* and specificity of *in vitro* effects that are predictive for toxicity. We also observe that these unspecific effects are embedded more frequently in clusters describing nuclear receptor activity and endocrine disruption. This observation aligns with findings in Chapter 3, where assays for endocrine disruption have shown broad correlations with cytotoxicity assays. Additionally, for compound to be toxic at low doses, they are required, according to the rules, to be more potent in assays (see Appendix B, Figure (B-3)).

In addition to the major bioactivity classes, further rules for hepatotoxicity involved other target assays. For example, the Novascreen target binding affinity assay ("A.6" and "B.22"), which appeared in rule sets at different dose levels, is related to the transporter SLC6A3 gene with relatively high information gain of 0.021 and 0.024 (see Figure (4-3) for distributions of information gain values) at 500mg/kg/day and 15mg/kg/day, respectively. The SLC6A3 gene encodes for the dopamine transporter (DAT). DAT-dependent neurological degeneration is linked with hepatic dysfunction related to ROS overproduction and mitochondrial impairment in rodents.[232] Another endpoint related to the toxicity threshold of 500mg/kg/day is the translocator protein (TSPO) (assay "A.9"), which is involved in the transport of cholesterol across mitochondrial membrane. The expression of TPSO has also been found to be associated with activating macrophages in chemical-induced liver injury and hence leads to cell death.[233,234] At a toxicity level 15mg/kg/day, upregulation of prostaglandin E2 (PGE2), in assay "B.1", was also used by rules. PGE2 is a lipid autocoid which protects against liver damage by downregulating the expression of inflammatory cytokines. [235,236] It has also been reported that PGE2 participates in liver regeneration upon injury. [237]

Therefore, we can overall conclude that hepatotoxicity can be initiated by a variety of mechanisms (also very likely beyond the ones covered in the assays used here), supporting the need for broad range of endpoints when screening for potential hepatotoxic compounds.

4.3.2.5   Combined bioactivity readouts

Our study goes beyond interpreting univariate associations into investigating rules constructed from multiple assay conditions. Previous studies has reported the benefit of combining multiple pathway perturbations to detect hepatotoxic compounds, but were narrowed to specific modes of action.[109,193] Some rules for hepatotoxicity involve two bioactivities at a time (linked by the symbol ⌐ in Tables (4-2) and (4-3)). There are seven and six rules including multiple bioactivity features at toxicity thresholds 500mg/kg/day and 15mg/kg/day, respectively. For example, multiple rules combined androgen receptor activity with a range of bioactivity assays, such as change in mitochondrial membrane potential, at threshold 500mg/kg/day, Cytochrome P (CYP2A1) activity at 15mg/kg/day, and PPAR activity at both thresholds. AR ligand, dihydrotestosterone, has shown to induce mitochondrial membrane potential[238], which establish a link between AR modulation and disrupting mitochondrial membrane. Additionally, CYP2A1, which metabolizes 90% of testosterone[239], is subject to inducers leading to a decrease in serum testosterone.[240] Hence, the dual activity of AR antagonist and CYP2A1 modulation can potentiate compounds to be hepatotoxic. There is an established bidirectional crosstalk between AR and PPAR isoforms by which each can influence the expression as well as the transcription activity of the other.[241,242] Primary hepatocytes of obese male AR-knockout mice had shown hepatic steatosis which is associated with altered PPAR-α and PPAR-γ expression.[243]

Other interactions which are seen at a level of 500mg/kg/day include multiple combinations Firstly, CYP2C19 appeared in conjunction with SLC6A3 (dopamine transporter) in rules (Table (4-2)). It is found that the antagonists of D2 dopaminergic receptors interfere with the regulation if CYP2C enzymes.[244,245] This can support the link between interfering dopamine transportation and CYP2C19 activity highlighted in the rules. Secondly is the combination of CD40 with IL-8 cytokine. Activation of CD40 increases the secretion of IL-8 in hepatic stellate cells resulting in an amplification of proinflammatory effects.[246] Also, the hypoxia inducible factor-1 (HIF-1) was combined with CCAAT/enhancer binding protein B (C/EBPB) (Table (4-2)). Studies have shown mutual regulation between these two transcription factors in expression and transcription.[247,248] Also, HIF-1 is one of the key transcription factors which binds to C/EBPB during liver regeneration.[249]

At 15mg/kg/day, multiple assay combinations predictive for hepatotoxicity can be seen in Table (4-3), including CYP2C6 with VDR and CXCL-9 with ER agonists. VDR involves in the metabolic liver damage and its expression correlates inversely with the severity of liver

steatosis[250] and fibrosis.[251] In response to xenobiotics, VDR directly induces the upregulation of CYP2C6.[252] Hence, it is plausible that compounds that combine activity against CYP2C6 and upregulation of VDR are more 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,[253] a cytokine associated with liver fibrosis.[254]

These observations support the importance of conditional associations in studying the translatability of *in vitro* activity into *in vivo* effects, and using rules we were able to suggest which assays are most predictive for hepatotoxicity when used in combination, based on the dataset used here.

4.3.2.6   The case study of troglitazone

As a case study for the benefit of considering bioactivity combinations in screening hepatotoxic compounds, we chose troglitazone, an antidiabetic drug that was withdrawn from the market in 2000 due to incidences of hepatotoxicity.[255] It was reported that the mechanism of troglitazone liver toxicity is initiated through the mitochondrial impairment, cellular stress[256] and triggering immunological responses.[257] It was also argued that risk factors including genetic and environmental factors, besides biological activity of the compound itself, play a contributing role.[258,259]

Troglitazone is labelled in our data as toxic at the LEL level of 500mg/mg/day but not at 15mg/kg/day as extracted from rat studies. This compound matched ten of the rules at the level of 500mg/kg/day and two rules at the level of 15mg/kg/day (see Appendix B, Table (B-4) and (B-5)). These are described by assays for mitochondrial toxicity, endocrine disruption and activity of immunological responses. At level of 500mg/kg/day, troglitazone fulfilled the rules describing the combinations of AR with PPAR-γ as well as IL-8 with CD40 (see "Combined bioactivity readouts"). Additionally, this compound also matched multiple rules in which cytotoxicity was combined with specific target activities, for both toxicity levels (see Appendix B, Table (B-4) and (B-5)). The average hepatotoxic compound, however, complied with only four rules at 500mg/mg/day and two rules at 15mg/kg/day (Appendix B, Figure (B-4)). Also, there is significant difference in number of satisfied rules by toxic and non-toxic compounds at both thresholds (Appendix B, Figure (B-4)). Given the number of rules satisfied by troglitazone, it had more bioactive liabilities compared to the average toxic compound at the level 500mg/kg/day and equivalent to the average liability at the level of 15mg/kg/day. Therefore, troglitazone's promiscuity in hepatotoxic rule space predict it to be likely hepatotoxic *in vivo*.

4.3.3 Comparison of prioritized bioactivities with commercial hepatotoxic assay endpoints

Further, we compared the bioactivities covered by prioritized rules with the *in vitro* endpoints conducted by commercial models for hepatotoxicity. Bale *et al*., reviewed a series of commercially available *in vitro* platforms for the detection of hepatotoxic compounds[260] which we will use here. From this review, we selected four *in vitro* models which screen against multiple readouts and endpoint classes, namely Cellciphr[®261], 3D Insight™[262] , Hepatopac[®263] and RegeneMed[112], which are summarized in Table (4-4).

There is an overall large overlap between bioassay endpoints (see Table (4-4)) screened within commercial models and assays prioritized by the rules generated in this study. Firstly, all models, as well as the rules derived in the current study, involve screening against mitochondrial impairment and cell stress, which are known as key signals for hepatotoxicity.[107] Secondly, phenotypic readouts associated with cell growth or morphology are used by CellCiphr[®], InSphero 3D Insight™ as well as in the rules, examples of which are apoptosis, cell loss and changes in nuclear size and mitochondrial mass. Additionally, screening for Cytochrome P activity is an endpoint used in InSphero 3D Insight™, Hepatopac[®], RegeneMed. 3D Insight™ and RegeneMed also screen for changes in cytokine levels, which is in agreement with the assays identified as important in the current study. Another relevant endpoint is the inhibition of the bile acid transporter which is screened in 3D Insight™and Hepatopac[®]. Although the original ToxCast assay set we used in this analysis did not include the inhibition of bile acid transporters, the rules detected activity against FXR and CAR, which directly regulate the expression of these transporters.[220] Moreover, inhibition of a set of proteins, such as albumin, urea, and fibrinogen is conducted by InSphero 3D Insight™, Hepatopac[®] and RegeneMed. The counterpart assay used in the prioritized rules screen for the decrease in total protein level in the cell ("A.21").

In addition to the above endpoints, our analysis also identified other assay readouts with association with hepatotoxicity, in particular nuclear receptor activity (Tables (4-2) and (4-3)), that at the current stage are less covered in commercial hepatotoxicity assays. The involvement of nuclear receptors in hepatotoxicity is supported by mechanistic studies - for example, Liu *et al*. reported that estrogen and androgen antagonism are related to proliferative lesions.[101] Additionally, Hu *et al*., demonstrated a significant univariate association of two ToxCast assays for androgen receptor activity with human hepatotoxicity.[80] Hence, although it is known that

estrogen and androgen disruption can cause hepatotoxicity[217,219], this endpoint is often not currently used in commercial assay setups (Table (4-4)). Therefore, while generally overlap between the features identified in our work and commercial assays for hepatotoxicity exists, it would be suggested that including endocrine activity among *in vitro* models can improve the coverage, and hence detection, of hepatotoxic compounds beyond the current state of the art.

Table (4-4) Comparison of bioactivities used in prioritized rules and assay endpoints adopted by hepatotoxicity *in vitro* models in four commercial setups. The table shows that the endpoints prioritized by rule models represent a combination of endpoint types used in the four commercial *in vitro* models. Nuclear receptor activity, and in particular endocrine disruption, was prioritized by rules in but not included in *in vitro* models. With reference to Tables (4-2) and (4-3), nuclear receptor activity represents a key independent cluster among hepatotoxic compounds.

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

Endpoint classes

4.3.4 Influence of physicochemical properties in improving *in vitro/in vivo* associations

We next analyzed the significance of physicochemical properties in the prioritized rule sets, to see to what extent those proxies for exposure [117,118] add value when attempting to anticipate the hepatotoxicity of compounds. For this, the deterioration in accuracy after removal of physicochemical condition combinations from prioritized rules was quantified with respect to the two toxicity thresholds, the results are shown in Appendix B, Figure (B-5). The absolute drop in accuracy by removing physicochemical conditions from all bioactivity rules ranged from 0% up to almost 20%. The overall drop in accuracy was more pronounced at a toxicity threshold of 15mg/kg/day, with an overall error rate usually around 6-11%. The distribution of error rate as a result of removing physicochemical conditions from rules at 500mg/kg/day was broad, ranging from no effect to up to 9% in most cases. Hence, incorporating physicochemical properties improve *in vitro-in vivo* associations, especially at a lower dose (and hence for the more potent toxicants).
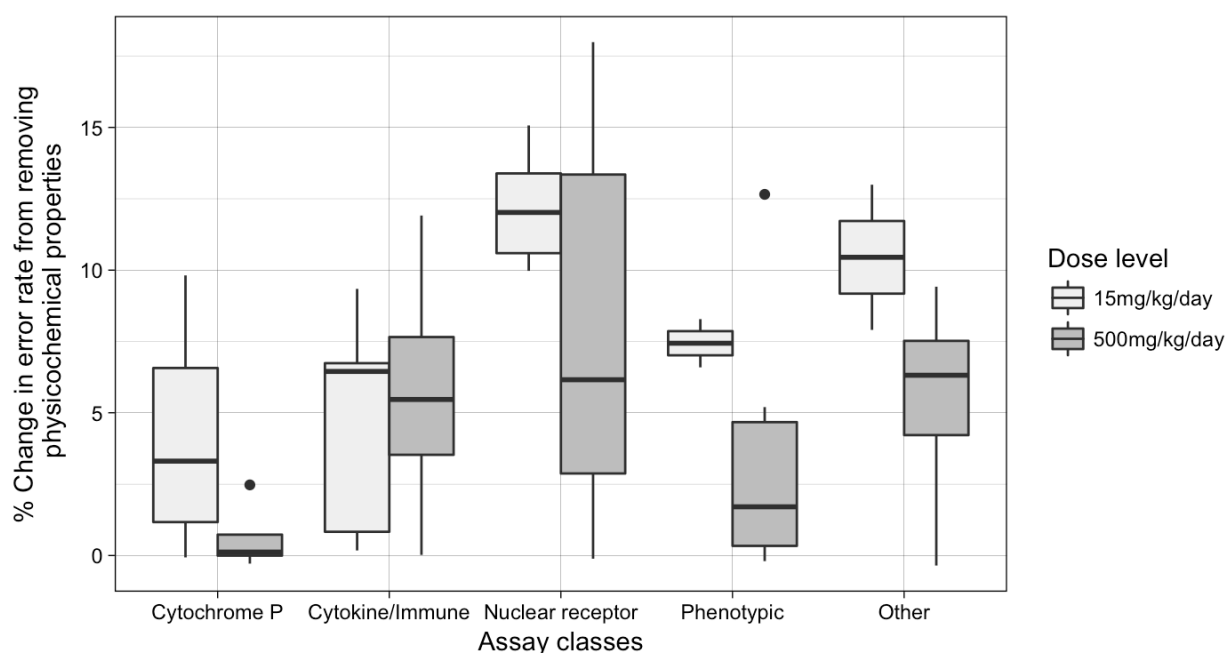
Figure (4-6) Percentage change in error rates as a result of removing physicochemical property conditions from best performing rules as a function of assay class. The overall deterioration of accuracy by removing physicochemical conditions varies with the assay type. The accuracy drop is minimal among rules of Cytochrome P activity and at the dose level of 500mg/kg/day, while very significant accuracy deterioration is seen in rules described by nuclear receptor activity, especially at a threshold of 15mg/kg/day.

The error rate varied not only by the toxic dose level but also by assay class (Figure (4-6)). The error rates of physicochemical conditions were minimal in rules describing Cytochrome P

activity, particularly at 500mg/kg/day. Immunological and phenotypic bioactivities had moderate accuracy deterioration from discarding physicochemical properties, at around 5%. Yet rules describing phenotypic assays showed a significant difference between toxicity thresholds, of more than 10% ($p<0.05$ Wilcoxon rank sum test). Rules related to nuclear receptor activity had the largest increase in error upon removing physicochemical conditions, especially at a threshold of 15mg/kg/day, reaching values higher than 10%. The peak in accuracy drop is for glucocorticoid receptor activity of almost 20%, followed by over 15% for androgen receptor activity at the toxicity level of 500mg/kg/day. Estrogen receptor activity showed an approximately 13% drop in accuracy at both toxicity thresholds.

Direct perturbation of nuclear receptors requires compounds to penetrate the nucleus membrane with sufficient permeability to be toxic at low doses. Overall, for an improved association with *in vivo* effect, some assay bioactivities, such as phenotypic and nuclear related effects, are more dependent on meeting physicochemical conditions, and hence require a proxy for exposure. This is more important for anticipating potent toxicants.

We next analyzed the most frequent physicochemical properties occurring in the rules (Table (4-5)). At a threshold of 15mg/kg/day, these were the number of rotatable bonds (which are equal to or below 6 in rules describing hepatotoxicity), the number of hydrogen bond donors (where hydrogen bond donors are required to be absent) and the number of aliphatic rings (which needs to be equal to or smaller than two). The number of rotatable bonds had the highest frequency, occurring in over a third of the rules. This physicochemical condition was associated with an accuracy deterioration (error rates) equivalent to 8% when it was removed from the rules. The number of rotatable bonds was, however, not abundant among rules at the level of 500mg/kg/day. Instead, the number of rings (<=3) was the most frequent physicochemical property in this case, and was associated with an increase in error rate of 6% and a frequency of 29%.

The above increase in error rates, when removing physicochemical properties from rules, is apparently due to a link between the physicochemical properties of compounds and *in vivo* pharmacokinetics parameters, such as bioavailability.[118,264] Bioavailability is governed by the extent and onset of absorption and distribution, which in turn are linked to some molecular properties such as membrane permeability and plasma protein binding. Good bioavailability *in vivo* means that, upon exposure, compounds achieve sufficient concentrations to achieve an effect at the site of action. Rotatable bond count, for example affects the magnitude of cell membrane permeability of compounds.[196] Majority of compounds that contain 6 rotatable bonds or fewer show oral bioavailability higher than 20%, irrespective to their molecular

weight.[196] Additionally, a number of aromatic rings greater than 3 in a compound is linked to higher plasma protein binding (PPB) (>90%), irrespective to cLogP.[265] Strong PPB may slow the rate of compound distribution among body compartments which will consequently affect the concentration at the site of action.[266,267] Additionally, higher numbers for aromatic rings are linked with Cytochrome P inhibition and higher lipophilicity, both of which are associated with toxicity.[119,265] DeGoey *et al.* have introduced a simple multiparametric scoring function to describe rat oral bioavailability consisting of three properties, namely cLogD, the number of rotatable bonds and the number of rings of a compound.[195] This score which involves summing the values of these properties (with a correction for cLogD). The authors reported a negative correlation between this score and bioavailability with correlation coefficient of -0.41, in agreement with our findings above.

Table (4-5) Physicochemical properties frequently present in prioritized hepatotoxicity rules. The most frequent physicochemical property in rules for threshold of 15mg/kg/day is number of rotatable bonds. Whereas, number of rings is the most frequent physicochemical property in rules at 500mg/kg/day level.

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

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

*Error rate % represents the deterioration in rule accuracy as a result of removing each of the physicochemical properties presented in the table.
** Frequency % is the percentage of rules containing the physicochemical property out of all prioritized rule set.

An additional observation, however, is that rules used different properties at different potency levels, but consistent properties within each potency threshold. Also, different assay types have shown variable magnitudes of dependence on physicochemical properties when measurements are extrapolated into *in vivo* effects. Still, it is apparent that also simple proxies for bioavailability are able to improve the prediction of hepatotoxic compounds on the dataset

employed here. Hence, given that in vivo exposure parameters (*i.e.* dose and Cmax) are often not available at primary stages of drug development, these simple proxies such as the number of rings and rotatable bonds may act as simple alternatives to exposure measures to anticipate the hepatotoxicity of compounds at an early stage.

4.4 Conclusion

We propose in this work a novel framework for generating interpretable rules for the hepatotoxicity of compounds, which use both *in vitro* bioactivity measurements and physicochemical properties. Rules generated from machine learning algorithms were pruned to remove statistically less informative and biologically less meaningful inactive assay conditions from rules describing toxicity. The resulting interpretable rules were used first to prioritize hepatotoxicity *in vitro* endpoints, considering 80% overall compound coverage. The resulting rules were compared with four commercial *in vitro* models for hepatotoxicity. Finally, the influence of physicochemical properties on the derived *in vitro- in vivo* associations were investigated separately for each assay class.

Our results suggest that a set of multiple ToxCast assays are needed for a sufficiently high coverage of hepatotoxic compounds, as no single assay can discriminate toxic from non-toxic compounds. This was also apparent from the information gain derived for single ToxCast assays alone. At two toxicity threshold levels of 15mg/kg/day and 500mg/kg/day, the best-performing modified rules, which cover 80% of toxic compounds, cluster into three major bioactivity classes, namely Cytochrome P activities, immunological responses, and nuclear receptor activities. While overall assays selected for predicting hepatotoxicity overlapped with endpoints used in *in vitro* models from commercial sources, nuclear receptor activity, which represented an independent mechanistic cluster, is not currently covered in this way.

Specific bioactivity combinations were seen in the rules, such as disruption of androgen receptors combined with activities against PPARs, Cytochrome P and increase in mitochondrial membrane potential. These describe perturbation in multiple biological pathways resulting in a greater likelihood of observing toxicity *in vivo*. Incorporating physicochemical properties, in general, also improved the accuracy of rules describing toxicity especially for potent toxicants, *i.e.* those toxic at the toxicity level of 15mg/kg/day. The likely explanation is that, for those compounds, bioavailability plays an important role for toxicity to

be observed, which to some extent can be anticipated by physicochemical properties. The most frequent physicochemical properties used in rules, namely the number of rotatable bonds and the number of rings, are linked to bioavailability parameters, such as membrane permeability and plasma protein binding, respectively. Hence, the likelihood of a compound to be hepatotoxic *in vivo* increases both if it is active in relevant bioassays as well as showing the necessary bioavailability.

There has been an increasing interest in understanding the molecular mechanisms responsible for initiating toxic side effects, in the form of adverse outcome pathways (AOP) frameworks. The assay endpoints screened in the ToxCast project can in principle describe key events in an AOP,[61] since they provide insights to the perturbation in biological processes in the cells by the screened compounds. Our proposed rule-based method can be used as a tool to generate molecular hypotheses so as to guide the identification of key events of the AOPs. In order to be practically successful in this direction, assay coverage in biological space, compound coverage in chemical space and a complete data matrix linking both domains are crucial.

Another application to this rule method framework is to optimize *in vitro* models for toxicity screening. For example, in order to improve the compound coverage of hepatotoxicity *in vitro* models, we recommend assays from three major bioactivity classes are incorporated when testing for hepatotoxicity, which are Cytochrome P activity, immune responses and endocrine disruption. This is in addition to phenotypic readouts such as cell viability, cell stress, mitochondrial impairment and changes in cellular organelles. The combination of assays from all areas will then allow for the better detection of hepatotoxic compounds. We also recommend considering physiochemical properties as simple proxies for *in vivo* exposure measures, such as C_{max}, when attempting to anticipate potential hepatotoxicity. While simple, those properties are fast to calculate and they are able to improve predictivity of *in vivo* hepatotoxicity, at least on the data based in this study.

The workflow presented here can finally be generalized to other types of toxicity, considering any type of chemical and biological input data, provided coverage in the chemical and biological domain for the toxic endpoint of interest is given.

5 DISCOVERING COMPLEX MECHANISMS IN ACUTE TOXICITY USING MULTISOURCE DATA AND EMERGING PATTERNS

5.1 Introduction

The use of *in vitro* screening for systemic acute toxicity, that are derived from a mechanistic understanding of modes of action, is as yet immature.[6] The current state of art for alternative approaches include *in vitro* assay for basal cytotoxicity using neutral red uptake (NRU) to identify starting doses in animal testing.[125] Some of the key target specific mechanisms identified for acute toxicity include the interference with neurotransmission, ion channels, energy haemostasis, antioxidant pathways and cellular integrity.[6] Several AOPs were proposed around these key events (KE) to describe acute mortality[127–129], which are published in the AOP database. These are represented by the modulation of ionotropic GABA (gamma-Aminobutyric acid) receptors[129] and ionotropic glutamate receptors[128], and the inhibition of acetylcholinesterase.[130] However, not much is understood on how perturbation of more than one biological pathway influences *in vivo* observations as the result of compounds' polypharmacology. Studying these interactions is especially essential in cases where complex mechanisms are involved such as acute toxicity.

In the case of acute organophosphate poisoning, the key mechanism driving acute toxicity is the inhibition of acetylcholine esterase. A study on the pathological pathways associated with the potent AChE inhibitor, chlorpyrifos-oxon, on zebra fish had shown multiple downstream effects including calcium ion dysregulation, immune and inflammatory responses and cytotoxicity, collectively referred as the cholinergic toxidrome.[268] Such complex downstream effects can be points of perturbation via diverse stressors due to compounds' polypharmacology, or exposure to bioactive compound mixtures. In this context, adverse effects should be captured by multiple pathway perturbations attributed to multiple key events.

One way to extract the complex associations that can discriminate between classes is emerging patterns. Emerging patterns are a form of rule models which identify feature sets frequent in

one class of the data, but not the others.[151] The main implementation of emerging patterns in cheminformatics is to derive structural alerts. This was first conducted by Auer and Bajorath[269] who used discretized physicochemical properties were used to describe target bioactivities. Sherhod has used emerging patterns[67] and jumping emerging patterns[76] to discover structural alerts for Ames mutagenicity and hERG activity. Emerging patterns were also implemented on gene expression profiles to predict toxic compounds[270] and identify diagnostic gene groups.[271] Yet, emerging patterns have limited applications in cheminformatics,[272] despite their ability to extract interpretable insights about patterns in data.

In this study, we analyse the complex associations between acute toxicity and potential KE/MIE represented as structural alerts and bioactivity properties. To the best our knowledge, this is the first attempt to use emerging patterns by combining chemical and bioactivity data to describe toxicity. This study uses data collected from Tox21 screening measurements, predicted biological targets, known toxicophores and Globally Harmonized System (GHS) labels for acute toxicity. In order to generate hypotheses about features' additive and synergistic interactions, we integrate emerging patterns with network analysis and mutual information. We illustrate how some features, which on their own show moderate correlations with acute toxicity, play vital roles to intensify the liability of other features through synergy. The conditional associations derived here provide novel insights into the mechanisms of acute toxicity. We believe that these findings can encourage interpreting potential toxicity as an assembly of interacting, rather than independent, key events that reflect multiple perturbations in the physiological system.

5.2 Materials and Methods

5.2.1 Datasets

Tox21 data The Tox21 project involve high throughput screening measurements of almost 10,000 compounds against phenotypic cytotoxicity and target-specific assays for toxicity pathway profiling.[53] Tox21 bioactivity data were collected from PubChem[273] (accessed 21 Jul 2017) for 8747 unique compounds labelled with PubChem ID against 110 assays (see Appendix Table (C-1)). The activity label used in this study is the percent activity compared with a positive reference determined from three replicates at 1μM. The conventional binarization procedure in PubChem is determined by a 40% activity threshold, above which a compound is considered active, *i.e.* 1 for active and 0 for inactive. Here we also added three thresholds levels, namely 20%, 60% and 80% activities (in addition to 40%). Therefore, each compound was described in an assay by four levels of bioactivity labels, resulting in 440 Tox21 assay descriptors. For example, if a compound has a bioactivity of 66% in a given assay, then this compound will show a positive label at levels of 20%, 40% and 60% but not 80%. This intervention was introduced so as to consider potency in Tox21 assays when generating the toxic describing rules.

Predicted bioactivities Compounds extracted from the Tox21 data were also annotated with their predicted bioactivities using an in-house predictive model "PIDGIN" (Prediction IncluDinG INactives) version 2.[274] PIDGIN was built using targets annotated against compounds in PubChem including 3394 unique targets in a Random Forest model. In this study, targets that have shown True Positive Ratio of 0.7 or higher were considered summing to 1790 targets. The annotations outcomes are in binary format (1 if active and 0 if inactive).

Chemical Descriptors Toxicophores were generated using the Online Chemical Modeling Environment (OCHEM) ToxAlerts server[73] against over 2300 structural alerts collected from 17 references. The server performs normalization and curation on query compounds after which it matches the curated compounds with the SMARTS of predefined toxicophores. The results were stored in form of binary matrix which represent the presence or the absence of the structural alert in the compound.

Additionally, substructures observed in at least 2% of the compounds in the dataset were generated using MoSS[275] node in KNIME.[200] The MoSS undergoes a graph based fragmentation followed by appropriate curation to avoid redundant detection of substructures.

GHS acute toxicity The Globally Harmonized System of Classification and Labelling (GHS)[276] is international standard system agreed by the United Nations to classify chemical compounds for their hazardous potential. The system includes a labelling system for toxicities such as acute toxicity, mutagenicity and organ specific toxicity. In this study, we used acute toxicity (oral, dermal and inhaled) to label compounds in the dataset. Compounds are assigned with a severity level from 1 to 5, where 1 is the most potent toxicant whereas, 5 demonstrate the least severity. These classes are derived from the LD_{50} for oral and dermal and equivalent LC_{50} for inhalation. The data were extracted from PubChem[273] for the compounds in the Tox21 dataset (Accessed 26 July 2017). The GHS classifications reported in PubChem are compiled from multiple sources, namely the European Chemcials Agency (ECHA), the EU regulation (EC), the Chemical Management Center (CMC) of Japan National Institute of Technology and Evaluation (NITE) and the Australian Hazardous Chemical Information System (HCIS). The GHS grading was discretised into an arbitrary two-class toxicity label, *i.e.* toxic and non-toxic. Compounds were considered toxic if the classification of either oral, dermal or inhalation were at levels 1,2 or 3, and non-toxic otherwise. The compounds selected for this study must show at least a valid classification for oral acute toxicity. Also, wherever a contradiction in the GHS labelling system was observed, the associated compounds were excluded. Compounds with valid labels sum up to 3573, 1261 toxic and 2312 are non-toxic.

Chemical curation Compound structures were extracted in SMILES format using the PubChem exchanger service for their corresponding PubChem CIDs. Chemical standardization of compounds was conducted in ChemAxon[198] considering the following specifications: add explicit hydrogens, clean 2D, remove fragments, neutralize, mesomerize and tautomerize

In order to generate a complete dataset across the different sources, only compounds that overlapped between the Tox21 data and GHS classification were used for analysis. Further, the compound set was curated in order to exclude compound mixtures and compounds including heavy metals. The overall integration of the dataset resulted in 2000 compounds, 993 toxic compounds and 1007 non-toxic. Only features which are present in 5 or more compounds were used. The total number of features used was 3732 representing 1789 predicted targets, 440 Tox21 assay activities, 1259 ToxAlert toxicophores and 242 MoSS derived substructures.

5.2.2   Emerging patterns generation

Emerging patterns (rules) were generated using CPAR algorithm developed by LUCS-KDD (Liverpool University Computer Science – Knowledge Discovery in DATAS). The CPAR

algorithm is explained in Chapter 2 section 2.2.2.2. All parameters were set as default (decay factor of 2/3 and similarity ratio of 1:0.99) except for the minimum gain by the split which was adjusted to 2.5. In order to generate meaningful and biologically interpretable emerging patterns, we applied data space pruning. The search within the descriptors space was restricted so as to align with causal or correlation links between these features and *in vivo* toxicity. The emerging pattern algorithm CPAR takes the input data in for of transaction, which are lists of features (see section 2.2). Hence each binary feature is described using two labels, *i.e.* the presence and the absence of that feature. To constrain the directional association for *in vitro* data, only the presence of an activity (and the presence of substructures) are included in the input set to generate toxic describing patterns. Similarly, patterns describing absence of toxicity used input data of variables describing absence of *in vitro* activity, and the presence of the substructure.

The training set used was randomly sampled to create 70% subsets on which the algorithm was run with 50 repeats. Emerging patterns performance was calculated as confidence and coverage using the all training set. Unique patterns were then selected according to a minimum of 70% confidence and minimum Fisher test p-value of 0.01 to exclude insignificant correlations. This process resulted in 7381 patterns describing toxicity and 3866 patterns describing absence of toxicity. In this study, terms emerging patterns and rules are used exchangeably.

5.2.3 Rule network analysis

Rules were visualized as networks. Feature conditions in rules are represented as nodes, whereas, edges connecte node features that co-occurred in the same rule at least once. Figure (5-1) demonstrates how groups of rules were converted into network representations. First, rules were represented as feature lists describing either toxicity or absence of toxicity, labelled as 1 or 0 in Figure (5-1), respectively. For each class label, the list representation is converted into a binary matrix where each row is equivalent to a rule and columns are the unique feature set. Next, the binary matrix was converted into an adjacency matrix to describe feature pairs that co-occurred in at least one rule. The numeric value in the adjacency matrix reflected the frequency of observing feature pairs in unique rule sets. Finally, this adjacency matrix was used to construct a network of features as nodes and adjacency as edges.

Visualization and analysis were conducted in Cytoscape software.[277] Cytoscape allows to perform number of analytical measures on networks to extract the topological properties. These properties include three key components which are i) network as whole, ii) nodes and iii) edges.

Networks can be described by their size, degree of saturation and betweenness (average shortest path to go from a node to another in the network through connected edges). Nodes are often described by centrality measures such as the degree, which describes the number of unique first neighbours. Edges can be directional or undirectional, and can be weighted.
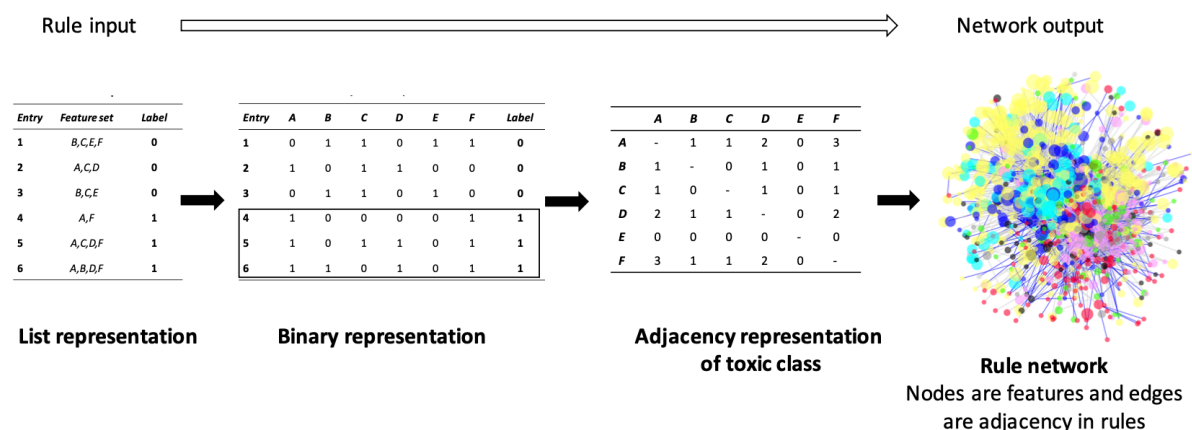
Rule input →————————————————————————————→ Network output

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

Figure (5-1) Workflow for generating rule networks. The generated rules are represented as lists of feature conditions describing presence or absence of acute toxicity. The list representation of rules was converted into a binary matrix of rule input in rows and feature set in columns, in which (0,1) reflect whether the feature is used or not used in a rule, respectively. The binary representation was then converted into an adjacency table, where the digits represented how many times feature pairs co-occurred in a unique set of rules. The adjacency table was used to generate a network of features as nodes and adjacency as edges. Networks were then generated for each class, *i.e.* presence and absence of toxicity.

5.2.4   Synergy measures

One way to analyse rules of multiple conditions is to measure the interactions between these conditions. Analysing these interactions is useful to understand how combining multiple features at a time would affect the odds for observing the outcome. For example, it is possible to detect synergy between features, where their combination acts beyond the additive effect. It should be noted, however, that the synergy in this context represents the improvement in probability or likelihood of observing toxicity not the increase in toxic intensity.

We measured synergy between all feature pairs co-occurred in rules. For rules composing of more than two features, all possible combination pairs were investigated. For example, in a rule of three conditions {a, b and c}, the investigated pairs are a-b, a-c and b-c. Two measures

were used. One is based on the normalized mutual information as described by Fang *et al* [278] and the second measure is based on odds ratios as reported by Cortina-Borja.[279] In both cases, the association statistic is measured for individual features in a rule as well as for feature pairs, and relative to the toxicity outcome.

I) Synergy using mutual information

For c_i ={a,b} pair in Rule i against toxicity label Z, where $c_i \subseteq$ Rule i, and a, b are feature conditions.

$$Synergy \ = \ MI(ci, Z) - [MI(ai, Z) \ + \ MI(bi, Z)] \qquad [10]$$

$$Improvement \ = \ MI(ci, Z) - \ max[MI(ai, Z) , MI(bi, Z)] \qquad [11]$$

Where MI is the normalized mutual information[278] calculated for a, b and their combination c_i.

II) Synergy Factor from odd ratios (equivalent to interaction weight in the regression equation)

$$Synergy \ Factor \ = \ ORci \ / \ (ORa \times ORb) \qquad [12]$$

Where OR is the odds ratio[279] calculated for a, b and their combination c_i. The criteria used to label a feature pair as synergistic are magnitudes greater than 0.001, using the MI statistic, and values greater than 1 for the OR-based synergy scores. The relationship between the two synergy measures is represented in Appendix C, Figure (C-1).

5.3 Results and Discussions

5.3.1 Univariate associations between biological and chemical properties and acute toxicity

First, in order to understand which classes of features are being captured, as single variables, the highest associations with acute toxicity, we examined the normalized mutual information (NMI) between all feature-toxicity label pairs. The distribution of NMI is visualized in Figure (5-2). The maximum possible NMI value for a perfect correlation is 1 (0 in $\log_{10}$ scale)
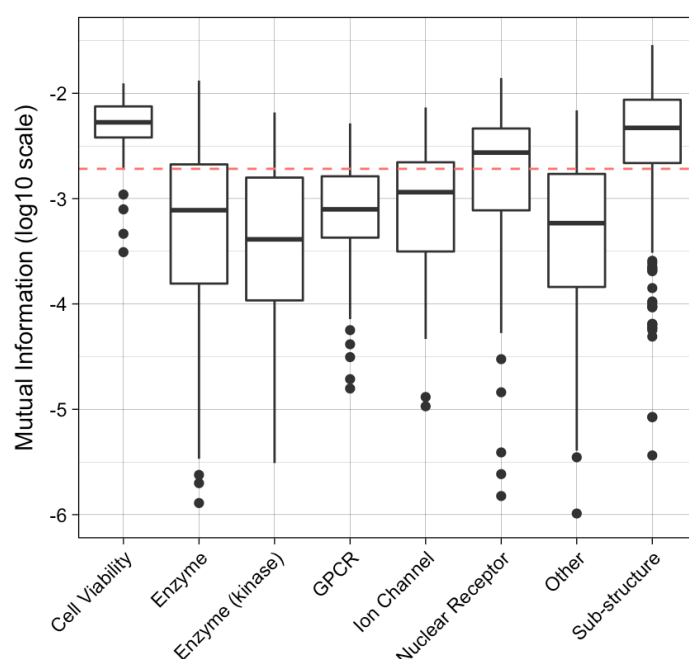

Figure (5-2) Distribution of the normalized mutual information (NMI) for each feature class against acute toxicity in $\log_{10}$ scale. The dashed line represents the median value of the overall NMI (-2.7). The majority of cell viability and substructure features show NMI values above median, which mean these features, as single variables, have the strongest association with acute toxicity. Except for nuclear receptor activity, target specific features, namely enzymes, ion channels and GPCR show average NMI values below median.

Features were grouped into 8 classes, namely substructures, cell viability, nuclear receptors, enzymes, kinases, ion channels, G protein-coupled receptors (GPCR) and 'other' targets, which include transporters. The median value of NMI as shown in Figure (5-2) is around -2.7 and a maximum of -1.5 in $\log_{10}$ scale. The difference between these values and the maximum possible value for NMI means that acute toxicity cannot be captured by a single variable. Figure (5-2) also shows that substructural and cell viability features had the top values for $\log_{10}$ NMI

of at an average of -2.3 (equivalent to NMI of 0.005). Cell viability had also shown the least spread in values compared to all other features. Following in the rank of the association strength were activity against nuclear receptors and ion channels, with $\log_{10}$ NMI of -2.6 and -2.9 respectively. Other features, including enzymes, kinases and GPCR, had an average $\log_{10}$ NMI values less than -3 (less than 0.001 NMI). Overall, cell viability and structural properties have stronger associations with acute toxicity than target-specific bioactivities.

5.3.2 Performance and structure of toxic and non-toxic rules

To unravel complex associations between acute toxicity and chemical and biological properties, emerging patterns were generated to capture these associations. Emerging patterns follow conventional rule measures of coverage and accuracy, and here the terms rules and emerging patterns were used exchangeably.

The first set of generated rules were filtered according to accuracy and significance in the Fischer test (see Methods). The characteristics of selected rules are shown in Table (5-1) with regards to counts, performance, average number of conditions in rules describing presence and absence of acute toxicity.

Table (5-1) Characteristics and performance of rules describing toxicity and rules describing non-toxicity. The average accuracy (confidence) of rules used in the analysis was above 80 % for both labels. Non-toxic rules described, in general, had more number of compounds per rule and more number of conditions per rule compared to toxic rules.

	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

The average confidence as in Table (5-1) was above 80%. Overall, non-toxic rules covered more number of compounds per rule, *i.e.* 34 compounds by non-toxic rule compared to 26 compounds by a toxic rule (Table (5-1)). The complexity of the rules can be estimated from the average number of conditions used per rule. The simplest rules consisted of one condition, however, these approximately constituted 8% and 4% of overall unique toxic and non-toxic

describing rules, respectively (Table (5-1)). Also, the overall number of conditions in rules describing non-toxicity was more than 3 which is greater than that observed in toxicity describing rules.
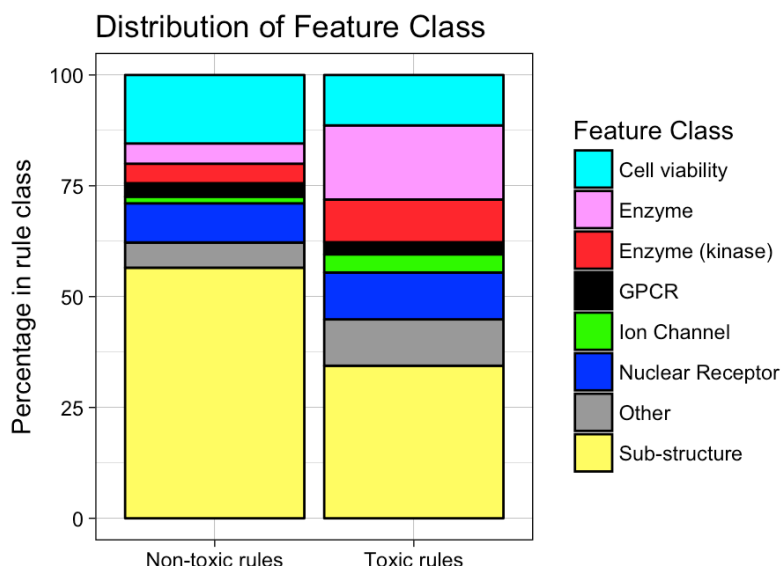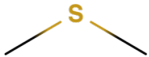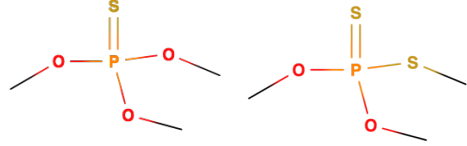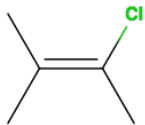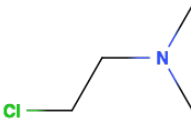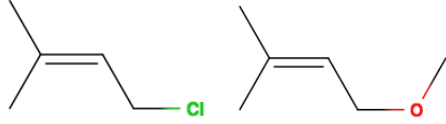

Figure (5-3) Percentages of feature classes used in toxic and non-toxic describing rules. Rules for toxicity use diverse set of features, whereas, rules capturing non-toxicity are mainly described by cell viability and structural features.

Next, we examined the distribution of feature classes in rules for the two toxicity levels represented in Figure (5-3). The unique features used by toxic and non-toxic rules intersect to a great extent, *i.e.* 512, which represent around 50% and 70% of all features used by toxic and non-toxic rules, respectively. However, the variability in feature frequency between rules resulted in the distribution seen in Figure (5-3). Toxic rules showed 3-fold higher proportions o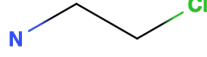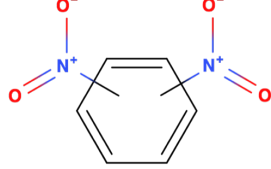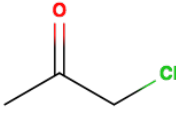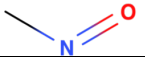f specific effects such as activity against enzymes, kinases and ion channel receptors. Non-toxic rules, on the other hand, had higher abundance of substructure- and cytotoxicity-describing features than toxic rules, which together represented over 70% of features in the former compared to less than half in the later. This pattern can be explained by the relatively high NMI magnitudes of these two feature classes with acute toxicity. This means they have higher generalizability to describe the presence or absence of *in vivo* acute toxic effects. Nuclear receptors and GPCR activity appeared in two rule classes at almost equal fractions. Overall, both rule types showed acceptable performance and share large proportions of unique features, however, the frequency at which these features were used by each rule class varied significantly.

5.3.3 Single condition rules capture known features of acute toxicity

Although the majority of rules had more than one condition, fewer than 10% were represented by single conditions (Table (5-1)). Single condition rules were described mainly by substructural features and potent cytotoxicity to a combined percentage of 75% (Appendix C, Figure (C-2). Examples of single feature rule with confidence higher than 80%, compound coverage greater than 15 and frequency equal to or higher than 2 are represented in Table (5-2). Substructure conditions such as sulfenic acid derivatives and organophosphates, observed in Table (5-2), are known structural alerts for potent toxicants.[280–282] These features also showed strong associations as estimated by log_{10} NMI at levels less than -2. Halogenated ketones, shown in Table (5-2) are examples of Michael acceptors, which is a known mechanism for chemical induced toxicity.[6,281] Single feature rules, in general, had an overall log_{10} NMI among the highest in comparison to average values of features in all rules (Figure (5-2)). Cytotoxicity is also a key endpoint in assessing the potential acute toxicity[6,19] which is used to estimate the starting dose in animal tests.[125] Therefore, we can conclude that features with strong univariate associations and features captured in single condition rules represent known structural alerts and biological properties of compounds inducing acute toxicity.

Table (5-2) Selected examples of single condition rules and the corresponding normalized mutual information (NMI). The selection was based on confidence higher than 80%, compound coverage greater than 15 and frequency equal to or higher than 2. Majority of single condition features were described by potent cytotoxicity or substructure features, which also show relevant high NMI values with reference to Figure (5-2).

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

5.3.4 Networks of multi-condition rules demonstrate complex feature-toxicity associations

To investigate rules of more than one condition, networks were generated to describe complex patterns in each class, *i.e.* toxic and non-toxic labels. The characteristic elements in the network structure are nodes and edges. Nodes in rule networks represent individual features, whereas, edges connect feature pairs co-occurred in one rule. Rule networks can help examine complex toxicity associations using network measures of connectivity as well as algorithms for detecting communities within the network. The aim is to get insights about important features related to acute toxicity which are beyond univariate associations.

Two rule networks were generated describing the presence and absence of toxicity. The characteristics of these networks are presented in Table (5-3), including number of nodes and edges, and network density. Network density is the proportion of edges counts present in the network to the edge counts of a saturated network. Toxic network consisted of 937 unique nodes and 8280 edges, and had a density of 0.019 (Table (5-3)). The non-toxic network involved 737 nodes connected via 7500 edges with a density of 0.028. The density is slightly higher in non- toxic network, reflecting rule complexity (number of conditions per rule) (Table (5-1)) and higher cross connections between the nodes (Table (5-3)).

Table (5-3) Topology characteristics of toxic and non-toxic networks. The first two clusters in both network classes had the largest number of unique features. Despite the similarity in size seen in cluster 1 and 2, the latter has double number of edges, and hence, double network density. Cluster 2 is dominated by target specific features in the toxic network and mixed feature classes in the non-toxic network. Cluster, 1 on the other hand, is mainly characterized by cell viability and nuclear receptor activity in toxic and non-toxic network alike (Figures (5-4) and (5-5)). Hence, target-specific features are involved in more complex associations with acute toxicity.

Toxic network					
Clusters	Overall	1	2	3	
N# nodes	937	376	328	184	
N# edges	8280	1748	3047	476	
N# synergy connections	1460	258	637	104	
Density	0.019	0.025	**0.057**	0.028	
Non-toxic network					
Clusters	Overall	1	2	3	4
N# nodes	737	279	230	132	48
N# edges	7500	1535	2356	269	63
N# synergy connections	336	64	93	31	4
Density	0.028	0.04	**0.09**	0.031	0.056

To detect groups characterized by similar features adjacency, community clustering was applied, based on Girvan-Newman method.[283] The algorithm finds communities of similar neighbours by cutting edges between node pairs exhibiting the highest betweenness (shortest path between a nodes pair in the network). Figures (5-4) and (5-5) show toxic and non-toxic networks, respectively. Clustering of toxic rule network resulted in 3 large clusters, with nodes sizes of 376, 328 and 184, and 9 clusters each has less than 20 nodes. Via applying community clustering to the non-toxic network, 15 groups were observed, the first 4 clusters involved 279, 230, 132 and 48 nodes, whereas each of the remaining clusters included less than 10 nodes. In further analysis, we will focus on the first three clusters in rule networks as they capture the majority of features.

The distribution of feature types varied between toxic and non-toxic networks. The content of clusters in toxic network showed specificity in the distribution of feature classes, whereas non-toxic clusters are more consistent in feature class distribution. Almost all clusters in both toxic and non-toxic networks included structural features. Cluster 1 in the toxic network is dominated by nodes describing cytotoxicity in Tox21 assays and nuclear receptor activity. Both cluster 2 and 3 have high proportions of specific biological targets under the enzyme and kinase classes. In addition, cluster 2 and 3 also include nodes describing activity against kinases, ion channels and GPCR. We can conclude two key points, firstly, there is a separation in toxicity induced by specific mechanisms and non-specific cell stress mechanisms. This aligns with previous reports that *in vivo* toxicity occurs as a result of either unspecific cytotoxicity and cell stress, or perturbation of specific pathways.[85] Secondly, the unspecific cytotoxic effect can be captured by the combination of cell viability assays and nuclear receptor disruption. This combination has previously been observed in ToxCast cluster analysis in Chapter 3 and in the nuclear disruption cluster of hepatotoxicity in Chapter 4.

Moreover, cytotoxicity and nuclear receptors occupied central positions in the first three clusters of non-toxic network, indicating critical roles in predicting acute toxicity. In other words, in order to ensure lack of acute toxicity, absence of cytotoxicity has to be ensured, as the absence of target specific perturbation may not be sufficient.

Figure (5-4) Toxic rule network clustered via Girvan-Newman algorithm. Nodes represent features and edges connect features co-occurred in a single rule. Nodes are coloured by feature class where the node size is proportional to the magnitude of mutual information with acute toxicity. Blue edges connect synergistic pairs calculated via normalized mutual information (NMI) (see Methods). The majority of features are captured in the first three clusters. Substructures occupy conserved spaces in each cluster. The dominant bioactivity classes in cluster 1 are cell viability and nuclear receptor activity, whereas, cluster 2 and 3 are dominated with target based features including enzymes and kinases. This means that the majority of rules capturing cytotoxicity, as predictors for acute toxicity, are independent of rules capturing specific target activity.
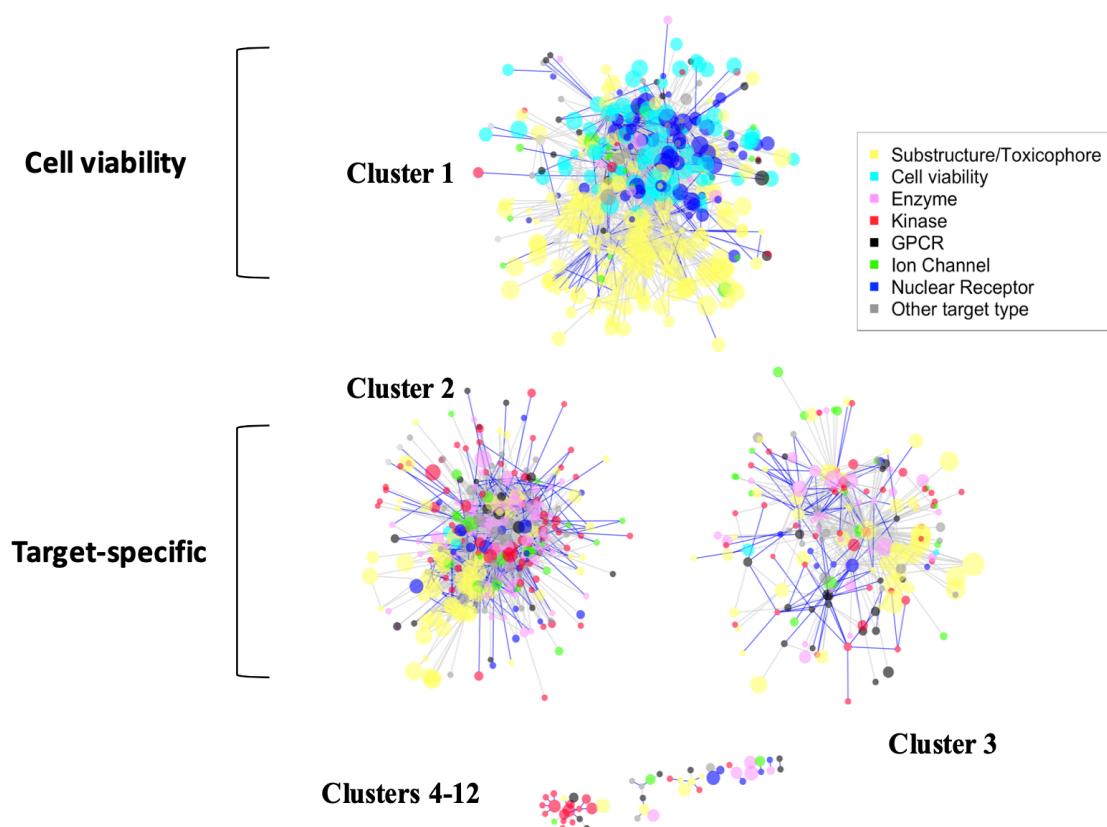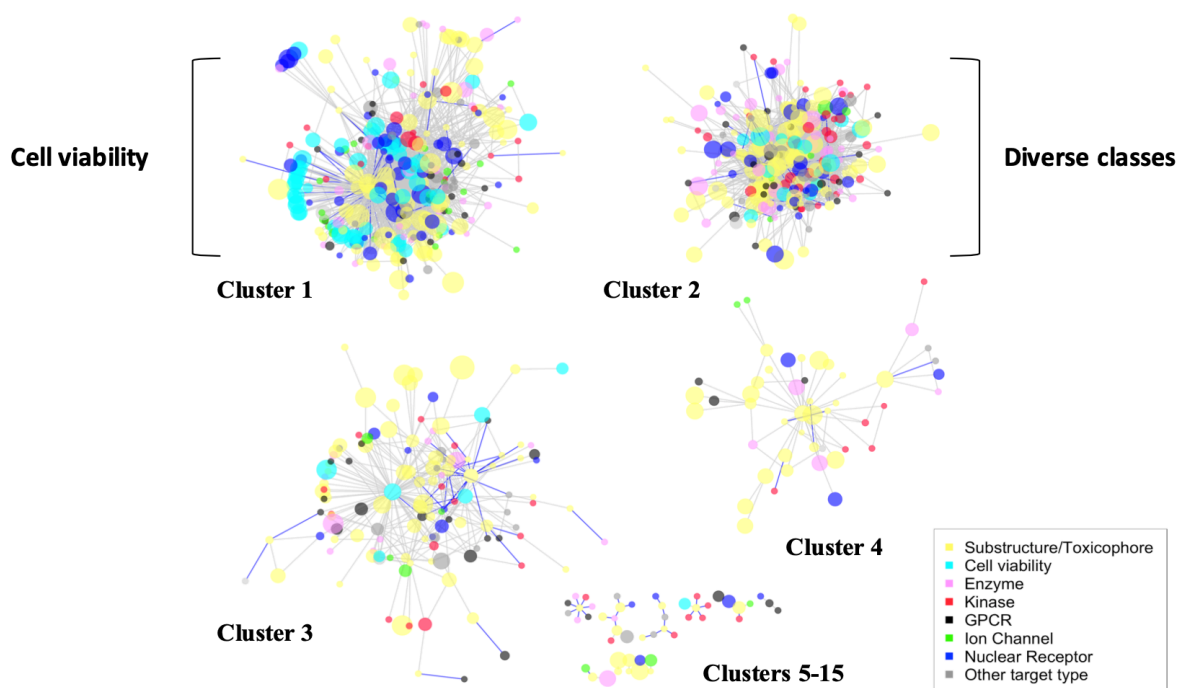
Figure (5-5)   Non-toxic rule network clustered via Girvan-Newman algorithm. Nodes represent features and edges connect features co-occurred in a single rule. Nodes are coloured by feature class. Fewer synergistic connections (blue edges) are seen in non-toxic network in comparison to toxic network (figure (5-4)). Majority of features are captured in the first four clusters. Substructure features are scattered in each cluster. The dominant bioactivity classes in cluster 1 are cell viability and nuclear receptor activity, cluster 2 show mixed bioactivity types, whereas, clusters 3 and 4 are mainly described by substructure features. This means that absence of toxicity is conditional in most cases by absence of cytotoxicity. Even when target specific activity is absent, this must be combined with absence of cytotoxicity as seen in cluster 2.

In order to examine the complexity of associations between the various feature classes with acute toxicity, densities of individual clusters were investigated. The level of edge density varied between clusters in both toxic and non-toxic networks (Table (5-3)). This variability, however, does not follow the cluster size. For example, cluster 1 had as double the number of nodes as cluster 3 in the toxic network. Yet, they had a similar level of density of 0.025 and 0.028, respectively. Cluster 2, however, which is similar in size to cluster 1 in toxic network has shown almost double number of edges and density, as well as, 2.5 folds more of synergistic connections. Hence, more dense clusters involve more complex associations between features used and *in vivo* toxicity. Therefore, the complexity in Cluster 2, which is dominated by target-specific features, indicates the need to address the effect of multiple perturbations for a proper extrapolation into *in vivo* toxicity. In other words, whereas, cytotoxicity and structural features

have direct univariate associations with toxicity, target specific features demand understanding the compounds' polypharmacology to predict toxicity.

Additionally, to investigate how different levels of potency in Tox21 assays behave in the toxic network, potency thresholds in assays were plotted against degree connectivity in the toxic network. Figure (5-6) shows that the most connected Tox21 assays in the toxic network were describing low potency thresholds. As the potency level increases, the number of connected nodes decreases. This means that when compounds show high potency in an assay fewer number of other conditions is required to translate into *in vivo* toxicity. However, in order to inform whether weak activity in an assay can lead to toxicity, more features are to be combined and considered. This observation is important because it indicates the significance of polypharmacology analysis for activities at low potency. Whereas the state of art for the study of off-target activity utilizes safety margin calculations using maximum plasma concentrations (C_{max}),[48] this approach fails to consider how combining multiple weak activities would translate into clinical effects.
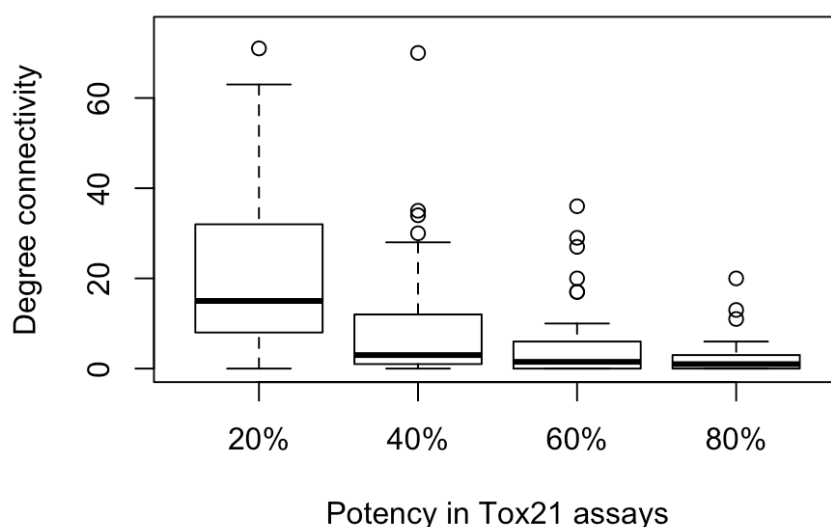
Figure (5-6)   Change in degree connectivity in the toxic network as a function of potency cutoff in Tox21 assays. As the potency level decreases, the number of associated conditions increases. Hence, to translate into *in vivo* toxicity, low potency in assays require a larger number of other conditions, and vice versa.

5.3.5   Important features in toxic and non-toxic networks

We next investigated the most important features deriving the associations in each cluster. These are represented by the nodes showing the highest number of unique connections (highest degree) as well as the highest mutual information values (Tables (5-4) and (5-5)). The features were selected for both structural and bioactivity properties.

5.3.5.1 Known mechanisms and key events in acute toxicity

Some trends in bioactivities in networks clusters reflect known mechanisms of acute toxicity such as cytotoxicity in cluster 1 of toxic networks. Cytotoxicity is a classical *in vitro* test for acute toxicity using neutral red dye uptake (NRU) which is conducted to determine the starting doses for *in vivo* oral acute toxicity tests.[125]

Other activities that are highly connected in one or both networks include glutamate receptors, disruptors of mitochondrial membrane potential and acetylcholinesterase (AChE) (Tables (5-4) and (5-5)). Glutamate is an excitatory neurotransmitter of which both overactivation and inhibition of its receptors are associated with neurotoxicity.[284,285] Impairment of mitochondrial membrane potential is a known biomarker for cytotoxicity[286] as well as *in vivo* acute toxicity.[287] Acetylcholine (ACh) is a carbamate derivative which is cleaved and deactivated by AChE into acetic acid and choline. The toxic effects of inhibiting AChE are initiated by the accumulation of ACh in neuromuscular junction leading to initial overstimulation followed by desensitization and ultimately muscle paralysis.[288] Cholinergic crisis can be a lethal result of the exposure to potent AChE inhibitors. It is also a classic mechanism in acute toxicity, mainly manifested as respiratory failure.[289] The inhibition of ionotropic glutamate receptors, AChE and disruption of mitochondrial membrane potential are known molecular initiating events (MIE) in acute mortality.[6] Therefore, the most connected target-specific activities in networks overlap with key events known to associate with acute toxicity.

5.3.1.1 Substructural features

The patterns in substructural features in the toxic and non-toxic networks showed generic and specific properties. Example of generic properties, in the toxic network, is the dominance of N and O atoms, and aromatic amines in cluster 2, and P and S in cluster 3, evident by their high connectivity. A possible explanation for the connectivity of aromatic amines and N and O atoms is the association of these features with compound promiscuity. It has been reported that compounds with basic centres and aromatic rings are promiscuous.[43] Additionally, promiscuity is negatively associated with the count of oxygen atoms.[42] Interestingly, these features are present in the most dense cluster in the toxic network, cluster 2, which describes target-specific activites. Additionally, in the non-toxic network, alcohols, unsaturated groups and carbonyl functionalities dominated clusters 1,2 and 3, respectively.

Specific substructural features were also seen. For example, carbon chain linked to a polar group such as oxygen and nitrogen has been observed as highly correlating with acute toxicity in cluster 1 in the toxic network as well as clusters 1 and 2 in the non-toxic network (Tables (5-4) and (5-5)). It was also seen as highly connected in toxic network in cluster 1. This feature resembles the structural backbone of surfactants, which constitute of a hydrophobic carbon chain linked to an ionic or non-ionic polarized functional group.[290] Surfactants are known to exert skin irritability as a function of solubility and level of organization in micellar structures.[291] Another feature, which had both high association with toxicity as well as connectivity to other conditions, was halogen functional groups, in cluster 2 of toxic network and cluster 3 of non-toxic network. Haloalkanes are liable to form free radicals[292] and hence exert toxic effects via binding and interfering with cellular molecules.[293] Heterocyclic substructural conditions were highly connected in cluster 1 in toxic network and cluster 3 in non-toxic network. The toxicity of heterocyclic compounds can vary according to adjacent functional groups and biological properties.[294] For example, heterocyclic aromatic amines in food contaminants can cause DNA damage which can result in apoptosis.[295] Nitrogen and sulfur mustard in cluster 2 of the toxic network represent a known cytotoxic substructures which are associated with gastrointestinal tract and kidney toxicity.[296] Michael reaction acceptors[6] in cluster 1, urethane derivatives and thiophosphoric acid derivatives[281] in cluster 3 in the toxic network represent classical alerts for acute toxicity. Urethane pesticides can induce toxicity by inhibiting acetylcholinesterase, a classic mechanism for systemic acute toxicity.[297] Therefore, the most important chemical features in the networks may represent either specific structural alerts/molecular initiating events or generic functionalities and compound promiscuity.

5.3.1.1 Modulatory nuclear receptors
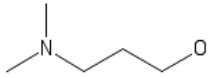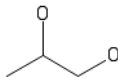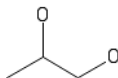
The most connected nodes in both toxic network and non-toxic describing networks do not always represent statistically the most associated features with acute toxicity (Tables (5-4) and (5-5)). For example, thyroid hormone receptor (TR) signalling pathway, vitamin D receptor (VDR) and arylhydrocarbon receptor (AhR), which have significantly high connectivity in rule networks, regulate homeostasis and defence against xenobiotic. These nuclear receptors are also common in regulating cellular oxidative stress. TR had shown connectivity of 216 in toxic network, which is 10 fold higher than the majority of features used in the toxic network. TR is involved in energy regulation, metabolism[298] and maintaining cardiac function.[299] Studies have shown that both hyperthyroidism and hypothyroidism are associated with oxidative stress.[300]

Table (5-4) Important substructural and bioactivity features in toxic network clusters selected the by normalized mutual information (NMI) and degree connectivity. Highly connected features in the network means that they are used frequently as conditions in the rules. It can be seen that most connected features in the network do not always represent features with the strongest associations with acute toxicity.

	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 reaction acceptor)	- oxygen- linked to aliphatic carbon chain (variable length) - nitrogen linked to saturated carbon chain (variable length)	- **TR (antagonist)** - **Nrf2** - CAR (antagonist) - Glutamate receptor (ion channel) - Disruptors of mitochondrial membrane 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

Highlighted features in the table had 150 or higher values for degree centrality in toxic networks. The median value of degree in network is 7. Features with the top mutual information had values higher than -2.2 (see figure 2 for overall distributions).

Table (5-5) Important substructural and bioactivity features in non-toxic network clusters selected by the normalized mutual information (NMI) and degree connectivity. It can be seen that the most connected structural features in cluster represent distinct functional groups. Also, an overlap in the most important bioactivity features can be observed with toxic network.

	Structural alerts		Bioactivity features	
	Top degree	*Top NMI*	*Top degree*	*Top NMI*
Custer 1	- **Alcohols (aliphatic and phenolic)** - aromatic amines - S -	- oxy- linked to aliphatic carbon chain - keto-linked to aliphatic chain - phenyl sulphonyl -	- Cell viability - ER (antagonist) - P53 (agonist) - Nrf2 - AhR - AMPK - Neuronal Ach ion channel	- CAR (antagonist) - ER (antagonist)
Custer 2	- **O** - **Aromatic compounds** - **Alkyl aryl ethers** - **Arenes** - **Benzoyl** -	- oxy- linked to aliphatic carbon chain - ethyl phenyl ether - -	- **AhR (activators)** - Troponin T cardiac - Disruptors of mitochondrial membrane potential - Glutamate receptor - Alpha-adrenergic receptor	- ER (antagonist) - ARE (antagonist) - Aldo-keto reductase - Carboxylic esterhydrolase - PPAR gamma
Custer 3	- Carbonyl compounds - Phenyl ketone - Amides - Five-member ring heterocycles	- oxy- linked to aliphatic carbon chain - alkyl halide - aliphatic amine - aliphatic amide	- **TR (antagonist)** - **Cell viability** - Alpha-adrenergic - NLRP3 inflammasome	- Cyclic phosphodiesterase - ARE (agonist)

Highlighted features in the table show 150 or higher values for degree centrality in non-toxic networks.

While hyperthyroidism activates reactive oxygen species (ROS) production, hypothyroidism is linked to the deficiency in antioxidant responses and altered lipid metabolism.[300] TRα-deficient mice have also shown higher apoptosis in pancreatic cells when stress in endoplasmic reticulum is induced.[301] AhR, which is a nuclear receptor activated by polycyclic hydrocarbons, regulates the expression of phase I and phase II metabolizing enzymes. AhR activation can either mediate oxidation or activates antioxidant responses, via Nuclear factor (erythroid-derived 2)-like 2 (Nrf2), depending on cell type or organ.[302] VDR is also known to maintain oxidative responses by inhibiting superoxide anion generation and regulating mitochondrial functions.[303] Therefore, the most connected nuclear receptors in both toxic and non-toxic network have high modulatory functions regulating oxidative stress.

Tables (5-4) and (5-5) also show estrogen (ER) and androgen (AR) receptors, which are related to chronic effects[304] as highly correlating with acute toxicity. These receptors as well as the nuclear receptors above, may not explain on their own how acute toxicity is triggered. As will be discussed later, nuclear receptors are involved in several synergistic interactions with other features leading to acute toxicity.

5.3.1.2 Antioxidant pathway

Our results highlighted three bioactivities involved in the antioxidant responsive element (ARE) pathway. These are ARE, Nuclear factor erythroid-derived 2 (Nrf2) and PPARγ in cluster 1 of toxic network and 1-3 clusters on non-toxic network (Tables (5-4) and (5-5)). ARE pathway is one of the first defence lines to protect the cell from ROS-induced DNA damage.[193] ROS activate the translocation if Nrf2 to the nucleus so as to bind to the ARE, resulting in the expression of detoxifying genes. It was also found that Nrf2 binds in the PPARγ promotor regions of ARE.[305] PPARγ regulates glucose metabolism and adepogenesis, and its disruption results in oxidative stress via an increase in lipid peroxidation.[193] Hence, targets of the antioxidant response pathway are key points in capturing the presence or absence of toxicity.

5.3.1.3 Kinase disruption

Kinase activity, which contributed mainly to cluster 2 and 3 in the toxic networks, as observed in Table (5-4), is associated with a number of adverse effects including cardiovascular, dermal and hepatotoxicities.[306,307] The well-studied drug doxorubicin is known for its severe cardiovascular effects which are attributed to the impairment of energy haemostasis, regulated

by kinases including AMPK (AMP-activated protein kinase) .[308] Doxorubicin was found to decrease AMPK levels, AMPK phosphorylation as well as the phosphorylation of its target Acetyl-CoA-carboxylase. [308] Another kinase was Ephrin receptor (EPHA2), which had the highest connectivity in the toxic network among kinases (187 node degree), is a subfamily of tyrosine kinases (Table (5-4)). Upon activation, EPHA2 inhibits neuronal repair after nervous system injury.[309] Ephrin receptors are overexpressed in a number of cancer cell lines, and responsible for cell adhesion and differentiation.[310] Chemical modulators of EPHA2 are cytotoxic and induce apoptosis in a number of cancer cell lines, attracting attention as a potential anticancer target.[310,311] The stress-activated protein kinase, JNK(c-Jun N-terminal kinase), which is activated by MAPK (mitogen-activated protein kinase), can mediate hepatotoxicity through impairing mitochondrial respiratory and increasing ROS release.[312] This, in turn, activates MAPK leading to further activate and sustain JNK pathway, which eventually leads to amplifying the toxic effects.[313] Inhibitors of tyrosine kinases, which were also observed in toxic networks, exert variable cardiovascular effects[306] such as ischemic heart diseases and vascular toxicity.[314] Phosphatidylinositol phosphate kinase (PIPK) contributed to cluster 3 in the toxic network, of which inhibitors are known to induce dermal and liver toxicities.[307] Calmodulin kinase (CaMK) is a multifunctional serine-threonine kinase which is activated when intracellular concentration of Ca ions increases.[315,316] The activation of CaMK induces myocardial dysfunction such as arrhythmias, which explains the toxic effects of cardiac glycosides.[317] The inhibition of CaMK is linked to neurotoxicity which was found to induce apoptosis in neurons and control neuronal structure and excitability.[315] Overall, kinase activity disruption plays a key role in describing acute toxicity.

5.3.1.4 Other targets

Toxic and non-toxic rule networks have shown significant connectivity of troponins, nitric oxide synthase (NOS) and NLRP3 inflammasome. Cardiac troponins are myofilament proteins that regulate cardiac muscle contraction and relaxation via binding to calcium ions.[318] Chemical compounds that increase or decrease calcium sensitization of myofilaments, via binding to troponins, can lead to changes in cardiac contractility.[319,320] Also mutation in cardiac troponin T was associated with a number of cardiomyopathies.[318,321] NOS is an enzyme that produces the vasodilator nitric oxide (NO), which is protective in ischemic events.[322] Activity against NOS is linked to cardiotoxicity as well as ROS production.[323] NLRP3 is a protein complex which activates a highly inflammatory cell death pathway, in response to non-microbial damage signals.[324] NLRP3 activates caspase 1, which is an IL-1β converting

enzymes, resulting in the release of a potent pro-inflammatory cytokine. This pathway is induced by reactive oxygen species (ROS) and environmental irritants. IL-1β mediates inflammatory responses that induce cardiac cell death linked to atherosclerosis, myocardial infarction and cardiac fibrosis.[324]

Other significant features in the networks include diamine oxidase (DAO), sirtulin-2 (SIR2) and tumor protein (p53). DAO is responsible for the degradation of a number of substances mainly histamine. The impaired breakdown of histamine can result in histamine accumulation which is itself cause acute toxicity at abnormal levels.[325] Symptoms of histamine toxicity include rash and hypotension.[326] SIR2 is a NAD-dependant deacetylase enzyme, a tumor suppressor which regulates DNA repair, apoptosis, neuroprotection and inflammation.[327] SIR2 deficiency is linked to mitotic cell death and gender-specific tumors.[328] P53 protein is a known biomarker in cancer [329] as it mediates several pathological pathways including doxorubicin-induced cardiotoxicity.[330] These findings, in general, demonstrate the high diversity in biological processes associated with acute toxicity.

Overall, the most important features captured in networks, which either have strong associations or high connectivity, may not explain alone the mechanisms initiating acute toxicity. Therefore, in order to further understand modes of action in acute toxicity, interactions between features should be analysed including the synergistic connections.

5.3.2 Novel insights into polypharmacology in acute toxicity through synergy interactions

High connectivity of some features discussed above demonstrates the importance of understanding toxicity with regards to compound polypharmacology. In order to determine which feature combinations are more important in acerbating toxic effects, we examined the statistical synergy of features pairs. This was performed using mutual information statistic and odds ratios on single features and compared with the pair combined (see Methods). A synergistic feature pair implies that the association of that pair with toxicity is significantly greater than individual features and their additive effect.

First, to understand why some feature pairs show synergistic interactions, we examined the similarity between connected pairs in terms of chemical profiles and biological functions, presented in Figure (5-6). As can be seen in Figure (5-6, A), synergy always occurred between features dissimilar in compound profiles, *i.e.* less than 0.2 and peak at 0.05. In other words, the

synergistic features often share small proportions of compounds in which both features are present. However, it is critical to note that not all dissimilar features, connected in the network, are synergistic (Figure (5-6, A)). Additionally, the similarity between target-based feature pairs was tested over KEGG biological pathways in Figure (5-6, B) as well as Gene Ontology (GO) molecular functions and GO biological processes (Appendix C, Figure (C-3)). It can be seen that the inverse relationship was reproduced, as highly synergistic bioactivity pairs had almost zero similarity in biological functions and pathways. This means that synergy resulted from combining biologically diverse activities, but not all biologically dissimilar activities are synergistic.
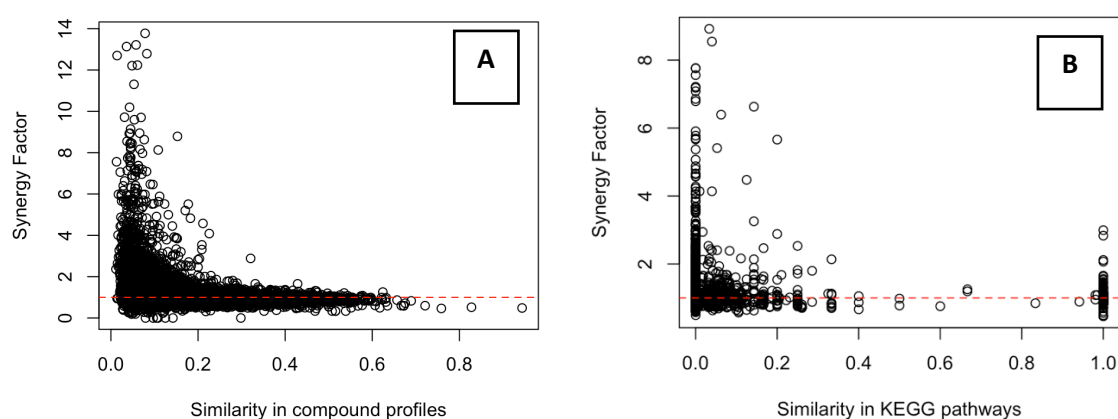
Figure (5-6) Similarity of connected feature pairs (co-occurred in rules) in compound profiles (A), and KEGG pathways of targets (B) against their corresponding synergy factor score (see methods). it can be seen that synergistic feature pairs (of which the synergy factor scores are higher than 1) are dissimilar in chemical profiles and biological function represented by pathways. However, not all dissimilar features are synergistic.

Next, we selected and analysed synergistic pairs in toxic networks which were present in two or more different rules, and summarized in Table (5-6).

5.3.2.1 Nuclear receptor disruption

Our results have shown co-clustering of cell viability assays with the disruption of number of nuclear receptors regulating proliferation such as estrogen and androgen receptors, in addition to nuclear receptors regulating expression of metabolizing enzymes. Proliferation and cell death is regulated, in general, by a network of nuclear receptors[331,332], which explain the observed co-clustering of these two classes. This combination was reproduced in assays predictive of hepatotoxicity, in Chapter 4. Although endocrine disruption is more associated with chronic effects such as carcinogenicity[304], the centrality of these effects in the network,

besides the relatively strong associations, indicate that compounds trigger acute toxicity are likely to produce chronic effects.

The antagonism of AR and ER were synergistic with the presence of six-member heterocycle substructures (Table (5-6)). Heterocycles steroids were reported to have an enhanced activity[333] and in some cases produce neurotoxicity and convulsions.[334] Synergy was observed between metabolizing enzyme Cytochrome 2C19 (CYP2C19) and a number of xenobiotic-sensing nuclear receptors such as VDR, constitutive androstane receptor (CAR) and AhR. In Chapter 4, rules describing hepatotoxicity combined the activity against CYP enzymes with VDR activity, an observation which was reproduced in rules describing acute toxicity. This synergy can be attributed to inhibiting clearance of toxic compounds via inhibiting the expression of xenobiotic metabolizing enzymes.[220,252,335] While AhR activation correlates with acute toxicity[132], the mechanism of AhR mediated acute toxicity was hypothesized to multiple routes. One study has demonstrated that AhR activation mediates the catabolism of vitamin D_3 via inducing Cytochrome P enzymes which may alter the protective effect of VDR signalling pathway.[336] This can be supported by the observed synergy between CYP enzymes, VDR and AhR in our study. Other studies have linked AhR activation with triggering inflammatory responses via inducing multiple cytokines[337], and promoting cellular morphological changes which correlate with activating JNK.[338] While specific synergies were observed for endocrine and xenobiotic-sensing receptors, nuclear receptors such as TR and VDR are involved in diverse synergistic interactions including neurotransmission and cholinergic toxidrome discussed below.

5.3.1.1 Neurosignalling disruption

Synergistic interactions were seen for the activity of two neuronal receptors, namely, glutamate and GABA receptors. The disruption of these receptors represents known key events in the AOP of acute toxicity.[128,129] However, when these stressors are combined with other specific bioactivities, the likelihood for *in vivo* toxicity is potentiated. Glutamate is an excitatory neurotransmitter and its activity is linked to neuronal damage in excessive excitation as well as inhibition.[284,285] There were multiple synergistic connections of glutamate receptor activity with TR, cytotoxicity and mitochondrial membrane potential. Studies have shown that toxicity is observed in chemicals which are characterized by mitochondrial toxicity and glutamate receptor activity.[339,340]

Table (5-6) Examples of important synergistic pairs in toxic network which occurred more than once in rules. Some of synergistic interactions in the table correspond to known key events (KE) in adverse outcome pathways (AOPs) of acute mortality, stored in the AOP wiki, namely Acetylcholinesterase (AOP ID 16), glutamate receptors (AOP ID 113 and 161) and GABA (AOP ID 10).

Key Event	Significantly associated Key Events	Odd Ratio comb	Synergy factor	N# toxic compounds	Comments
Cell viability (general)	Nuclear receptor activity (general)	2.8 ± 1.4 (Avr)	0.94 0 ± .32 (Avr)	58 ± 28 (Avr)	Cell proliferation and apoptosis are directly regulated by nuclear receptors[331,332]
AR antagonism ER antagonism	Six-member ring heterocycles	6.5 2.9	**3.8** 1.5	37 47	Heterocycles steroids have enhanced activity[333] and can produce neurotoxicity and convulsions[334]
Glutamate receptor	TR antagonism	3.7	**2.2**	46	Thyroid hormone (T3) activates glutamic neuronal reuptake. Mitochondrial toxicity potential toxicity of glutamate disruptors.[339,340]
	Cell viability	3.9	**2.0**	37	
	Disruption of mitomembrane potential	2.4	1.5	68	
GABA receptor	ARE agonist	21.7	**6.8**	21	TR and ROS control GABA reuptake.[341,342343] Vitamin D3 via VDR regulate GABA expression.[344]
	TR antagonism	22.8	**5.4**	21	
	Disruption of mitomembrane potential	23.8	**6.1**	22	
	HIF2	5.8	**2.5**	23	
	VDR	2.4	1.4	62	
Cyp2C19	CAR antagonism	1.4	1.0	327	CAR, VDR and AhR regulate the expression of Cytochrome P enzymes.[220,252,335]
	VDR antagonism	1.5	1.1	308	
	AhR activation	1.5	0.9	326	
AChE	Derivatives of carbamates	17.4	**4.9**	17	Cholinergic toxidrome involve Ca ion dysregulation and inflammation.[268] Interference with calcium sensitization of troponin[319,320] and inflammatory responses of NLRP3[324] are associated with cardiovascular effects. Depletion of PIP2 mediated the inhibition of ACh K^+ ion channels[345] via PI5P4K inhibition.
	Phophstidyl inositol 5 phosphate kinase	6.4	**2.6**	37	
	VDR antagonism	1.8	1.2	164	
	AhR activation	1.8	1.1	141	
	Troponin T cardiac	1.6	1.0	243	
	NLRP3	1.7	1.0	233	
NOS	Retinal dehydrogenase	2.1	1.3	99	VDR[346] and retinal dehydrogenase[347] activities can induce NOS expression.
	VDR	2.0	1.2	160	
	Alkyl halides	3.2	1.2	61	
α-adrenergic	Cell viability	9.2	**3.8**	18	PKC contribute in the α-adrenergic mediated contraction of vascular smooth muscles.[348,349]
	Aromatic primary and secondary amine	4.3	**3.2**	25	
	Protein kinase C (Tyrosine kinase)	4.1	**2.7**	16	
	AhR activator	2.8	1.4	40	

Additionally, thyroid hormone (T3) activates glutamate neuronal reuptake which results in protecting against toxicity triggered by glutamate excessive stimulation.[350] Therefore, we can conclude that glutamate receptor-mediated toxicity should not only be tested via activity against this receptor, but also combined with mitochondrial disruption, cytotoxicity as well as thyroid receptor disruption.

The excitatory neurotransmitter Glutamate is a precursor of the inhibitory transmitter GABA.[351] The activity against GABA had multiple synergistic links in the toxic network including TR antagonism, ARE, VDR and hypoxia-inducible factor-2 (HIF2). Similar to glutamate, TR regulates the synthesis, degradation, release and expression of GABA.[343]

There is clinical evidence that GABA related nervous disorders are linked to thyroid dysfunction.[343,352,353] Therefore, TR is highly connected because it regulates neurosignalling, including serotonergic, dopaminergic, glutamatergic and GABAergic networks,[350] in addition to oxidative stress and inflammatory processes.[300] Cellular oxidative stress, which induces the antioxidant response element pathway, reduces GABA reuptake and hence increases its levels,[341,342] which consequently interfere with normal neurotransmitter signalling. This explains the synergy between GABA modulators and activators of ARE. It was also reported that mitochondrial-derived ROS regulate postsynaptic GABA receptors in cerebral stellar cells.[354] Additionally, the deficiency of vitamin D3, the endogenous ligand of VDR, lowers the expression of genes regulating GABA neurotransmission,[344] and hence, explains the detected synergy between VDR and GABA activities. Hypoxia inducible factors (HIF-1 and HIF-2) are transcription factors that regulate haemostasis in response to low oxygen levels.[355] A possible association between HIF and GABA can be explained by the induction of GABA catabolism (GABA shunt) in hypoxic environment, which is an anaerobic route for energy production when the Krebs cycle is compensated.[356] The activation of the GABA catabolism produces succinate as a metabolite which activates HIF-1 and induces the inflammatory cytokine IL-1β.[357] HIFs also regulates glutamate signalling via upregulating glutamate transporters.[358] Overall, in order to capture neurotoxicity, a network of biological effects should be considered.

5.3.1.2 Cholinergic toxidrome

Cholinergic toxidrome represents a set of effects associated with AChE inhibition and linked with Ca ions dysregulation, immune and inflammatory responses and cytotoxicity.[268] The current analysis has shown that AChE involved in a diverse set of synergies (Table (5-6)) and was also ranked as one of the most connected nodes in toxic networks (Table (5-4)). AChE

have shown synergies with cardiac troponin, NLRP3 inflammasome and VDR, which are all involved in cardiovascular functions. AChE inhibitors also increase serum troponin[359,360], which is the most sensitive biomarker for myocardial infarction.[361] Although, levels of serum troponin may not be an indication for direct bioactivity against the protein[362,363], our study has shown synergy between AChE and troponin bioactivities. These findings support the possible cardiotoxic effect having these properties combined in compounds. Another bioactivity that synergistically connected to AChE and troponin was NLRP3 inflammasome (NACHT, LRR and PYD domains-containing protein 3). Inhibition of NLRP3 in models of ischemic injury reduces myocardial infarction and serum troponin.[364] The severity of toxification by organophosphates, which are potent AChE inhibitors, correlated with electrocardiogram (ECG) grading and cardiovascular effects such as ischemic changes and arrhythmias.[365] Therefore, combining activities against cardiac troponin, AChE and NLRP3 can induce cardiotoxic effects as well as potentiate the signs of cholinergic toxidrome. Another synergy interactions was observed between AChE activity and carbamate derivatives. The data has shown that around 50% of compounds possessing the carbamate substructure are toxic. Whereas, over 90% of carbamate derivatives in the dataset, and also are predicted to be AChE inhibitors, are toxic. Moreover, AChE has shown synergy with phosphatidylinositol-5-phosphate-4-kinase (PI5P4K), which participate in the synthesis of the second messenger phosphatidylinositol-4,5-diphosphate (PIP2).[366] It was reported that the depletion of PIP2 inhibits ACh K^+ ion channels.[345] Therefore, the observed synergy can result from the aggregated sensitization by ACh accumulation and PIP2 depletion leading to rapid onset of muscle paralysis and respiratory failure. Additionally, this study has shown synergy between AhR and AChE. A study has shown that 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) a potent AhR activator suppresses the activity of AChE through transcriptional down-regulation.[367] Overall, AChE is a key event in acute toxicity, which when present relevant bioactivity profiles should be screened so as to evaluate the possible *in vivo* effects.

5.3.1.3 Nitric oxide synthase activity

The rules have picked multiple isoforms of this enzyme, most notably the inducible and endothelial isoforms, iNOS and eNOS, respectively. NOS bioactivity was frequently and synergistically connected with VDR antagonism in rules for acute toxicity. VDR induces the expression of NOS via Vitamin D3.[346] This is supported by a study on VDR mutant mice, which had shown low levels of NO, reduced NOS expression, and associated with atrial stiffness and impaired heart function.[368] Our study has shown that VDR antagonism on its own

does not exert high mutual information with acute toxicity, however, it was one of the top bioactivities linked synergistically with diverse biological properties. This is likely due to the high regulatory function of VDR in haemostasis and cardiovascular function.[369] Moreover, NOS has shown synergy with alkyl halides and retinal dehydrogenase. Studies have reported that organochlorides induce the expression of NOS as well as ROS production in endothelial cells while decreasing NO levels.[370] Retinal dehydrogenase-1 (ALDH1) is an enzyme that oxidizes retinal into retinoic acid, and hence, regulating the metabolism of vitamin A.[371] ALDH1 metabolite, retinoic acid, increases the expression of NOS and suppresses T lymphocytes proliferation in liver dendritic cells to revert pathological activation of immune responses.[347] Also, ALDH1 is involved in the defence against oxidative stress.[372] Therefore, inhibiting ALDH1 may interfere with NOS-mediated immunological response.

5.3.1.4 α-Adrenergic receptor activity

Inhibition of α-adrenergic receptors can cause hypotension via decreasing constriction of vascular smooth muscles.[373] Synergy was observed between α-adrenergic receptors and protein kinase C (PKC), AhR and cytotoxicity (Table (5-6)). PKC contribute in the α-adrenergic mediated contraction of vascular smooth muscles.[348,349] α-adrenergic receptors also control the expression of PKC via the accumulation of downstream second messenger diacylglycerol (DAG).[374] Blocking PKC activity via gene deletion results in hypotension and decreased vascular contractility.[375] Therefore, the simultaneous blocking of α-adrenergic receptors and PKC can manifest in severe hypotension. Additionally, studies in AhR (-,-) mice have shown a higher expression of α1D-adrenergic receptors and an increase in the maximal effect in aorta contraction by α-receptors agonists; noradrenaline and phenylephrine.[133] Therefore, it is possible that AhR modulator would modulate basal α-adrenergic effects, which may explain the synergy between Ahr and α-receptors. Cytotoxicity also has shown synergy with α-adrenergic receptors, which may be a direct or an independent phenotype. The stimulation of α1A-receptors is associated with proliferative as well as antiproliferative effects.[376] Additionally, the analysis has demonstrated synergy between the presence of an aromatic amine substructure and α-adrenergic activity.

For the mechanistic interpretation of acute toxicity, our results indicate the importance of considering multiple bioactivities at a time. Single features on their own may not be strong

alerts for toxicity. Given the pairwise analysis conducted, the interpretation here may miss important interactions resulted from more than two properties at a time.

5.4 Conclusions

In this study, we investigated mechanistic interactions in acute toxicity from a feature space collected from structural properties, Tox21 assays and predicted targets. In order to understand complex interactions, we used emerging patterns which are generated for both the presence and absence of *in vivo* toxicity.

Rule patterns held associations of variable complexities as captured by single and multiple conditions, respectively. Chemical features had simpler and less complex interactions with acute toxicity. In contrast, target specific features involved complex and diverse connected conditions. The implications of complex bioactivity associations, observed in this study, may translate into complex predictive models for toxicity using bioactivity data. Hence, we anticipate that this complexity can explain the low performance of bioactivity-based models reported in the literature.[98,99,146] Also, understanding possible interactions, and hence significant polypharmacology is particularly important for weak assay activities.

Via rules, we gained novel insights into how specific polypharmacology profiles have higher odds of acute toxicity. For example, disruption of neurosignalling by GABA and glutamine receptors are more likely to trigger neurotoxicity when combined with thyroid receptor disruption. Another example is the cholinergic toxidrome which is characterized by inhibiting acetylcholinesterase AChE, and downstream effects including inflammation and Ca ion dysregulation.[268] we demonstrated that the likelihood of acute toxicity incidences is significantly greater when compounds inhibit with AChE also interfere with the calcium sensitization of troponin[319,320] and inflammatory responses of NLRP3[324], which are associated with cardiovascular effects.

In conclusion, the framework described in this study can be used to generate hypotheses about interacting key events in the adverse outcome pathway.

6   UNDERSTANDING PROMISCUITY BIAS OF *IN SILICO* MODELS PREDICTING ACUTE TOXICITY

6.1 Introduction

As discussed in earlier chapters, understanding the confounders and biases in *in vitro* data is essential for interpreting *in vitro* measurements. For example, overly activated pathways as a result of the cytotoxicity burst phenomenon can result in compounds inaccurately flagged as positives, *i.e.* false positives.[85] Consequently in these cases, and without consideration of factors deriving compounds' promiscuity, rationalization of *in vitro* measurements can be inaccurate. Additionally, it is likely that such confounders can result in spurious associations derived by statistical predictive models. It has been debated whether incorporating *in vitro* assay measurements as descriptors to predict toxicity is beneficial.[100] Mixed outcomes were observed as such leading to improving performance in some cases[146] and deteriorating in others.[99] This mean care should be taken when using bioactivity data to train statistical models for toxicity.

The aim of this study is to examine biases in predictive models that occur as a result of using bioactivity data. In a conformal prediction framework (see section 2.3), the predictivity of rule models, described in Chapter 5, will be investigated against Random Forest models trained on chemical descriptors (without bioactivity data). The distribution of predictions made by the models in the chemical space are compared with experimental toxicity labels to explore possible biases.

6.2 Materials and Methods

6.2.1 Datasets

The dataset used in this chapter was adapted from Chapter 5, including Tox21 assays, predicted targets, substructures and GHS acute toxicity labels (see section 5.1.1). The data represents 2000 compounds, 993 toxic compounds and 1007 non-toxic, against 3732 features. These represent 1789 predicted targets, 440 Tox21 assay activities, 1259 ToxAlert toxicophores and 242 MoSS derived substructures. Additionally, using the chemical structures of the 2000 compounds, chemical descriptors were calculated using RDKit[199] node in KNIME[200] to generate 117 physicochemical properties as well as circular fingerprints with a radius of 2 and 2048-bit length.

6.2.2 Emerging patterns generation

Emerging patterns were generated using CPAR algorithms following the data and specification described in Chapter 5 (see section 5.1.2).

6.2.3 Random Forest models

Classification models were built based on Random Forest algorithm, which was trained on chemical descriptors (physicochemical properties and circular fingerprints) for 2000 compounds. The model was generated in R (version 3.3.3) using the package 'RandomForest' setting the number of trees to 500, while all other parameters are set to default.

6.2.4 Conformal prediction

A conformal prediction framework (Chapter 2, section 2.3) was applied to evaluate the reliability of predictions made by rules and represented in Figure (6-1). In the conformal prediction framework, the data are split into proper training and test sets in a 4: 1 ratio (see Figure (6-1)). The proper training set is then split into training and calibration sets at a 4:1 ratio. The training set is used to train the models, whereas predictions are made for calibration and test sets. A key component of the conformal framework is the non-conformity measure (α), which is used to estimate how a test set compares to an independent set, *i.e.* the calibration set.

With regards to emerging patterns, each test (or calibration) compound can satisfy i number of toxic describing emerging patterns, where $i \in \{0, 1, 2, 3 ,.... , m\}$, and m is the total number of toxic describing patterns. Similarly, compounds can satisfy j number of non-toxic describing emerging patterns, where $j \in \{0, 1, 2, 3 ,.... , n\}$, and n is the total number of non-toxic describing patterns. The non-conformity score is the difference in compound promiscuity over emerging patterns describing toxicity and non- toxicity, called the rule promiscuity score (see equation below).

Rule promiscuity score
= N# satisfied toxicity describing patterns – N# satisfied non_toxicity describing patterns
$= i - j$ [13]

In order to examine if promiscuity over rules outperform promiscuity over important features, we also used promiscuity over significant features as a non conformity score. Significant features were determined through univariate associations with acute toxicity (Fischer test p-value < 0.01 on features used to train rule models).

As for Random Forest models, the probability values were used as the non-conformity score (see Figure (6-1)). The non-conformity score is calculated for the calibration set and test sets. Then the rank of the test set among the calibration set determines the significance (p-value). In a Mondrian conformal models, non-conformity scores are calculated for both labels (α_{toxic} and $\beta_{non\text{-}toxic}$), and hence, two independent ranks are produced for each test set in a two-class label setup. The non-conformity scores for the non-toxic label were obtained as follows for the emerging patterns models:

$$\beta non\_toxic = -[Rule\ promiscuity\ score] = -[\alpha] \qquad [14]$$

And for the Random forest model:

$$\beta non\_toxic = 1 - [P] = 1 - [\alpha] \qquad [15]$$

According to the p-values obtained from non-conformity score rankings and a predefined confidence levels, four labels can be assigned by the conformal models, namely, 'Toxic', 'Non-toxic', Both' or 'None'. The process to derive these predictions is explained in section 2.3.

Conformal models generation and predictions were conducted in R[176] environment (version 3.3.2).

6.2.5 Model validation

We applied 5-fold cross-validation for both models (see data split in Figure (6-1). Within each fold, another 5-fold cross conformal was applied on the training set as described by Sun *et al.*[377] The training data were split randomly into proper training and calibration sets in a 4:1 ratio 5 times against one test set. Therefore, each test instance had 5 P-value scores, which were then averaged. The averaged P-value was used to make predictions with respect to the confidence level used.

6.2.6 Chemical space analysis

In order to compare the distribution of experimental toxicity labels against predicted labels, the compounds were visualized in chemical space and labelled accordingly. Circular fingerprints of compounds were and plotted into two dimensions using multidimensional scaling analysis (MDS). MDS was conducted in R using 'cmdscale' function on the Euclidean distances of fingerprint matrix. The generated weights (eigenvalues) for each compound against the two compressed dimensions represented the scales in the 2D plot.

6.2.7 Cluster analysis

To detect which chemical and biological properties were driving the coefficients in the MDS 2D plot, hierarchical clustering was applied to these coefficients against physicochemical properties and compounds promiscuity. The similarity was calculated using the Pearson correlation statistic and the similarity matrix was visualized in heatmaps using 'heatmap.2' function of 'ggplot2'[205] package in R[176] environment (version 3.3.3) at default settings.
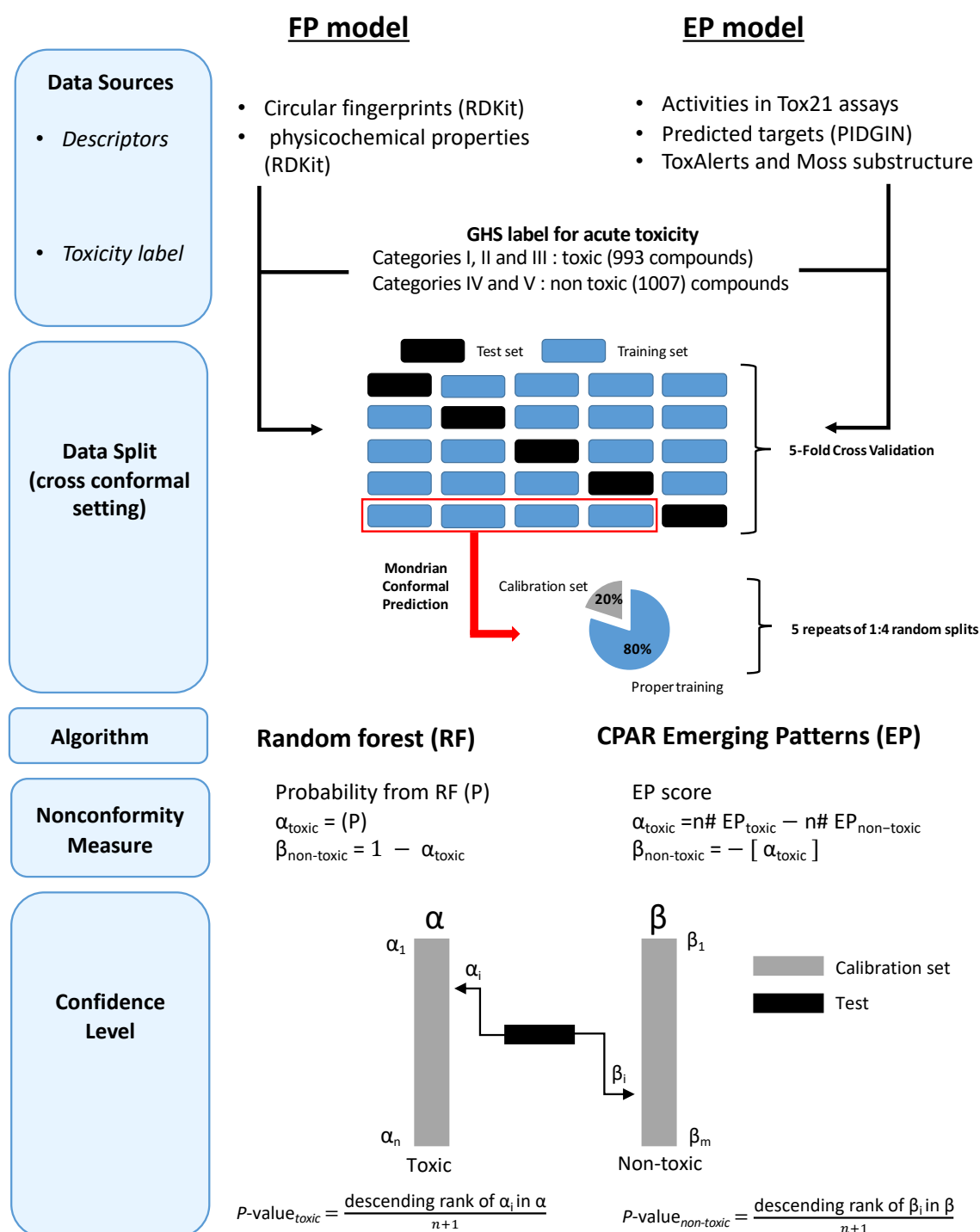
Figure (6-1) Cross-conformal workflow for chemical fingerprint (FP) models and emerging patterns (EP) model. FP model used circular descriptors and physicochemical properties as input, whereas, rule models use bioactivities from Tox21 measurements, predicted targets in addition to chemical substructures. The toxicity label was generated from GHS classification grading for acute toxicity. The data split involved 5-fold cross-validation of training and test set. At each fold, the training data were partitioned into calibration and proper training sets in 5 random repeats of 1:4 splits. The proper training was used to build the predicted models via random forest in the FP model and CPAR-based emerging patterns for EP model. The nonconformity scores (NMC) used to rank compounds were label probability and rule promiscuity score for FP and EP models, respectively (see Methods). According to the descending ranks of test data points in the calibration sets for each label class, the significant level (p-value) is computed of which confidence is estimated.

6.3 Results and Discussions

6.3.1 Conformal Predictions

Three measures were used to examine the performance of the conformal model, namely, validity and efficiency (see Section 2.3) and correct classification rate. Firstly, the validity was tested across multiple confidence thresholds. We compared three sets of models generated from i) conventional Random Forest model using calculated circular fingerprints and physiochemical properties, ii) emerging patterns model and iii) model using the promiscuity over statistically significant features. Figures (6-2) shows that the error rates align with the corresponding confidence levels for all models. This means that the models were capable of producing valid predictions ('Toxic', 'Non-toxic' and 'Both'). While the overall validity trends aligned well with the identity line, a slight fluctuation can be seen for the validity of individual class labels due to statistical noise.[378]
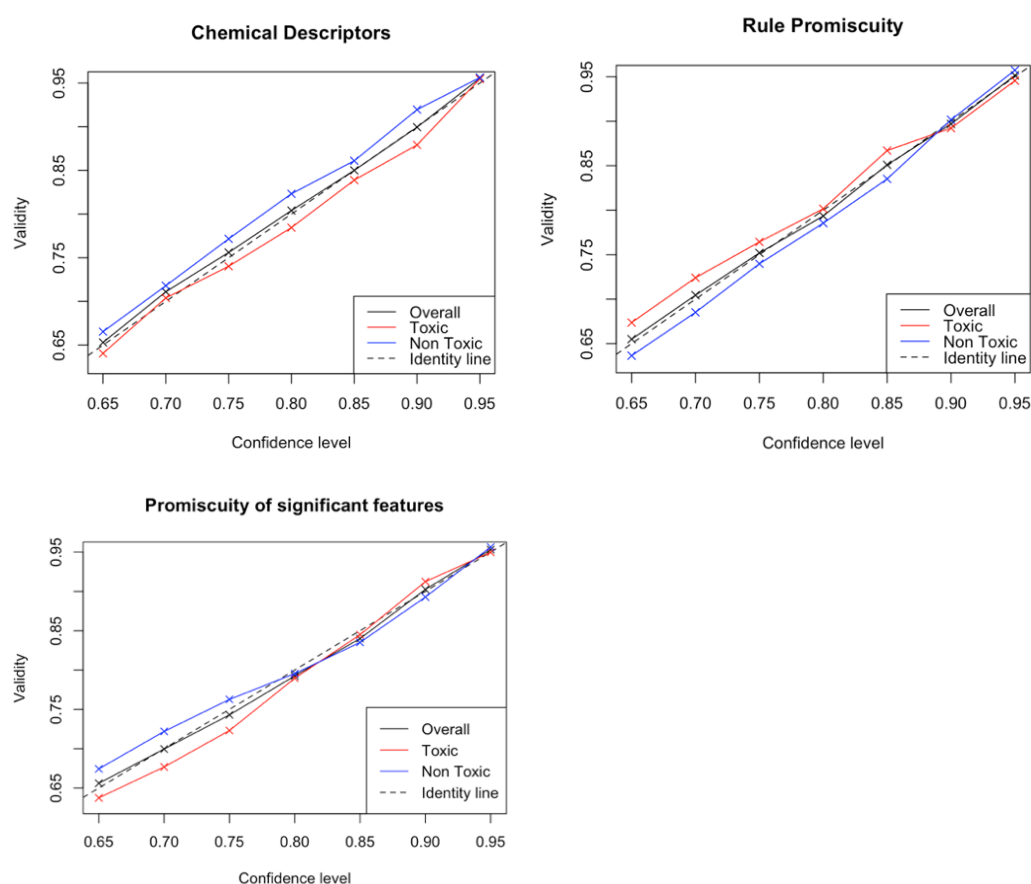
Figure (6-2) Validity of conformal models. The validity (conformal accuracy) reflects the percentage of valid predictions ('toxic', 'non-toxic' and 'both') to all predictions. A correct prediction in the conformal model setting is given to classes where the label is classified correctly as well as points predicted to satisfy both classes. The plots show that all models generated valid predictions at the given confidence levels (percentage of valid predictions is equal to the predefined confidence level).

The second step in assessing the performance of conformal models is to examine the percentages of single predictions made by each model in a given confidence level, which reflects the models' efficiency. Table (6-1) represents the percentages of single predictions made by each model at three confidence thresholds, namely 90%, 80% and 70%. As a general trend, the percentage of single predictions increased as the confidence level decreased. This means that the model was allowed to assign more single labels to new instances when the expected error rate was higher. In other words, at higher degrees of confidence, the model can only assign single labels to fewer instances. The higher the number of single predictions, the more efficient the model. Chemical descriptors and rule promiscuity models had shown the highest efficiency up to around 90% and 82% at 70% compared to significant feature promiscuity with 64% efficiency at the same level. At the higher levels of confidence, namely 80% and 90%, the difference in efficiency by chemical descriptors was about 2 folds higher compared to feature promiscuity (absolute differences of 24% and 21%, respectively) , whereas rule promiscuity is 50% higher than feature promiscuity (absolute differences of 7% and 11%, respectively).

Table (6-1) Percentages of single class predictions, namely 'toxic' and 'non-toxic' labels.

	Confidence level		
	90%	**80%**	**70%**
Calculated Descriptors	42.75	67.1	90.4
Rule Promiscuity	31.55	60.35	81.85
Feature Promiscuity	21.5	43.3	64.3

Next, the distribution of predicted labels generated by the models across the three confidence levels are presented in Figure (6-3). It can be observed that feature promiscuity-based models produced the highest proportion of instances predicted as 'Both'. This means that the latter had the least ability to discriminate between classes as compared to the other two. Moreover, chemical descriptors and rule promiscuity models were almost similar in the percentages of Toxic class predictions especially at 80 and 70% confidence, which was equivalent to above 30% and 40%, respectively, of all predictions. Given that rule and feature promiscuity-based models arise from similar concepts, the former significantly outperformed the latter in efficiency.
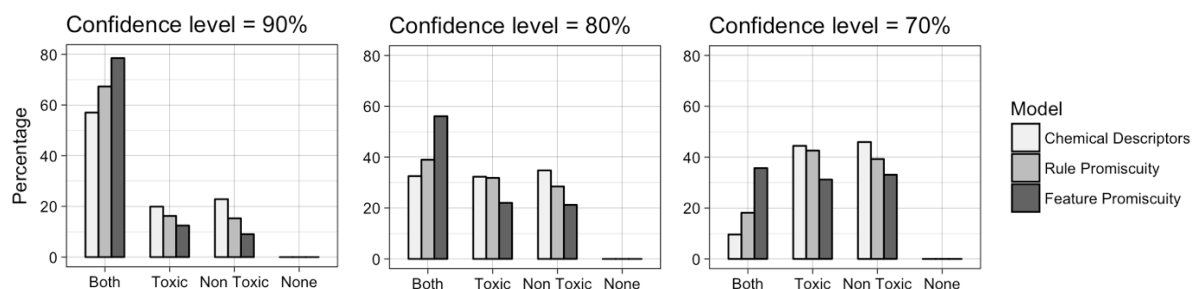
Figure (6-3) The percentage of predicted/assigned class labels at confidence levels of 90%, 80% and 70%. At the level of 90% confidence, the majority of predictions were classified as 'both'. As the threshold for confidence decreases, the models produce more single prediction labels, namely 'toxic' and 'non-toxic'. In general, chemical descriptor models (FP) always generated higher numbers of single label predictions, whereas models based on feature promiscuity had the least number of single class predictions across the confidence levels.

The third step in performance is measuring the correct classification rate of the single predictions, which are reported in Table (6-2). The table shows that both chemical descriptors and rule promiscuity, at 90% confidence had an overall accuracy higher than 70%. However, the correct classification rate of feature promiscuity models was almost random. This means that using the absolute promiscuity over significant features was less useful in discriminating between toxic and non-toxic compounds. Overall, the accuracy deteriorated as the confidence level decreases, where higher error rates are expected. This decrease was steeper for chemical descriptors models and rule-based models. The accuracy was also higher in predicting the toxic compounds across all models, especially at higher confidence. For example, at 90% confidence, the accuracies of the toxic class predictions by chemical descriptors and rule models were around 80% and 73%, respectively. Lower accuracy can be seen for classifying non-toxic compounds at the same threshold to around 75% and 69% for chemical descriptor and rule-based models, respectively. The class-wise variability in performance diminished as the confidence decrease, so that the accuracy of predicting both classes is almost the same at 70% confidence for chemical descriptor models and in favour of non-toxic compounds using rule models. In average and across all threshold levels, rule models are only 5% less accurate than chemical descriptor models. In further analysis, only rule and Random Forest models are examined due to the low performance of feature promiscuity models.

Table (6-2) Correct classification rate of single predictions (FP: chemical descriptors, EP: emerging patterns/ rule models, SP: promiscuity over significant features). As the confidence level decreases, the accuracy of single class predictions decreases. Overall, FP models had shown the highest accuracies. The difference between the EP model and FP model accuracies was consistent around 5%. Using target promiscuity as a nonconformity sore in the models had almost random performance.

	Confidence level								
	90%			**80%**			**70%**		
	FP	*EP*	*SP*	*FP*	*EP*	*SP*	*FP*	*EP*	*SP*
Overall	77.07	70.99	54.65	71.31	66.86	53.46	68.03	63.89	53.26
Toxic	79.64	72.92	56.62	72.76	67.34	53.96	68.05	62.79	55.13
Non toxic	74.84	68.95	51.93	69.97	66.31	52.94	68.01	65.09	51.51

6.3.2 Distribution of model predictions across chemical space

In order to visualize predictions made by Random Forest (FP) and rule models in chemical space, 2D multidimensional scaling analysis (MDS) plots were generated using circular fingerprint descriptors. As a reference, the toxicity labels from the GHS classification were also visualized in Figure (6-4). As can be observed in Figure (6-4, a), the distribution of toxic and non-toxic labels, from GHS data, in the chemical space is almost random especially across the second dimension in the MDS plot (D2). At the negative scale of the horizontal D1 axis, the ratio of two labels is in favour of toxic compounds, which is also captured in the corresponding density plot. With regards to predictive models (Figure (6-4, b and c)), a clear pattern can be seen in the distribution of predicted toxic and non-toxic classes in the chemical space. For example, across the negative scale of D1, the imbalance between the two labels is greater than GHS toxicity labels. Also, along D2, there is higher density of toxic compounds at the positive scale (precisely at ranges higher than 0.5). The majority of compounds at the negative scale of D2 were predicted as non-toxic. The imbalance in distribution and segregation in predicted labels across D2 is more pronounced in rule models (Figure (6-4, c)). Therefore, models tend to amplify trends observed in the original data at variable extremes and depending on the accuracy of the model.
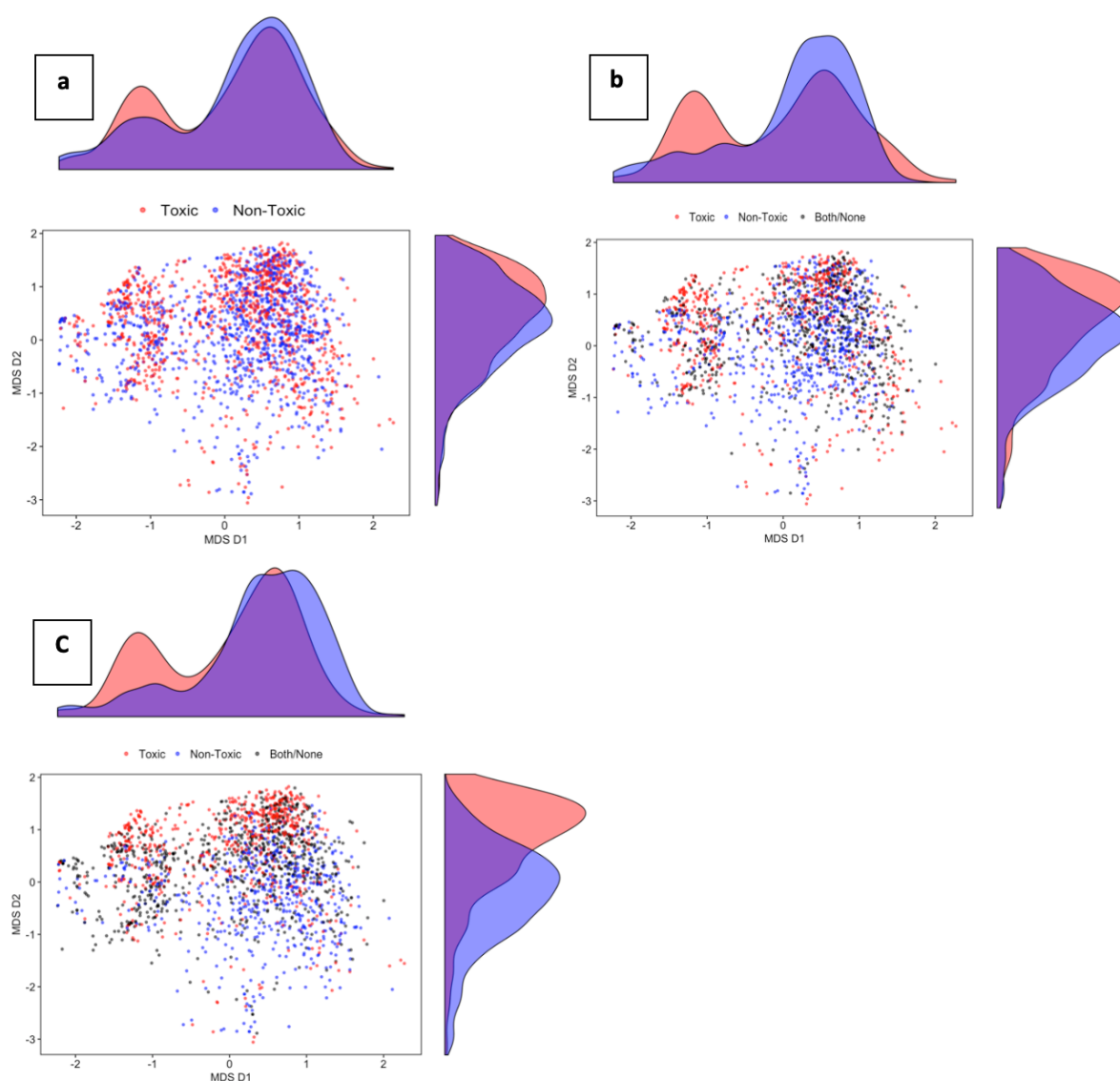
Figure (6-4) Visualization of the chemical space labelled by toxicity classes (a), chemical descriptors (FP) model predictions (b) and emerging patterns (EP) model predictions (c) in multidimensional scaling (MDS) plot using circular fingerprints. Toxicity labels are red and non-toxicity labels are blue. Compounds predicted to be either 'both' or 'none' are labelled black. The density plot at each axis shows the distribution of toxic and non-toxic label with regards to dimension 1 in the x-axis of the MDS and dimension 2 in the y-axis. Toxicity labels in the data had random distribution in the chemical space (a), however, both models, FP and EP, had unbalance distributions of the predicted labels, especially across the y-axis (D2) in b and c. Compounds predicted as toxic populated more at the positive scale of the y-axis whereas non-toxic predictions were denser at scales <0.5 in the y-axis. This is also reflected in the density plots of the y-axes, as the distributions of predicted classes are more separated than the actual toxicity labels. Abbreviations: MDS D1, dimension 1 in chemical space plot; MDS D2, dimension 2 in chemical space plot.

In order to interpret the features deriving the bias in the model, we investigated how the coefficients of D1 and D2 correlate with physicochemical properties as well as promiscuity over bioactivity and structural features. The pairwise Pearson correlations were calculated and visualized as a square hierarchical clustering heatmap in Figure (6-5). D1 in chemical space is highly correlated with the number of aromatic rings (0.75 Pearson coefficient). The majority of compounds clustered at the negative range of D1 have no rings (results not shown). D2, on the other hand, had shown correlations with promiscuity over significant bioactivity features at 0.65 Pearson coefficient. Also, D2 had negative correlations with the number of atoms and promiscuity over structural features at -0.67 and -0.6 correlation magnitudes, respectively. This means that compounds of high D2 coefficients are associated with high promiscuity over targets whereas low D2 coefficients indicate that compounds have a higher number of atoms as well as high diversity over structural alerts. SLogP lipophilicity is a property frequently linked with promiscuity over targets. However, this was not observed in our study, nor with D2 coefficients. One interpretation for this is that the bioactivity in our dataset was derived from cell-based assays and hence permeability can be a key condition for intracellular activity. We found that SLogP values of the most promiscuous compound peaked around 3 (results not shown), which was reported for good permeability.[195]

D2 coefficients, as well as promiscuity over bioactivities, also negatively correlated with the number of saturated carbons and number of oxygen atoms (-0.69 and -0.62, respectively). The trends seen here aligns with previous studies that reported an association between compound promiscuity and the fraction of sp3 hybridised carbons to the total number of carbons.[379] Promiscuity also negatively correlated with the number of oxygen atoms.[42] Although the presence of aromatic amines in the chemical structure is associated with promiscuity[43], the number of rings and number of nitrogen atoms, on their own, did not show here significant correlations with promiscuity here. Promiscuity over significant bioactivities had shown a negative correlation with molecular weight and promiscuity over alerts at –0.51 and –0.37 Pearson coefficients, respectively. The low promiscuity of structurally complex compounds was previously reported[380] and can be attributed to the higher specificity in interactions with biological targets.[381] Overall, we can conclude that the observed bias in the predicted labels across chemical space (in Figure (6-4)) is greatly associated with compounds promiscuity over bioactivities as well as chemical features driving this promiscuity.
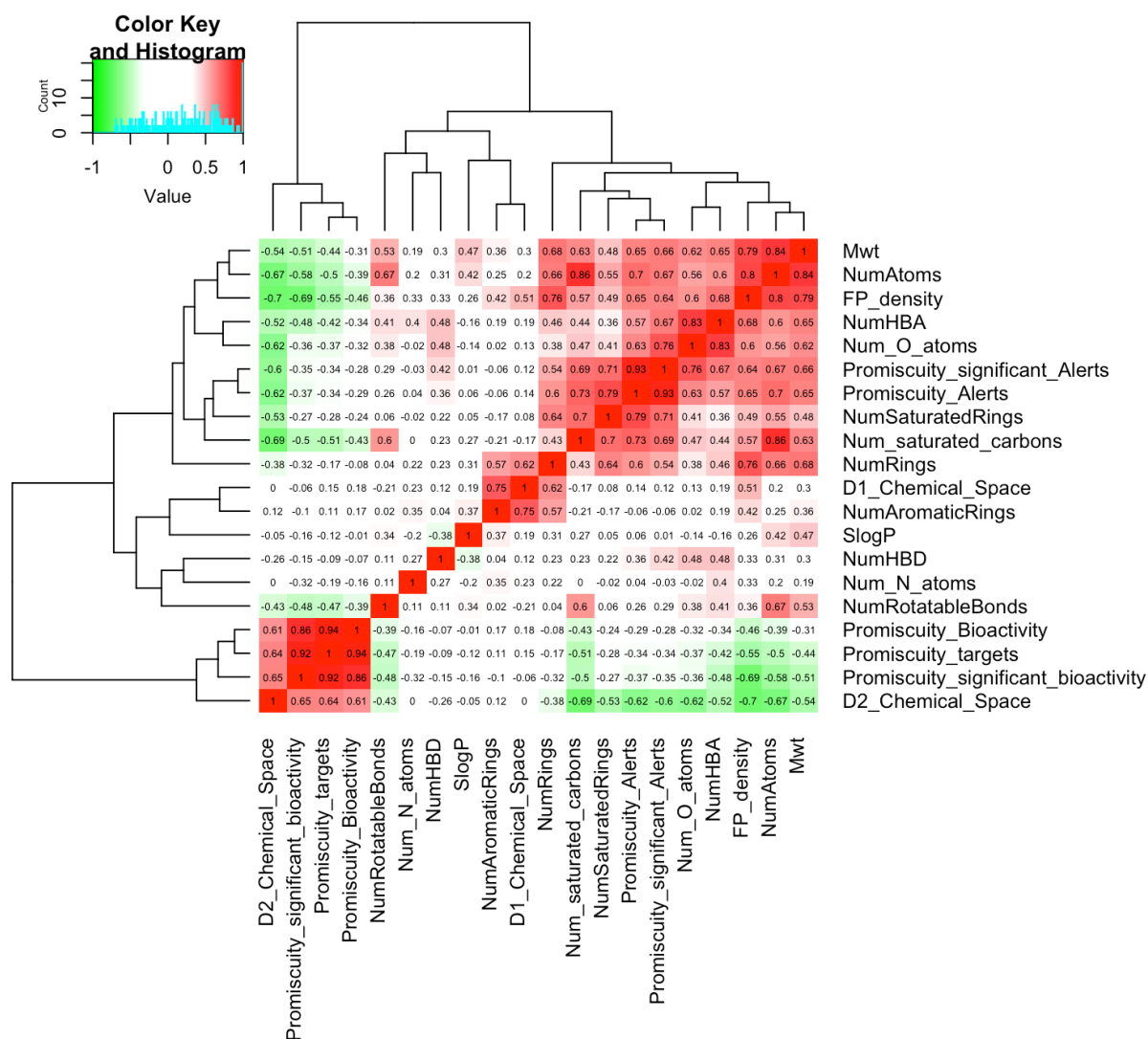
Figure (6-5) Heatmap plot for the cluster analysis of the MDS weights from Figure (6-4) against compounds promiscuity and physicochemical properties. Pairwise Pearson correlation was calculated between variables, and used as distances to perform hierarchical clustering. The second dimension D2 positively correlated with promiscuity over bioactivities but negatively correlated with the count of unique substructures. Promiscuity over bioactivity features negatively correlated with the number of saturated carbons and number of atoms, reproducing findings in literature.[379] Abbreviations: D1_Chemical_Space, dimension 1 in chemical space plot; D2_Chemical_Space, dimension 2 in chemical space plot; FP_density, Count of present bits in chemical fingerprint; Mwt, molecular weight; Num_N_atoms, number of nitrogen atoms; Num_O_atoms, number of oxygen atoms; Num_saturated_carbons, number of saturated carbons; NumAromaticRings, number of aromatic rings; NumAtoms, total number of atoms; NumHBA, number of hydrogen bond acceptors; NumHBD, number of hydrogen bond donors; NumRings, number of rings; NumRotatableBonds, number of rotatable bonds; NumSaturatedRings, number of saturated rings; Promiscuity_Alerts, total number of substructural features; Promiscuity_Bioactivity, total number of positive activities in Tox21 assays and predicted targets; Promiscuity_significant_Alerts, total number of significant substructural features (Fischer < 0.05 with acute toxicity label); Promiscuity_significant_bioactivity, total number of positive activities in significant Tox21 assays and predicted targets (Fischer < 0.05 with acute toxicity label); Promiscuity_targets total number of positive activities in predicted targets.

6.3.3 Promiscuity biased predictions

The models' bias over promiscuity was assessed by measuring the performance (the number of single predictions and accuracy) across three promiscuity intervals, namely high, moderate and low. The boundaries of these intervals were determined by discretising the ranked promiscuity values into equal sized groups. Figure (6-6) compares the number of single label predictions with the distribution of GHS toxicity classes across promiscuity intervals. Both Random Forest (FP) and rule-based models produced significantly higher fractions of toxic predictions compared to non-toxic predictions for highly promiscuous compounds. The differences in prediction labels were 7 fold and 2 fold difference for rule and FP models respectively. Yet, only 15% difference in the distribution was observed for GHS toxicity labels at the high promiscuity level. In contrast, at low promiscuity, there were significantly larger numbers of non-toxic predictions produced by rule and FP models to almost 6 and 2 folds, respectively, compared to toxic label predictions. This means that both models over-predicted high promiscuous compounds as toxic and low promiscuous compounds as non-toxic, although the maximum variability in class proportions is not exceeding 30%, at any promiscuity level. At the moderate promiscuity level, both models produced balanced counts of predicted toxic and non-toxic labels. FP models, however, generated around 55 more single predictions for each class than rule models at the same promiscuity level.
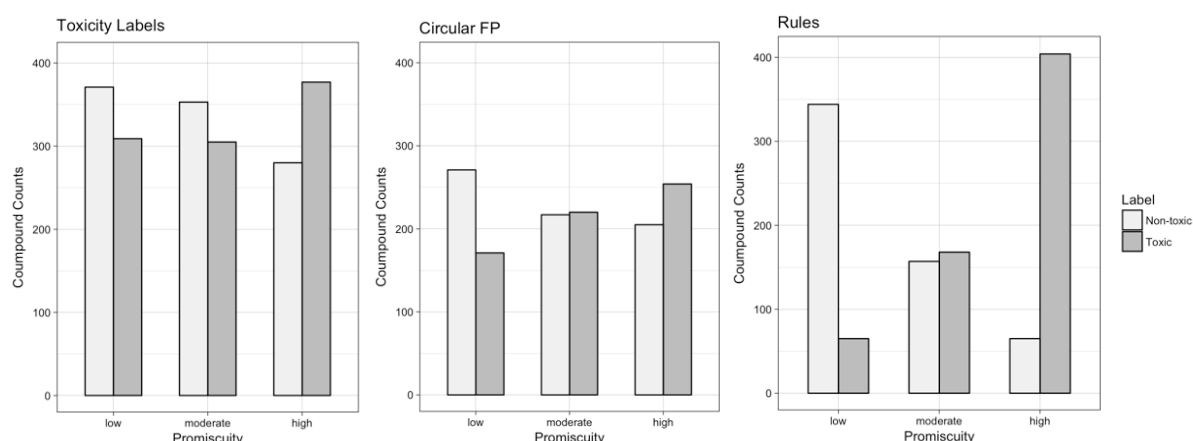
Figure (6-6) The distribution of toxicity labels (a) and single class predictions by FP model (b) and EP model (c) against three levels of promiscuity. The variability in toxicity label proportions in the data did not exceed 30% at any promiscuity level. With regards to predictions, the imbalance in class labels was in favour of non-toxic at low promiscuity levels and toxic at high promiscuity levels. EP models showed the highest bias to the level of compound promiscuity, as this model used predicted targets as input data and promiscuity score as the predicting function.

Next, we quantified the ability of the models to generate correct predictions for toxic and nontoxic labels across promiscuity levels, which is presented in Figure (6-7). We found that

the promiscuity level has also reflected on the true positives rates (TPR) and true negatives rates (TNR). For low promiscuity compounds, FP model gave the most accurate single labels predictions at an average of 75%. Rules also correctly assigned 77% of predicted toxic labels for low promiscuity compounds, and below 65% correct predictions of single non-toxic labels. Whereas, for high promiscuity compounds, FP model had over 70% TPR but less than 65% TNR. Rules had shown higher TNR for high promiscuous compounds than FP models, exceeding 70%. At moderate promiscuity levels, both models had similar rates of correct classification of almost 70% for both classes wherever a single class prediction was made.
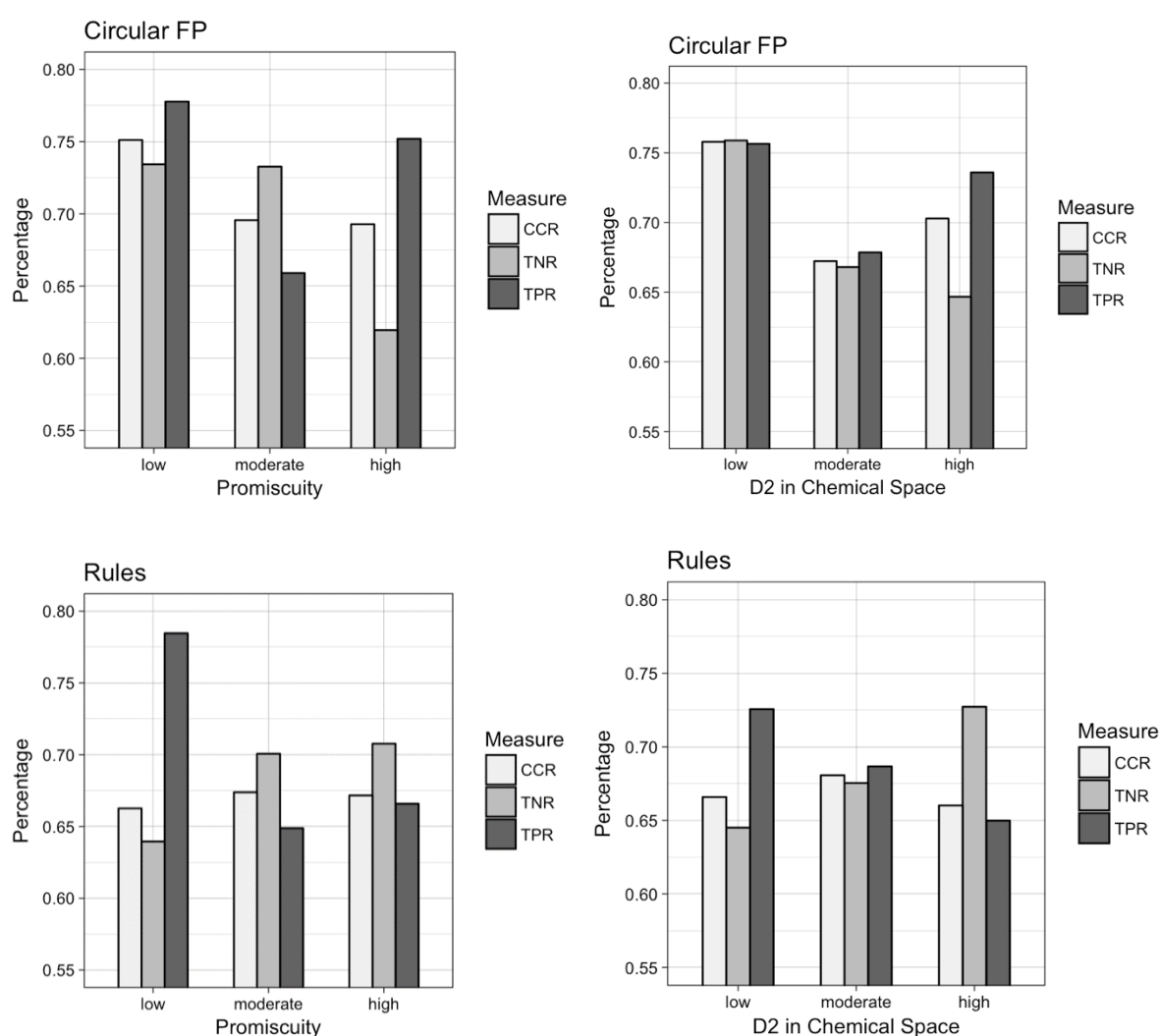
Figure (6-7) Accuracy of single class predictions across different levels of D2 in chemical space and promiscuity, measured as the true positive rate (TPR), true negative rate (TNR) and correct classification rate (CCR). FP model generated the most accurate single class prediction for low promiscuous compounds. The accuracy of single class predictions of both models was almost similar for compounds of moderate promiscuity ranging between 65% and 70%. However, whenever a non-

toxic label was generated by EP model for a high promiscuous compound, there are greater chances of correct predictions compared to FP models, with a difference of TNR of about 6%.

The factors deriving this promiscuity based biased can be attributed to the nature of the model and data used. Rule models used predicted target annotations, and hence, the bioactivity profiles of compounds may not be 100% accurate. Also, rules models predicted toxicity labels as a function of promiscuity over rules. For example, compounds that satisfy a higher number of rules describing toxicity are likely to be predicted as toxic. Hence, it is likely that promiscuous compounds will satisfy a greater number of rules even if they were non-toxic. Another factor that can affect performance is the lack of physicochemical properties in the data used to generate the rules. While perturbing multiple pathways has a greater likelihood of observing *in vivo* toxicity compared to no perturbation, rapid onset is also a key factor. Cell permeability and bioavailability, which are governed by physicochemical properties, contribute to whether an acute effect will be observed.

FP models, on the other hand, did not use target-based features but still showed bias over promiscuity. Chemical features can be associated with compound promiscuity over targets if the features represent a pharmacophore for target binding. Hence the chemical feature, in this case, substituted target activity. Additionally, the bias of FP models was greater over specific chemical features (captured by D2 in chemical space) than target promiscuity. The FP models performed overall better for compounds with low (negative) coefficients in D2 of chemical space (Figure (6-7)). D2 in chemical space also strongly correlated with fingerprint density, *i.e.* 0.7 Pearson coefficient (see Figure (6-5)). Since the model learning is affected by data sparsity, complex compounds are denser in the fingerprint space allowing for better predictivity.

6.4 Conclusions

Conformal prediction models provide an assessment of confidence for a given prediction. We used the conformal prediction framework to evaluate the performance of rules used in Chapter 5, given their interpretability. For a reference, we compared the performance of these rules with two models i) Random Forest algorithm on circular fingerprints and physicochemical properties, and ii) models based on promiscuity over statistically significant features. The overall performance of rule models outperformed promiscuity over features and showed an average accuracy 5% lower than the benchmark Random Forest model. Given the interpretability of rule models in comparison to Random Forest, rules can aid data-based

decision making by providing mechanistic insights about predicted toxicity. The results also show that the mere application of promiscuity in toxicity assessment should be interpreted with caution, aligning with previous reports.[382] Specific feature combinations capturing synergistic interactions, as shown in Chapter 5, explain why promiscuity over rules outperforms promiscuity over significant features.

Upon investigating the models' performances with regards to the chemical space, we found a clear bias in performance and in association with compound promiscuity. Rule models displayed larger biases due to two key factors. First, the promiscuity is part of the prediction formula, and hence the higher the promiscuity, the higher the likelihood that a compound will be classified as toxic. Second, unlike chemical descriptor models, physicochemical properties were not used, resulting in missing kinetics-based factors. Nonetheless, the Random Forest model on chemical descriptors has also shown the promiscuity bias trend. We anticipate that the promiscuity bias, in this case, is indirect and captured by the structural complexity, a property directly linked to promiscuity. Overall, in order to improve the predictive performance of rules, it is recommended to incorporate physicochemical properties, and also to optimize the prediction algorithm. For example, as an alternative for rule promiscuity, predictions can be made using the best rule or averaging the performances of the best k-number of rules for each class label.

Variability was seen between the true positive rates and true negative at the same space of chemistry or promiscuity. Therefore, assigning a correct prediction within a given chemical space is influenced by the class type. We, hence, recommend that measures of confidence and applicability domain should be class specific.

The results here report compounds promiscuity as a confounding factor for classifying *in vivo* toxicity. Irrespective to data type used, models often predict promiscuous compounds as toxic. This promiscuity is governed by the chemical structure and physicochemical properties, such as number of oxygen atoms and number of saturated carbons. Therefore, we recommend that compounds' intrinsic promiscuity should be considered when rationalizing outcomes from statistical models as well as from *in vitro* assays.

CONCLUSION

Toxicity mode of action analysis is a key component of the integrated approaches in testing and assessment (IATA). Repositories of *in vitro* bioactivity measurements such as Tox21[53] and ToxCast[52] , combined with advancement in computational power allowed for the generation of statistical models for *in vivo* toxicity[52,53], as well as the derivation of hypotheses on possible modes of action. However, perturbation mechanisms at the molecular level can be highly complex due to multiple routes for downstream effects as a result of interfering with a single pathway. Additionally, compounds rarely have a single biological action and are rather characterized by polypharmacology profiles rendering the task even harder. The adverse outcome pathway (AOP) framework describes adverse events as directional series of key events across the biological systems. This directional association is widely held when interpreting mechanisms of toxicity, however, its application in modeling toxicological data is rare.[144] To tackle this, we used rule-based models which allow mining associations between bioactivity and chemical conditions against toxicity. In order to improve the interpretability of rules, we applied constraints on the direction of the expected associations between *in vitro* activity and *in vivo* toxicity. First, rules were used to prioritize assays for hepatotoxicity *in vitro* models using ToxCast and data from animal studies (Chapter 4). We detected known endpoints such as activity against Cytochrome P, immunological responses and mitochondrial toxicity. Further, we found that endocrine disruption, such as modulation of estrogen and androgen receptors, are key bioactivities in hepatotoxicity. These were missed from four commercial *in vitro* models for hepatotoxicity. Second, we explored the role of chemical properties in the improving the translatability of *in vitro* readouts into *in vivo* hepatotoxicity. A specific set of physicochemical properties, such as number of rotatable bonds showed improved odds of hepatotoxicity when combined with *in vitro* endocrine disruption (Chapter 4). These physicochemical conditions are related to kinetic parameters such permeability,[196] and hence can act as proxies for *in vivo* exposure measures. Despite endocrine disruption being commonly associated with chronic effects such as reproductive toxicity, we showed that endocrine disrupting compounds bearing a heterocycle ring can trigger acute toxicity (Chapter 5). Third, rules also facilitate gaining novel insights into the influence of polypharmacology on toxicity. We found that not mere promiscuity is triggering toxicity, but specific combinations of activities which can be synergistic and lead to toxicity. For example, thyroid hormone receptor modulation, when combined with the disruption of ionotropic GABA and glutamate receptors,

increases the odds of acute toxicity significantly (Chapter 5). Also, we provided novel insights into the cholinergic toxidrome which is characterized by inhibiting acetylcholinesterase AChE, and downstream effects including inflammation and calcium ion dysregulation. [268] The probability of acute toxicity is significantly higher when AChE inhibitors also interfere with the calcium sensitization of troponin[319,320] and inflammatory responses of NLRP3[324], resulting in cardiovascular effects.

Finally, we tested the predictive power of rules for acute toxicity, in a conformal prediction framework, against Random Forest as a benchmark model using chemical descriptors (Chapter 6). The performance of rule models was comparable with Random Forest at 80% confidence with an overall accuracy of 66.86% and 71.31%, respectively for. Additionally, we observed that both models exerted a bias over compound promiscuity. Despite the even distribution of toxicity classes across different levels of compound promiscuity, promiscuous compounds were frequently predicted as toxic, whereas low promiscuous compounds were often predicted as non-toxic.

Overall, our findings provide novel insights into the molecular mechanisms of toxicity and polypharmacology. The workflows proposed in this work can be generally used to improve the utility and design of alternative methods for toxicity. For example, assays predictive of toxicity can be prioritized so as to sufficiently cover the biological space of toxic compounds. Also, via understanding the synergistic interactions between key events, cost and time effective iterative screening protocols can be developed. In this way, initial screening of major bioactivity classes is followed up by screening relevant endpoints where interactions are expected to trigger toxicity.

Nonetheless, in order to accelerate regulatory acceptance of alternative methods for toxicity, interpretable and robust models should be developed which consider a number of factors including chemical reactivity, polypharmacology, bioavailability and exposure. Further analysis is to be conducted to understand complex mechanisms of toxicity and how perturbation of a pathway can progress into *in vivo* effect. Uncertainty of prediction should be assessed with care, and with regards to chemical space and compound promiscuity. Above all, collaboration and harmonization across the academic, governmental and private sectors is needed to accelerate the acceptance of alternative methods in regulatory toxicology.

REFERENCES

(1) Cruelty Free International (2018) Facts and figures on animal testing, http://www.crueltyfreeinternational.org/why-we-do- (accessed 2018-10-18).

(2) Pauwels, M., and Rogiers, V. (2010) Human health safety evaluation of cosmetics in the EU: a legally imposed challenge to science. *Toxicol. Appl. Pharmacol. 243*, 260–274.

(3) Ahuja, V., Bokan, S., and Sharma, S. (2017) Predicting toxicities in humans by nonclinical safety testing: an update with particular reference to anticancer compounds. *Drug Discov. Today 22*, 127–132.

(4) Shanks, N., Greek, R., and Greek, J. (2009) Are animal models predictive for humans? *Philos. Ethics, Humanit. Med. 4*, 1–20.

(5) Waring, M. J., Arrowsmith, J., Leach, A. R., Leeson, P. D., Mandrell, S., Owen, R. M., Pairaudeau, G., Pennie, W. D., Pickett, S. D., Wang, J., Wallace, O., and Weir, A. (2015) An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat. Rev. Drug Discov. 14*, 475–486.

(6) Hamm, J., Sullivan, K., Clippinger, A. J., Strickland, J., Bell, S., Bhhatarai, B., Blaauboer, B., Casey, W., Dorman, D., Forsby, A., Garcia-Reyero, N., Gehen, S., Graepel, R., Hotchkiss, J., Lowit, A., Matheson, J., Reaves, E., Scarano, L., Sprankle, C., Tunkel, J., Wilson, D., Xia, M., Zhu, H., and Allen, D. (2017) Alternative approaches for identifying acute systemic toxicity: Moving from research to regulatory testing. *Toxicol. In Vitro 41*, 245–259.

(7) Basic Information | Risk Assessment Portal | US EPA. https://www.epa.gov/risk/about-risk-assessment#whatisrisk (accesed 2015-09-14).

(8) OECD Guidelines for the Testing of Chemicals, Section 4:Health Effects.

(9) OECD Guidelines for the Testing of Chemicals, Section 4, Test No. 421: Reproduction/Developmental Toxicity Screening Test.

(10) OECD Guidelines for the Testing of Chemicals, Section 4, Test No. 451: Carcinogenicity Studies.

(11) OECD Guidelines for the Testing of Chemicals, Section 4, Test No. 452: Chronic Toxicity Studies.

(12) Andrade, E. L., Bento, A. F., Cavalli, J., Oliveira, S. K., Schwanke, R. C., Siqueira, J. M., Freitas, C. S., Marcon, R., and Calixto, J. B. (2016) Non-clinical studies in the process of new drug development - Part II: Good laboratory practice, metabolism, pharmacokinetics, safety and dose translation to clinical studies. *Brazilian J. Med. Biol. Res. 49*, e5646.

(13) EC. (2008) Council Regulation (EC) No. 440/2008 of 30 May 2008 Laying Down Test Methods Pursuant to Regulation (EC) No. 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorization, and Restriction of Chemicals (REACH). *Off. J. Eur. Union*.

(14) Mayr, L. M., and Bojanic, D. (2009) Novel trends in high-throughput screening. *Curr. Opin. Pharmacol. 9*, 580–8.

(15) Judson, R. S., Houck, K. A., Kavlock, R. J., Knudsen, T. B., Martin, M. T., Mortensen, H. M., Reif, D. M., Rotroff, D. M., Shah, I., Richard, A. M., and Dix, D. J. (2010) In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. *Environ. Health*

Perspect. 118, 485–492.

(16) Rang, H. P., and Vasella, D. (2006) Drug Discovery and Development: Technology in Transition (Hill, R., Ed.). Elsevier.

(17) Russell, W. M. S., and Burch, R. L. (1959) The principles of humane experimental technique. Methuen & Co. Ltd., London.

(18) Törnqvist, E., Annas, A., Granath, B., Jalkesten, E., Cotgreave, I., and Öberg, M. (2014) Strategic focus on 3R principles reveals major reductions in the use of animals in pharmaceutical toxicity testing. *PLoS One 9*, 101638.

(19) Strickland, J., Clippinger, A. J., Brown, J., Allen, D., Jacobs, A., Matheson, J., Lowit, A., Reinke, E. N., Johnson, M. S., Quinn, M. J., Mattie, D., Fitzpatrick, S. C., Ahir, S., Kleinstreuer, N., and Casey, W. (2018) Status of acute systemic toxicity testing requirements and data uses by U.S. regulatory agencies. *Regul. Toxicol. Pharmacol. 94*, 183–196.

(20) Burden, N., Mahony, C., Müller, B. P., Terry, C., Westmoreland, C., and Kimber, I. (2015) Aligning the 3Rs with new paradigms in the safety assessment of chemicals. *Toxicology 330*, 62–66.

(21) OECD. (2015) Guidance Document on Revisions to OECD Genetic Toxicology Test Guidelines. Genetic Toxicology Guidance Document: Second Commenting Round.

(22) OECD. (2009) Table 1 . Adopted OECD Test Guidelines on Short , and Long-Term Toxicity Testing with 3R Relevance. 1–4.

(23) OECD. (2015) Performance-Based Test Guideline for Human Recombinant Estrogen Receptor (hrER) In Vitro Assays to Detect Chemicals with ER Binding Affinity, OECD Guidelines for Testing Chemicals; 493 1–64.

(24) OECD. (2011) H295R Steroidogenesis Assay, OECD Guidelines for Testing Chemicals; 456.

(25) Rouquié, D., Heneweer, M., Botham, J., Ketelslegers, H., Markell, L., Pfister, T., Steiling, W., Strauss, V., and Hennes, C. (2015) Contribution of new technologies to characterization and prediction of adverse effects. *Crit. Rev. Toxicol. 45*, 172–183.

(26) Worth, A. P., and Patlewicz, G. (2016) Integrated approaches to testing and assessment, in *Validation of alternative methods for toxicity testing* (Eskes, C., and Whelan, M., Eds.), pp 317–342. Springer International Publishing.

(27) Bernauer, U., Oberemm, A., Madie, S., and Gundert-Remy, U. (2005) The use of in vitro data in risk assessment. *Basic Clin. Pharmacol. Toxicol. 96*, 176–181.

(28) Jennings, P. (2015) The future of in vitro toxicology. *Toxicol. In Vitro 29*, 1217–1221.

(29) Hertzberg, R. P., and Pope, A. J. (2000) High-throughput screening: new technology for the 21st century. *Curr. Opin. Chem. Biol. 4*, 445–451.

(30) Sittampalam, G. S., Kahl, S. D., and Janzen, W. P. (1997) High-throughput screening: advances in assay technologies. *Curr. Opin. Chem. Biol. 1*, 384–391.

(31) Broach, J. R., and Thorner, J. (1996) High-throughput screening for drug discovery. *Nature 384*, 14–6.

(32) Inglese, J., Johnson, R. L., Simeonov, A., Xia, M., Zheng, W., Austin, C. P., and Auld, D. S. (2007) High-throughput screening assays for the identification of chemical probes. *Nat. Chem. Biol. 3*, 466–79.

(33) Bleicher, K. H., Böhm, H.-J., Müller, K., and Alanine, A. I. (2003) Hit and lead generation: beyond high-throughput screening. *Nat. Rev. Drug Discov. 2*, 369–78.

(34) Scheiber, J., Chen, B., Milik, M., Sukuru, S. C. K., Bender, A., Mikhailov, D., Whitebread, S., Hamon, J., Azzaoui, K., Urban, L., Glick, M., Davies, J. W., and Jenkins, J. L. (2009) Gaining insight into off-target mediated effects of drug candidates with a comprehensive systems chemical biology analysis. *J. Chem. Inf. Model. 49*, 308–17.

(35) Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B. A., Wang, J., Yu, B., Zhang, J., and Bryant, S. H. (2016) PubChem substance and compound databases. *Nucleic Acids Res. 44*, D1202–D1213.

(36) Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., and Overington, J. P. (2012) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res. 40*, D1100–D1107.

(37) Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., and Woolsey, J. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res. 34*, D668–D672.

(38) Seiler, K. P., George, G. A., Happ, M. P., Bodycombe, N. E., Carrinski, H. A., Norton, S., Brudz, S., Sullivan, J. P., Muhlich, J., Serrano, M., Ferraiolo, P., Tolliday, N. J., Schreiber, S. L., and Clemons, P. A. (2007) ChemBank: a small-molecule screening and cheminformatics resource database. *Nucleic Acids Res. 36*, D351–D359.

(39) Whitebread, S., Hamon, J., Bojanic, D., and Urban, L. (2005) Keynote review: in vitro safety pharmacology profiling: an essential tool for successful drug development. *Drug Discov. Today 10*, 1421–33.

(40) Whitebread, S., Dumotier, B., Armstrong, D., Fekete, A., Chen, S., Hartmann, A., Muller, P. Y., and Urban, L. (2016) Secondary pharmacology: screening and interpretation of off-target activities – focus on translation. *Drug Discov. Today 21*, 1232–1242.

(41) Bowes, J., Brown, A. J., Hamon, J., Jarolimek, W., Sridhar, A., Waldron, G., and Whitebread, S. (2012) Reducing safety-related drug attrition: the use of in vitro pharmacological profiling. *Nat. Rev. Drug Discov. 11*, 909–22.

(42) Azzaoui, K., Hamon, J., Faller, B., Whitebread, S., Jacoby, E., Bender, A., Jenkins, J. L., and Urban, L. (2007) Modeling promiscuity based on in vitro safety pharmacology profiling data. *ChemMedChem 2*, 874–880.

(43) Peters, J. U., Hert, J., Bissantz, C., Hillebrecht, A., Gerebtzoff, G., Bendels, S., Tillier, F., Migeon, J., Fischer, H., Guba, W., and Kansy, M. (2012) Can we discover pharmacological promiscuity early in the drug discovery process? *Drug Discov. Today 17*, 325–335.

(44) Wang, X., and Greene, N. (2012) Comparing measures of promiscuity and exploring their relationship to toxicity. *Mol. Inform. 31*, 145–159.

(45) Mencher, S. K., and Wang, L. G. (2005) Promiscuous drugs compared to selective drugs (promiscuity can be a virtue). *BMC Clin. Pharmacol. 5*, 1:7.

(46) Muthas, D., Boyer, S., and Hasselgren, C. (2013) A critical assessment of modeling safety-related drug attrition. *Medchemcomm 4*, 1058–1065.

(47) Papoian, T., Chiu, H.-J., Elayan, I., Jagadeesh, G., Khan, I., Laniyonu, A. A., Xinguang Li, C., Saulnier, M., Simpson, N., Yang, B., Li, C. X., Saulnier, M., Simpson, N., and Yang, B. (2015) Secondary pharmacology data to assess potential off-target activity of new drugs: a regulatory perspective. *Nat. Rev. Drug Discov. 14*, 294.

(48) Muller, P. Y., and Milton, M. N. (2012) The determination and interpretation of the therapeutic index in drug development. *Nat. Rev. Drug Discov. 11*, 751–761.

(49) Kavlock, R., Chandler, K., Houck, K., Hunter, S., Judson, R., Kleinstreuer, N., Knudsen, T., Martin, M., Padilla, S., Reif, D., Richard, A., Rotroff, D., Sipes, N., and Dix, D. (2012) Update on EPA's ToxCast program: Providing high throughput decision support tools for chemical risk management. *Chem. Res. Toxicol. 25*, 1287–1302.

(50) Dix, D. J., Houck, K. A., Martin, M. T., Richard, A. M., Setzer, R. W., and Kavlock, R. J. (2007) The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol. Sci. 95*, 5–12.

(51) Krewski, D., Acosta, D., Andersen, M., Anderson, H., Bailar, J. C., Boekelheide, K., Brent, R., Charnley, G., Cheung, V. G., Green, S., Kelsey, K. T., Kerkvliet, N. I., Li, A. A., McCray, L., Meyer, O., Patterson, R. D., Pennie, W., Scala, R. A., Solomon, G. M., Stephens, M., Yager, J., and Zeise, L. (2010) Toxicity testing in the 21st century: a vision and a strategy. *J. Toxicol. Environ. Health. B. Crit. Rev. 13*, 51–138.

(52) Richard, A. M., Judson, R. S., Houck, K. A., Grulke, C. M., Volarath, P., Thillainadarajah, I., Yang, C., Rathman, J., Martin, M. T., Wambaugh, J. F., Knudsen, T. B., Kancherla, J., Mansouri, K., Patlewicz, G., Williams, A. J., Little, S. B., Crofton, K. M., and Thomas, R. S. (2016) ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology. *Chem. Res. Toxicol. 29*, 1225–1251.

(53) Tice, R. R., Austin, C. P., Kavlock, R. J., and Bucher, J. R. (2013) Improving the human hazard characterization of chemicals: a Tox21 update. *Environ. Health Perspect. 121*, 756–65.

(54) OECD. (2017) Guidance Document for the Use of Adverse Outcome Pathways in Developing Integrated Approaches to Testing and Assessment (IATA), OECD Guidelines for the Testing of Chemicals 1–31.

(55) Sewell, F., Doe, J., Gellatly, N., Ragan, I., and Burden, N. (2017) Steps towards the international regulatory acceptance of non-animal methodology in safety assessment. *Regul. Toxicol. Pharmacol. 89*, 50–56.

(56) Blaauboer, B. J., Barratt, M. D., and Houston, J. B. (1999) The Integrated Use of Alternative Methods in Toxicological Risk Evaluation - ECVAM Integrated Testing Strategies Task Force Report 1. *Altern. Lab. Anim. 27*, 229–37.

(57) Borgert, C. J., Quill, T. F., McCarty, L. S., and Mason, A. M. (2004) Can mode of action predict mixture toxicity for risk assessment? *Toxicol. Appl. Pharmacol. 201*, 85–96.

(58) Ankley, G. T., Bennett, R. S., Erickson, R. J., Hoff, D. J., Hornung, M. W., Johnson, R. D., Mount, D. R., Nichols, J. W., Russom, C. L., Schmieder, P. K., Serrrano, J. A., Tietge, J. E., and Villeneuve, D. L. (2010) Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environ. Toxicol. Chem. 29*, 730–741.

(59) Villeneuve, D. L., Crump, D., Garcia-Reyero, N., Hecker, M., Hutchinson, T. H., LaLone, C. A., Landesmann, B., Lettieri, T., Munn, S., Nepelska, M., Ottinger, M. A., Vergauwen, L., and Whelan, M. (2014) Adverse outcome pathway (AOP) development I: Strategies and principles. *Toxicol. Sci. 142*, 312–320.

(60) Villeneuve, D. L., Crump, D., Garcia-Reyero, N., Hecker, M., Hutchinson, T. H., LaLone, C. A., Landesmann, B., Lettieri, T., Munn, S., Nepelska, M., Ottinger, M. A., Vergauwen, L., and Whelan, M. (2014) Adverse Outcome Pathway Development II: Best Practices. *Toxicol. Sci. 142*, 321–330.

(61) Wittwehr, C., Aladjov, H., Ankley, G., Byrne, H. J., de Knecht, J., Heinzle, E., Klambauer, G.,

Landesmann, B., Luijten, M., MacKay, C., Maxwell, G., Meek, M. E. B., Paini, A., Perkins, E., Sobanski, T., Villeneuve, D., Waters, K. M., and Whelan, M. (2017) How adverse outcome pathways can aid the development and use of computational prediction models for regulatory toxicology. *Toxicol. Sci. 155*, 326–336.

(62) Patlewicz, G., Simon, T. W., Rowlands, J. C., Budinsky, R. A., and Becker, R. A. (2015) Proposing a scientific confidence framework to help support the application of adverse outcome pathways for regulatory purposes. *Regul. Toxicol. Pharmacol. 71*, 463–477.

(63) OECD. (2013) Guidance document on developing and assessing Adverse Outcome Pathways. Series on Testing and Assessment, OECD Publications *184*, 1–32.

(64) Ashby, J. (1985) Fundamental structural alerts to potential carcinogenicity or noncarcinogenicity. *Environ. Mutagen. 7*, 919–921.

(65) Dehaspe, L., Toivonen, H., and King, R. D. (1998) Finding frequent substructures in chemical compounds, in *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, pp 30–36. AAAI Press.

(66) Alves, V. M., Muratov, E. N., Capuzzi, S. J., Politi, R., Low, Y., Braga, R. C., Zakharov, A. V., Sedykh, A., Mokshyna, E., Farag, S., Andrade, C. H., Kuz'min, V. E., Fourches, D., Tropsha, A., Luechtefeld, T., Maertens, A., Marty, S., Naciff, J. M. J. M., Palmer, J., Pamies, D., Penman, M., Richarz, A.-N. A.-N., Russo, D. P. D. P., Stuard, S. B. S. B., Patlewicz, G., Ravenzwaay, B. Van, Wu, S., Zhu, H., and Hartung, T. (2016) Alarms about structural alerts. *Green Chem. 18*, 4348–4360.

(67) Sherhod, R., Judson, P., Hanser, T., Vessey, J., Webb, S. J., and Gillet, V. (2014) Emerging Pattern Mining to Aid Toxicological Knowledge Discovery Emerging Pattern Mining to Aid Toxicological Knowledge Discovery. *J. Chem. Inf. Model 54*, 1864–1879.

(68) Liu, R., Yu, X., and Wallqvist, A. (2015) Data-driven identification of structural alerts for mitigating the risk of drug-induced human liver injuries. *J. Cheminform. 7*, 4.

(69) Lepailleur, A., Poezevara, G., and Bureau, R. (2013) Automated detection of structural alerts (chemical fragments) in (eco)toxicology. *Comput. Struct. Biotechnol. J. 5*, e201302013.

(70) Barber, C., Hanser, T., Judson, P., and Williams, R. (2017) Distinguishing between expert and statistical systems for application under ICH M7. *Regul. Toxicol. Pharmacol. 84*, 124–130.

(71) Marchant, C. A., Briggs, K. A., and Long, A. (2008) In Silico Tools for Sharing Data and Knowledge on Toxicity and Metabolism: Derek for Windows, Meteor, and Vitic. *Toxicol. Mech. Methods 18*, 177–187.

(72) Patlewicz, G., Jeliazkova, N., Safford, R. J., Worth, A. P., and Aleksiev, B. (2008) An evaluation of the implementation of the Cramer classification scheme in the Toxtree software. *SAR QSAR Environ. Res. 19*, 495–524.

(73) Sushko, I., Salmina, E., Potemkin, V. A., Poda, G., and Tetko, I. V. (2012) ToxAlerts: A web server of structural alerts for toxic chemicals and compounds with potential adverse reactions. *J. Chem. Inf. Model. 52*, 2310–2316.

(74) Cortes-Ciriano, I. (2016) Bioalerts: a python library for the derivation of structural alerts from bioactivity and toxicity data sets. *J. Cheminform. 8*, 13.

(75) Métivier, J. P., Lepailleur, A., Buzmakov, A., Poezevara, G., Crémilleux, B., Kuznetsov, S. O., Goff, J. Le, Napoli, A., Bureau, R., and Cuissart, B. (2015) Discovering structural alerts for mutagenicity using stable emerging molecular patterns. *J. Chem. Inf. Model. 55*, 925–940.

(76) Sherhod, R., Gillet, V. J., Judson, P. N., and Vessey, J. D. (2012) Automating Knowledge

Discovery for Toxicity Prediction Using Jumping Emerging Pattern Mining. *J. Chem. Inf. Model. 52*, 3074–3087.

(77) Garcia-Serna, R., Vidal, D., Remez, N., and Mestres, J. (2015) Large-Scale Predictive Drug Safety: From Structural Alerts to Biological Mechanisms. *Chem. Res. Toxicol. 28*, 1875–1887.

(78) Plunkett, L. M., Kaplan, A. M., and Becker, R. A. (2015) Challenges in using the ToxRefDB as a resource for toxicity prediction modeling. *Regul. Toxicol. Pharmacol. 72*, 610–614.

(79) Martin, M. T., Dix, D. J., Judson, R. S., Kavlock, R. J., Reif, D. M., Richard, A. M., Rotroff, D. M., Romanov, S., Medvedev, A., Poltoratskaya, N., Gambarian, M., Moeser, M., Makarov, S. S., and Houck, K. A. (2010) Impact of environmental chemicals on key transcription regulators and correlation to toxicity end points within EPA's ToxCast program. *Chem. Res. Toxicol. 23*, 578–90.

(80) Hu, B., Gifford, E., Wang, H., Bailey, W., and Johnson, T. (2015) Analysis of the ToxCast Chemical-Assay Space using the Comparative Toxicogenomics Database. *Chem. Res. Toxicol. 28*, 2210–2223.

(81) Svensson, F., Zoufir, A., Mahmoud, S., Afzal, A. M., Smit, I., Giblin, K. A., Clements, P. J., Mettetal, J. T., Pointon, A., Harvey, J. S., Greene, N., Williams, R. V, and Bender, A. (2018) Information-Derived Mechanistic Hypotheses for Structural Cardiotoxicity. *Chem. Res. Toxicol. 31*, 1119–1127.

(82) Piersma, A. H., Hessel, E. V., and Staal, Y. C. (2017) Retinoic acid in developmental toxicology: Teratogen, morphogen and biomarker. *Reprod. Toxicol. 72*, 53–61.

(83) Baker, N., Hunter, S., Franzosa, J., Richard, A., Judson, R., and Knudsen, T. (2016) Toxcast Chemical and Bioactivity Profiles for In Vitro Targets in the Retinoid Signaling System, in *SOT 2016 Annual Meeting*, p 2680.

(84) Fay, K. A., Villeneuve, D. L., Swintek, J., Edwards, S. W., Nelms, M. D., Blackwell, B. R., and Ankley, G. T. (2018) Differentiating Pathway-Specific From Nonspecific Effects in High-Throughput Toxicity Data: A Foundation for Prioritizing Adverse Outcome Pathway Development. *Toxicol. Sci. 163*, 500–515.

(85) Judson, R., Houck, K., Martin, M., Richard, A. M., Knudsen, T. B., Shah, I., Little, S., Wambaugh, J., Woodrow Setzer, R., Kothya, P., Phuong, J., Filer, D., Smith, D., Reif, D., Rotroff, D., Kleinstreuer, N., Sipes, N., Xia, M., Huang, R., Crofton, K., and Thomas, R. S. (2016) Analysis of the Effects of Cell Stress and Cytotoxicity on In Vitro Assay Activity Across a Diverse Chemical and Assay Space. *Toxicol. Sci. 152*, 323–339.

(86) Wambaugh, J. F., Hughes, M. F., Ring, C. L., MacMillan, D. K., Ford, J., Fennell, T. R., Black, S. R., Snyder, R. W., Sipes, N. S., Wetmore, B. A., Westerhout, J., Setzer, R. W., Pearce, R. G., Simmons, J. E., and Thomas, R. S. (2018) Evaluating in vitro-in vivo extrapolation of toxicokinetics. *Toxicol. Sci. 163*, 152–169.

(87) Wambaugh, J. F., Wetmore, B. A., Pearce, R., Strope, C., Goldsmith, R., Sluka, J. P., Sedykh, A., Tropsha, A., Bosgra, S., Shah, I., Judson, R., Thomas, R. S., and Woodrow Setzer, R. (2015) Toxicokinetic Triage for Environmental Chemicals. *Toxicol. Sci. 147*, 55–67.

(88) Yoon, M., Campbell, J. L., Andersen, M. E., and Clewell, H. J. (2012) Quantitative in vitro to in vivo extrapolation of cell-based toxicity assay results. *Crit. Rev. Toxicol. 42*, 633–652.

(89) Wetmore, B. A. (2015) Quantitative in vitro-to-in vivo extrapolation in a high-throughput environment. *Toxicology 332*, 94–101.

(90) Wetmore, B. A., Wambaugh, J. F., Ferguson, S. S., Sochaski, M. A., Rotroff, D. M., Freeman, K., Clewell, H. J., Dix, D. J., Andersen, M. E., Houck, K. A., Allen, B., Judson, R. S., Singh, R.,

Kavlock, R. J., Richard, A. M., and Thomas, R. S. (2012) Integration of Dosimetry, Exposure, and High-Throughput Screening Data in Chemical Toxicity Assessment. *Toxicol. Sci. 125*, 157–174.

(91) Punt, A. (2018) Toxicokinetics in Risk Evaluations. *Chem. Res. Toxicol. 31*, 285–286.

(92) MacDonald, M. L., Lamerdin, J., Owens, S., Keon, B. H., Bilter, G. K., Shang, Z., Huang, Z., Yu, H., Dias, J., Minami, T., Michnick, S. W., and Westwick, J. K. (2006) Identifying off-target effects and hidden phenotypes of drugs in human cells. *Nat. Chem. Biol. 2*, 329–37.

(93) Sipes, N. S., Martin, M. T., Kothiya, P., Reif, D. M., Judson, R. S., Richard, A. M., Houck, K. A., Dix, D. J., Kavlock, R. J., and Knudsen, T. B. (2013) Profiling 976 ToxCast chemicals across 331 enzymatic and receptor signaling assays. *Chem. Res. Toxicol. 26*, 878–95.

(94) Houck, K. A., Dix, D. J., Judson, R. S., Kavlock, R. J., Yang, J., and Berg, E. L. (2009) Profiling bioactivity of the ToxCast chemical library using BioMAP primary human cell systems. *J. Biomol. Screen. 14*, 1054–66.

(95) Kleinstreuer, N. C., Yang, J., Berg, E. L., Knudsen, T. B., Richard, A. M., Martin, M. T., Reif, D. M., Judson, R. S., Polokoff, M., Dix, D. J., Kavlock, R. J., and Houck, K. A. (2014) Phenotypic screening of the ToxCast chemical library to classify toxic and therapeutic mechanisms. *Nat. Biotechnol. 32*, 583–91.

(96) Benigni, R., Bossa, C., Giuliani, A., and Tcheremenskaia, O. (2010) Exploring in vitro/in vivo correlation: lessons learned from analyzing phase I results of the US EPA's ToxCast Project. *J. Environ. Sci. Health. C. Environ. Carcinog. Ecotoxicol. Rev. 28*, 272–86.

(97) Benigni, R., Bossa, C., Tcheremenskaya, O., and Giuliani, A. (2010) In Vitro / In Vivo Relationship in the Light of Toxcast Phase 1. *ALTEX Altern. zu Tierexperimenten 27*, 269–274.

(98) Abdelaziz, A., Sushko, Y., Novotarskyi, S., Körner, R., Brandmaier, S., and Tetko, I. V. (2015) Using Online Tool (iPrior) for Modeling ToxCast™ Assays Towards Prioritization of Animal Toxicity Testing. *Comb. Chem. High Throughput Screen. 18*, 420–38.

(99) Thomas, R. S., Black, M. B., Li, L., Healy, E., Chu, T.-M., Bao, W., Andersen, M. E., and Wolfinger, R. D. (2012) A comprehensive statistical analysis of predicting in vivo hazard using high-throughput in vitro screening. *Toxicol. Sci. 128*, 398–417.

(100) Dix, D. J., Houck, K. A., Judson, R. S., Kleinstreuer, N. C., Knudsen, T. B., Martin, M. T., Reif, D. M., Richard, A. M., Shah, I., Sipes, N. S., and Kavlock, R. J. (2012) Incorporating biological, chemical, and toxicological knowledge into predictive models of toxicity. *Toxicol. Sci. 130*, 440–441.

(101) Liu, J., Mansouri, K., Judson, R. S., Martin, M. T., Hong, H., Chen, M., Xu, X., Thomas, R. S., and Shah, I. (2015) Predicting hepatotoxicity using ToxCast in vitro bioactivity and chemical structure. *Chem. Res. Toxicol. 28*, 738−751.

(102) Huang, R., Xia, M., Sakamuru, S., Zhao, J., Shahane, S. A., Attene-Ramos, M., Zhao, T., Austin, C. P., and Simeonov, A. (2016) Modelling the Tox21 10 K chemical profiles for in vivo toxicity prediction and mechanism characterization. *Nat. Commun. 7*, 10425.

(103) Liam, R., Robinson, M., Palczewska, A., Palczewski, J., Kidley, N., Marchese Robinson, R. L., Palczewska, A., Palczewski, J., and Kidley, N. (2017) Comparison of the Predictive Performance and Interpretability of Random Forest and Linear Models on Benchmark Data Sets. *J. Chem. Inf. Model. 57*, 1773−1792.

(104) Cook, D., Brown, D., Alexander, R., March, R., Morgan, P., Satterthwaite, G., and Pangalos, M. N. (2014) Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat. Rev. Drug Discov. 13*, 419–431.

(105) Kaplowitz, N. (2004) Drug-Induced Liver Injury. *Clin. Infect. Dis. 38*, S44–S48.

(106) Vinken, M. (2015) Adverse Outcome Pathways and Drug-Induced Liver Injury Testing. *Chem. Res. Toxicol. 28*, 1391–7.

(107) Russmann, S., Kullak-Ublick, G., and Grattagliano, I. (2009) Current Concepts of Mechanisms in Drug-Induced Hepatotoxicity. *Curr. Med. Chem. 16*, 3041–3053.

(108) Michalopoulos, G. K. (2007) Liver regeneration. *J. Cell. Physiol. 213*, 286–300.

(109) Shah, F., Leung, L., Barton, H. A., Will, Y., Rodrigues, A. D., Greene, N., and Aleo, M. D. (2015) Setting Clinical Exposure Levels of Concern for Drug-Induced Liver Injury (DILI) Using Mechanistic in vitro Assays. *Toxicol. Sci. 147*, 500–514.

(110) Thompson, R. A., Isin, E. M., Li, Y., Weidolf, L., Page, K., Wilson, I., Swallow, S., Middleton, B., Stahl, S., Foster, A. J., Dolgos, H., Weaver, R., and Kenna, J. G. (2012) In vitro approach to assess the potential for risk of idiosyncratic adverse reactions caused by candidate drugs. *Chem. Res. Toxicol. 25*, 1616–1632.

(111) Sakatis, M. Z., Reese, M. J., Harrell, A. W., Taylor, M. A., Baines, I. A., Chen, L., Bloomer, J. C., Yang, E. Y., Ellens, H. M., Ambroso, J. L., Lovatt, C. A., Ayrton, A. D., and Clarke, S. E. (2012) Preclinical strategy to reduce clinical hepatotoxicity using in vitro bioactivation data for >200 compounds. *Chem. Res. Toxicol. 25*, 2067–2082.

(112) Kostadinova, R., Boess, F., Applegate, D., Suter, L., Weiser, T., Singer, T., Naughton, B., and Roth, A. (2013) A long-term three dimensional liver co-culture system for improved prediction of clinically relevant drug-induced hepatotoxicity. *Toxicol. Appl. Pharmacol. 268*, 1–16.

(113) O'Brien, P. J., Irwin, W., Diaz, D., Howard-Cofield, E., Krejsa, C. M., Slaughter, M. R., Gao, B., Kaludercic, N., Angeline, A., Bernardi, P., Brain, P., and Hougham, C. (2006) High concordance of drug-induced human hepatotoxicity with in vitro cytotoxicity measured in a novel cell-based model using high content screening. *Arch. Toxicol. 80*, 580–604.

(114) Persson, M., Loye, A. F., Jacquet, M., Mow, N. S., Thougaard, A. V., Mow, T., and Hornberg, J. J. (2014) High-content analysis/screening for predictive toxicology: Application to hepatotoxicity and genotoxicity. *Basic Clin. Pharmacol. Toxicol. 115*, 18–23.

(115) Persson, M., Løye, A. F., Mow, T., and Hornberg, J. J. (2013) A high content screening assay to predict human drug-induced liver injury during drug discovery. *J. Pharmacol. Toxicol. Methods 68*, 302–313.

(116) Bisgin, H. (2014) Toward predictive models for drug-induced liver injury in humans : are we there yet ? *Biomarkers Med 8*, 201–213.

(117) Kumar, R., Sharma, A., and Varadwaj, P. K. (2011) A prediction model for oral bioavailability of drugs using physicochemical properties by support vector machine. *J. Nat. Sci. Biol. Med. 2*, 168–73.

(118) Khojasteh, S. C., Wong, H., and Hop, C. E. C. A. (2011) ADME Properties and Their Dependence on Physicochemical Properties, in *Drug Metabolism and Pharmacokinetics Quick Guide*, pp 165–181. Springer New York, New York, NY.

(119) Chen, M., Borlak, J., and Tong, W. (2013) High lipophilicity and high daily dose of oral medications are associated with significant risk for drug-induced liver injury. *Hepatology 58*, 388–396.

(120) Chen, M., Suzuki, A., Borlak, J., Andrade, R. J., and Lucena, M. I. (2015) Drug-induced liver injury: Interactions between drug properties and host factors. *J. Hepatol. 63*, 503–514.

(121) Nakayama, S., Atsumi, R., Takakusa, H., Kobayashi, Y., Kurihara, A., Nagai, Y., Nakai, D., and Okazaki, O. (2009) A zone classification system for risk assessment of idiosyncratic drug toxicity using daily dose and covalent binding. *Drug Metab. Dispos. 37*, 1970–1977.

(122) ECHA. (2015) Guidance on information requirements and chemical safety assessment Chapter R.7a: Endpoint specific guidance.

(123) OECD. (2008) Test No. 425: Acute Oral Toxicity: Up-and-Down Procedure, OECD Guidelines for the Testing of Chemicals, Section 4.

(124) OECD. (2002) Test No. 420: Acute Oral Toxicity - Fixed Dose Procedure, OECD Guidelines for the Testing of Chemicals.

(125) Stokes, W. S., Casati, S., Strickland, J., and Paris, M. (2008) Neutral red uptake cytotoxicity tests for estimating starting doses for acute oral toxicity tests. *Curr. Protoc. Toxicol. 36*, 20.4.1-20.4.20.

(126) Ducharme, N. A., Reif, D. M., Gustafsson, J.-A., and Bondesson, M. (2015) Comparison of toxicity values across zebrafish early life stages and mammalian studies: Implications for chemical testing. *Reprod. Toxicol. 55*, 3–10.

(127) Russom, C. L., Lalone, C. A., Villeneuve, D. L., and Ankley, G. T. (2014) Development of an adverse outcome pathway for acetylcholinesterase inhibition leading to acute mortality. *Environ. Toxicol. Chem. 33*, 2157–2169.

(128) Sachana, M., Munn, S., and Bal-Price, A. (2017) Adverse Outcome Pathway on binding of agonists to ionotropic glutamate receptors in adult brain leading to excitotoxicity that mediates neuronal cell death, contributing to learning and memory impairment. OECD Series on Adverse Outcome Pathways, No. 6.

(129) Gong, P., Hong, H., and Perkins, E. J. (2015) Ionotropic GABA receptor antagonism-induced adverse outcome pathways for potential neurotoxicity biomarkers. *Biomark. Med. 9*, 1225–1239.

(130) Russom, C. L., Lalone, C. A., Villeneuve, D. L., and Ankley, G. T. (2014) Development of an adverse outcome pathway for acetylcholinesterase inhibition leading to acute mortality. *Environ. Toxicol. Chem. 33*, 2157–2169.

(131) Viney, J. L. (1995) Transgenic and gene knockout mice in cancer research. *Cancer Metastasis Rev. 14*, 77–90.

(132) Wang, Y., Wang, Q., Wu, B., Li, Y., and Lu, G. (2013) Correlation between TCDD acute toxicity and aryl hydrocarbon receptor structure for different mammals. *Ecotoxicol. Environ. Saf. 89*, 84–88.

(133) Villalobos-Molina, R., Vázquez-Cuevas, F. G., López-Guerrero, J. J., Figueroa-García, M. C., Gallardo-Ortiz, I. A., Ibarra, M., Rodríguez-Sosa, M., Gonzalez, F. J., and Elizondo, G. (2008) Vascular α1D-adrenoceptors are overexpressed in aorta of the aryl hydrocarbon receptor null mouse: Role of increased angiotensin II. *Auton. Autacoid Pharmacol. 28*, 61–67.

(134) Knapen, D., Vergauwen, L., Villeneuve, D. L., and Ankley, G. T. (2015) The potential of AOP networks for reproductive and developmental toxicity assay development. *Reprod. Toxicol. 56*, 52–55.

(135) Knapen, D., Angrish, M. M., Fortin, M. C., Katsiadaki, I., Leonard, M., Margiotta-Casaluci, L., Munn, S., O'Brien, J. M., Pollesch, N., Smith, L. C., Zhang, X., and Villeneuve, D. L. (2018) Adverse outcome pathway networks I: Development and applications. *Environ. Toxicol. Chem. 37*, 1723–1733.

(136) Villeneuve, D. L., Angrish, M. M., Fortin, M. C., Katsiadaki, I., Leonard, M., Margiotta-Casaluci, L., Munn, S., O'Brien, J. M., Pollesch, N. L., Smith, L. C., Zhang, X., and Knapen, D. (2018) Adverse outcome pathway networks II: Network analytics. *Environ. Toxicol. Chem. 37*, 1734–1748.

(137) Perkins, E. J., Antczak, P., Burgoon, L., Falciani, F., Garcia-Reyero, N., Gutsell, S., Hodges, G., Kienzler, A., Knapen, D., McBride, M., and Willett, C. (2015) Adverse outcome pathways for regulatory applications: Examination of four case studies with different degrees of completeness and scientific confidence. *Toxicol. Sci. 148*, 14–25.

(138) Garcia-Reyero, N. (2015) Are adverse outcome pathways here to stay? *Environ. Sci. Technol. 49*, 3–9.

(139) ICH. (2009) ICH harmonised tripartite guideline guidance on nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals. ICH Harmonised Tripartite Guideline.

(140) Burden, N., Sewell, F., and Chapman, K. (2015) Testing Chemical Safety: What Is Needed to Ensure the Widespread Application of Non-animal Approaches? *PLoS Biol. 13*, e1002156.

(141) Kuhn, M., and Johnson, K. (2013) Applied Predictive Modeling. Springer, New York.

(142) Ahlberg, E., Carlsson, L., and Boyer, S. (2014) Computational derivation of structural alerts from large toxicology data sets. *J. Chem. Inf. Model. 54*, 2945–2952.

(143) Aiken, L. S., and Stephen, G. (1991) Multiple regression: Testing and interpreting interactions. Sage Publications.

(144) Lazic, S. E., Edmunds, N., and Pollard, C. E. (2017) Predicting drug safety and communicating risk: benefits of a Bayesian approach. *Toxicol. Sci. 162*, 89–98.

(145) Judson, R., Elloumi, F., Setzer, R. W., Li, Z., and Shah, I. (2008) A comparison of machine learning algorithms for chemical toxicity classification using a simulated multi-scale data model. *BMC Bioinformatics 9*, 241.

(146) Allen, C. H. G., Koutsoukas, A., Cortés-Ciriano, I., Murrell, D. S., Malliavin, T. E., Glen, R. C., and Bender, A. (2016) Improving the prediction of organism-level toxicity through integration of chemical, protein target and cytotoxicity qHTS data. *Toxicol. Res. (Camb). 5*, 883–894.

(147) Li, X., Chen, L., Cheng, F., Wu, Z., Bian, H., Xu, C., Li, W., Liu, G., Shen, X., and Tang, Y. (2014) In silico prediction of chemical acute oral toxicity using multi-classification methods. *J. Chem. Inf. Model. 54*, 1061–1069.

(148) Unterthiner, T., Mayr, A., Klambauer, G., and Hochreiter, S. (2015) Toxicity Prediction using Deep Learning. *Front. Environ. Sci. 3*, 80.

(149) Dong, G., and Bailey, J. (2013) Contrast data mining : concepts, algorithms, and applications. CRC Press.

(150) Kumbhare, T. A., and Chobe, S. V. (2014) An Overview of Association Rule Mining Algorithms. *Int. J. Comput. Sci. Inf. Technol. 5*, 927–930.

(151) García-Borroto, M., Martínez-Trinidad, J. F., and Carrasco-Ochoa, J. A. (2012) A survey of emerging patterns for supervised classification. *Artif. Intell. Rev. 42*, 705–721.

(152) Kuhn, M., Weston, S., Coulter, N., Culp, M. C., and Maintainer, R. Q. (2015) Package "C50" Decision Trees and Rule-Based Models code for C5.0, R package version 0.1.2. https://CRAN.R-project.org/package=C50.

(153) Quinlan, J. R. (John R., and Ross, J. (1993) C4.5 : programs for machine learning. Morgan Kaufmann Publishers.

(154) Shannon, C. E. (1948) The mathematical theory of communication. *Bell Syst. Tech. J. 27*, 379–423.

(155) Yin, X. & Han, J. (2003) CPAR : Classification based on Predictive Association Rules, in *Proceedings of the SIAM International Conference on Data Mining*, pp 369–376. SIAM Press, San Francisco, CA.

(156) Quinlan, J. R., and Cameron-Jones, R. M. (1993) FOIL: A midterm report. Springer, Berlin, Heidelberg.

(157) Breiman, L. (2001) Random forests. *Mach. Learn. 45*, 5–32.

(158) Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984) Classification and Regression Trees. Chapman & Hall/CRC, Taylor & Francis Group.

(159) Svetnik, V., Liaw, A., Tong, C., Christopher Culberson, J., Sheridan, R. P., and Feuston, B. P. (2003) Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *J. Chem. Inf. Comput. Sci. 43*, 1947–1958.

(160) Mitchell, J. B. O. (2014) Machine learning methods in chemoinformatics. *WIREs Comput. Mol. Sci. 4*, 468–481.

(161) Klopman, G. (1984) Artificial intelligence approach to structure-activity studies. Computer automated structure evaluation of biological activity of organic molecules. *J. Am. Chem. Soc. 106*, 7315–7321.

(162) Fourches, D., Muratov, E., and Tropsha, A. (2010) Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. *J. Chem. Inf. Model. 50*, 1189–1204.

(163) Enoch, S. J. (2010) Recent Advances in QSAR Studies. *Program*. Springer Netherlands, Dordrecht.

(164) Klingspohn, W., Mathea, M., ter Laak, A., Heinrich, N., and Baumann, K. (2017) Efficiency of different measures for defining the applicability domain of classification models. *J. Cheminform. 9*, 44.

(165) Worth, A. P., Van Leeuwen, C. J., and Hartung, T. (2004) The prospects for using (Q)SARs in a changing political environment--high expectations and a key role for the european commission's joint research centre. *SAR QSAR Environ. Res. 15*, 331–343.

(166) Hanser, T., Barber, C., Marchaland, J. F., and Werner, S. (2016) Applicability domain: towards a more formal definition. *SAR QSAR Environ. Res. 27*, 865–881.

(167) Norinder, U., Carlsson, L., Boyer, S., and Eklund, M. (2015) Introducing conformal prediction in predictive modeling for regulatory purposes. A transparent and flexible alternative to applicability domain determination. *Regul. Toxicol. Pharmacol. 71*, 279–284.

(168) Eklund, M., Norinder, U., Boyer, S., and Carlsson, L. (2013) The application of conformal prediction to the drug discovery process. *Ann. Math. Artif. Intell. 74*, 117–132.

(169) Svensson, F., Norinder, U., and Bender, A. (2017) Modelling compound cytotoxicity using conformal prediction and PubChem HTS data. *Toxicol. Res. (Camb). 6*, 73–80.

(170) Ji, C., Svensson, F., Zoufir, A., Bender, A., and Valencia, A. (2018) eMolTox: prediction of

molecular toxicity with confidence. *Bioinformatics 34*, 2508–2509.

(171) Norinder, U., Carlsson, L., Boyer, S., and Eklund, M. (2014) Introducing Conformal Prediction in Predictive Modeling. A Transparent and Flexible Alternative to Applicability Domain Determination. *J. Chem. Inf. Model. 54*, 1596–1603.

(172) Vovk, V., Lindsay, D., Nouretdinov, I., and Gammerman, A. (2003) Mondrian Confidence Machine. *Tech. Rep.* 1–19.

(173) EPA, U. S. (2016) Toxicity ForeCaster (ToxCastTM) Data, https://www.epa.gov/chemical-research/toxicity-forecaster-toxcasttm-data (accessed 2017-07-19).

(174) Kuhn, M. (2008) Building Predictive Models in R Using the caret Package. *J. Stat. Softw. 28*, 1–26.

(175) Meyer, D., Dimitriadou, E., Hornik, K., Andreas, Leisch, and Friedrich, W. (2015) Package 'e1071': Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.7-0.1. https://CRAN.R-project.org/package=e1071.

(176) R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

(177) Lê, S., Josse, J., and Husson, F. (2008) FactoMineR : An R package for multivariate analysis. *J Stat Softw 25*, 1–18.

(178) Lantz, B. (2015) Machine Learning with R. Second Edi. Packt Publishing, Birmingham, UK.

(179) Everitt, B., and Hothorn, T. (2011) An Introduction to Applied Multivariate Analysis with R. Springer-Verlag New York.

(180) Lehman, A., O'Rourke, N., Hatcher, L., and Stepanski, E. (2013) JMP for Basic Univariate and Multivariate Statistics: Methods for Researchers and Social Scientists, Second Edi. SAS Institute Inc.

(181) Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., and Venables, B. "gplots": Various R Programming Tools for Plotting Data. R package version 3.0.1.1. https://CRAN.R-project.org/package=gplots.

(182) Levin, E. R. (2005) Integration of the extranuclear and nuclear actions of estrogen. *Mol. Endocrinol. 19*, 1951–9.

(183) Roy, A. K., Lavrovsky, Y., Song, C. S., Chen, S., Jung, M. H., Velu, N. K., Bi, B. Y., and Chatterjee, B. (1998) Regulation of Androgen Action. *Vitam. Horm. 55*, 309–352.

(184) Lu, N. Z., Wardell, S. E., Burnstein, K. L., Defranco, D., Fuller, P. J., Giguere, V., Hochberg, R. B., McKay, L., Renoir, J.-M., Weigel, N. L., Wilson, E. M., McDonnell, D. P., and Cidlowski, J. A. (2006) International Union of Pharmacology. LXV. The Pharmacology and Classification of the Nuclear Receptor Superfamily: Glucocorticoid, Mineralocorticoid, Progesterone, and Androgen Receptors. *Pharmacol. Rev. 58*, 782–797.

(185) Flamant, F., Baxter, J. D., Forrest, D., Refetoff, S., Samuels, H., Scanlan, T. S., Vennström, B., and Samarut, J. (2006) International Union of Pharmacology. LIX. The pharmacology and classification of the nuclear receptor superfamily: thyroid hormone receptors. *Pharmacol. Rev. 58*, 705–11.

(186) Patisaul, H. B., Fenton, S. E., and Aylor, D. (2018) Animal models of endocrine disruption. *Best Pract. Res. Clin. Endocrinol. Metab. 32*, 283–297.

(187) Gore, A. C., Chappell, V. A., Fenton, S. E., Flaws, J. A., Nadal, A., Prins, G. S., Toppari, J., and

Zoeller, R. T. (2016) EDC-2 : The Endocrine Society ' s Second Scientific Statement on Endocrine-Disrupting Chemicals. *Endocr. Rev. 36*, 1–150.

(188) Dunand-Sauthier, I., Walker, C. A., Narasimhan, J., Pearce, A. K., Wek, R. C., and Humphrey, T. C. (2005) Stress-activated protein kinase pathway functions to support protein synthesis and translational adaptation in response to environmental stress in fission yeast. *Eukaryot. Cell 4*, 1785–93.

(189) Webster, M., Witkin, K. L., and Cohen-Fix, O. (2009) Sizing up the nucleus: nuclear shape, size and nuclear-envelope assembly. *J. Cell Sci. 122*, 1477–1486.

(190) Shadel, G. S. (2008) Expression and maintenance of mitochondrial DNA: new insights into human disease pathology. *Am. J. Pathol. 172*, 1445–56.

(191) Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., and Darnell, J. (2000) Cell-Cell Adhesion and Communication, in *Molecular Cell Biology* Forth Edi. W. H. Freeman, New York.

(192) Eberwine, J. (1999) Regulation of Transcription by Transcription Factors, in *Basic Neurochemistry: Molecular, Cellular and Medical Aspects* 6th edi. Lippincott-Raven, Philadelphia.

(193) Kim, M. T., Huang, R., Sedykh, A., Wang, W., Xia, M., and 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*, 634–41.

(194) Wetmore, B. A., Wambaugh, J. F., Allen, B., Ferguson, S. S., Sochaski, M. A., Setzer, R. W., Houck, K. A., Strope, C. L., Cantwell, K., Judson, R. S., LeCluyse, E., Clewell, H. J., Thomas, R. S., and Andersen, M. E. (2015) Incorporating High-Throughput Exposure Predictions with Dosimetry-Adjusted In Vitro Bioactivity to Inform Chemical Toxicity Testing. *Toxicol. Sci. 148*, 121–136.

(195) Degoey, D. A., Chen, H. J., Cox, P. B., and Wendt, M. D. (2018) Beyond the Rule of 5: Lessons Learned from AbbVie's Drugs and Compound Collection. *J. Med. Chem. 61*, 2636–2651.

(196) Veber, D. F., Johnson, S. R., Cheng, H., Smith, B. R., Ward, K. W., and Kopple, K. D. (2002) Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J. Med. Chem. 45*, 2615–2623.

(197) Martin, M. T., Judson, R. S., Reif, D. M., Kavlock, R. J., and Dix, D. J. (2009) Profiling Chemicals Based on Chronic Toxicity Results from the U.S. EPA ToxRef Database. *Environ. Health Perspect. 117*, 392–399.

(198) Standardizer was used for structure canonicalization and transformation, JChem 15.12.14.0, 2015, ChemAxon (http://www.chemaxon.com).

(199) Landrum, G. RDKit: Open source chemoinformatics; http://www.rdkit.org.

(200) Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Ohl, P., Thiel, K., and Wiswedel, B. (2009) KNIME - the Konstanz information miner, in *Data Analysis, Machine Learning and Applications*, pp 319–326. Springer, Berlin, Heidelberg.

(201) Calculator Plugins: Calculator Plugins were used for structure property prediction and calculation, Marvin 15.12.14.0, 2015, ChemAxon (http://www.chemaxon.com).

(202) ChemAxon. Instant JChem was used for structure database management, search and prediction, Instant JChem 15.12.14.0, 2015, ChemAxon (http://www.chemaxon.com).

(203) Real, R., and Vargas, J. M. (1996) The probabilistic basis of Jaccard's index of similarity. *Syst. Biol. 45*, 380–385.

(204) Murtagh, F., and Legendre, P. (2014) Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *J. Classif. 31*, 274–295.

(205) Wickham, H., and Chang, W. (2016) 'ggplot2': Elegant Graphics for Data Analysis. Springer-Verlag New York.

(206) Huang, R., Xia, M., Sakamuru, S., Zhao, J., Shahane, S. A., Attene-Ramos, M., Zhao, T., Austin, C. P., and Simeonov, A. (2016) Modelling the Tox21 10 K chemical profiles for in vivo toxicity prediction and mechanism characterization. *Nat. Commun. 7*, 10425.

(207) Wienkers, L. C., and Heath, T. G. (2005) Predicting in vivo drug interactions from in vitro drug discovery data. *Nat. Rev. Drug Discov. 4*, 825–833.

(208) Murugaiyan, G., Martin, S., and Saha, B. (2007) CD40-induced countercurrent conduits for tumor escape or elimination? *Trends Immunol. 28*, 467–473.

(209) Zhai, Y., Shen, X. D., Gao, F., Zhao, A., Freitas, M. C., Lassman, C., Luster, A. D., Busuttil, R. W., and Kupiec-Weglinski, J. W. (2008) CXCL10 regulates liver innate immune response against ischemia and reperfusion injury. *Hepatology 47*, 207–214.

(210) Schmitz, V., Dombrowski, F., Prieto, J., Qian, C., Diehl, L., Knolle, P., Sauerbruch, T., Caselmann, W. H., Spengler, U., and Leifeld, L. (2006) Induction of murine liver damage by overexpression of CD40 ligand provides an experimental model to study fulminant hepatic failure. *Hepatology 44*, 430–439.

(211) Shu, U., Kiniwa, M., Wu, C. Y., Maliszewski, C., Vezzio, N., Hakimi, J., Gately, M., and Delespesse, G. (1995) Activated T cells induce interleukin-12 production by monocytes via CD40-CD40 ligand interaction. *Eur. J. Immunol. 25*, 1125–1128.

(212) Karthivashan, G., Arulselvan, P., and Fakurazi, S. (2015) Pathways involved in acetaminophen hepatotoxicity with specific targets for inhibition/downregulation. *RSC Adv. 5*, 62040–62051.

(213) Bone-Larson, C. L., Hogaboam, C. M., Evanhoff, H., Strieter, R. M., and Kunkel, S. L. (2001) IFN-gamma-inducible protein-10 (CXCL10) is hepatoprotective during acute liver injury through the induction of CXCR2 on hepatocytes. *J. Immunol. 167*, 7077–7083.

(214) Jaeschke, H., Williams, C. D., Ramachandran, A., and Bajt, M. L. (2012) Acetaminophen hepatotoxicity and repair: the role of sterile inflammation and innate immunity. *Liver Int. 32*, 8–20.

(215) Saiman, Y., and Friedman, S. L. (2012) The role of chemokines in acute liver injury. *Front. Physiol. 3*, 213.

(216) McMahan, R. H., Porsche, C. E., Edwards, M. G., and Rosen, H. R. (2016) Free fatty acids differentially downregulate chemokines in liver sinusoidal endothelial cells: Insights into non-alcoholic fatty liver disease. *PLoS One 11*, e0159217.

(217) Yamamoto, Y. (2006) Estrogen Receptor Mediates 17 -Ethynylestradiol Causing Hepatotoxicity. *J. Biol. Chem. 281*, 16625–16631.

(218) Mueller, K. M., Kornfeld, J.-W., Friedbichler, K., Blaas, L., Egger, G., Esterbauer, H., Hasselblatt, P., Schlederer, M., Haindl, S., Wagner, K.-U., Engblom, D., Haemmerle, G., Kratky, D., Sexl, V., Kenner, L., Kozlov, A. V, Terracciano, L., Zechner, R., Schuetz, G., Casanova, E., Pospisilik, J. A., Heim, M. H., and Moriggl, R. (2011) Impairment of hepatic growth hormone and glucocorticoid receptor signaling causes steatosis and hepatocellular carcinoma in mice. *Hepatology 54*, 1398–409.

(219) Thole, Z., Manso, G., Salgueiro, E., Revuelta, P., and Hidalgo, A. (2004) Hepatotoxicity induced by antiandrogens: A review of the literature. *Urol. Int. 73*, 289–295.

(220) Zhang, J., Huang, W., Qatanani, M., Evans, R. M., and Moore, D. D. (2004) The constitutive androstane receptor and pregnane X receptor function coordinately to prevent bile acid-induced hepatotoxicity. *J. Biol. Chem. 279*, 49517–49522.

(221) Lee, F. Y., de Aguiar Vallim, T. Q., Chong, H. K., Zhang, Y., Liu, Y., Jones, S. A., Osborne, T. F., and Edwards, P. A. (2010) Activation of the Farnesoid X Receptor Provides Protection against Acetaminophen-Induced Hepatic Toxicity. *Mol. Endocrinol. 24*, 1626–1636.

(222) Collin, M., Abdelrahman, M., and Thiemermann, C. (2004) Endogenous ligands of PPAR-γ reduce the liver injury in haemorrhagic shock. *Eur. J. Pharmacol. 486*, 233–235.

(223) Shan, W., Nicol, C. J., Ito, S., Bility, M. T., Kennett, M. J., Ward, J. M., Gonzalez, F. J., and Peters, J. M. (2007) Peroxisome proliferator-activated receptor-β/δ protects against chemically induced liver toxicity in mice. *Hepatology 47*, 225–235.

(224) Osawa, Y., Oboki, K., Imamura, J., Kojika, E., Hayashi, Y., Hishima, T., Saibara, T., Shibasaki, F., Kohara, M., and Kimura, K. (2015) Inhibition of Cyclic Adenosine Monophosphate (cAMP)-response Element-binding Protein (CREB)-binding Protein (CBP)/β-Catenin Reduces Liver Fibrosis in Mice. *EBIOM 2*, 1751–1758.

(225) AlSharari, S. D., Al-Rejaie, S. S., Abuohashish, H. M., Ahmed, M. M., and Hafez, M. M. (2016) Rutin Attenuates Hepatotoxicity in High-Cholesterol-Diet-Fed Rats. *Oxid. Med. Cell. Longev. 2016*, 1–11.

(226) Dooley, S., and Ten Dijke, P. (2012) TGF-β in progression of liver disease. *Cell Tissue Res. 347*, 245–256.

(227) Wagner, M., Zollner, G., and Trauner, M. (2011) Nuclear receptors in liver disease. *Hepatology 53*, 1023–1034.

(228) Zuniga, S., Firrincieli, D., Housset, C., and Chignard, N. (2011) Vitamin D and the vitamin D receptor in liver pathophysiology. *Clin. Res. Hepatol. Gastroenterol. 35*, 295–302.

(229) Saini, S. P. S., Zhang, B., Niu, Y., Jiang, M., Gao, J., Zhai, Y., Hoon Lee, J., Uppal, H., Tian, H., Tortorici, M. A., Poloyac, S. M., Qin, W., Venkataramanan, R., and Xie, W. (2011) Activation of liver X receptor increases acetaminophen clearance and prevents its toxicity in mice. *Hepatology 54*, 2208–17.

(230) Lemasters, J. J., Qian, T., Bradham, C. A., Brenner, D. A., Cascio, W. E., Trost, L. C., Nishimura, Y., Nieminen, A. L., and Herman, B. (1999) Mitochondrial dysfunction in the pathogenesis of necrotic and apoptotic cell death. *J. Bioenerg. Biomembr. 31*, 305–19.

(231) Jaeschke, H., Gores, G. J., Cederbaum, A. I., Hinson, J. A., Pessayre, D., and Lemasters, J. J. (2002) Mechanisms of hepatotoxicity. *Toxicol. Sci. 65*, 166–176.

(232) Vairetti, M., Ferrigno, A., Rizzo, V., Ambrosi, G., Bianchi, A., Richelmi, P., Blandini, F., and Armentero, M. T. (2012) Impaired hepatic function and central dopaminergic denervation in a rodent model of Parkinson's disease: A self-perpetuating crosstalk? *Biochim. Biophys. Acta - Mol. Basis Dis. 1822*, 176–184.

(233) Veenman, L., and Gavish, M. (2012) The Role of 18 kDa Mitochondrial Translocator Protein (TSPO) in Programmed Cell Death, and Effects of Steroids on TSPO Expression. *Curr. Mol. Med. 12*, 398–412.

(234) Hatori, A., Yui, J., Xie, L., Yamasaki, T., Kumata, K., Fujinaga, M., Wakizaka, H., Ogawa, M., Nengaki, N., Kawamura, K., and Zhang, M. R. (2014) Visualization of acute liver damage induced by cycloheximide in rats using PET with [18F]FEDAC, a radiotracer for translocator protein (18 kDa). *PLoS One 9*, e86625.

(235) Rincón-Sánchez, A. R., Covarrubias, A., Rivas-Estilla, A. M., Pedraza-Chaverrí, J., Cruz, C., Islas-Carbajal, M. C., Panduro, A., Estanes, A., and Armendáriz-Borunda, J. (2005) PGE2 alleviates kidney and liver damage, decreases plasma renin activity and acute phase response in cirrhotic rats with acute liver damage. *Exp. Toxicol. Pathol. 56*, 291–303.

(236) Takano, M., Nishimura, H., Kimura, Y., Washizu, J., Mokuno, Y., Nimura, Y., and Yoshikai, Y. (1998) Prostaglandin E2 protects against liver injury after Escherichia coli infection but hampers the resolution of the infection in mice. *J Immunol 161*, 3019–3025.

(237) Rudnick, D. A., Perlmutter, D. H., and Muglia, L. J. (2001) Prostaglandins are required for CREB activation and cellular proliferation during liver regeneration., in *Proceedings of the National Academy of Sciences of the United States of America*, pp 8885–8890.

(238) Zhang, H., Liu, Y., Wang, L., Li, Z., Zhang, H., Wu, J., Rahman, N., Guo, Y., Li, D., Li, N., Huhtaniemi, I., Tsang, S. Y., Gao, G. F., and Li, X. (2013) Differential effects of estrogen/androgen on the prevention of nonalcoholic fatty liver disease in the male rat. *J. Lipid Res. 54*, 345–57.

(239) Hanioka, N., Gonzalez, F. J., Lindberg, N. A., Liu, G., Gelboin, H. V., and Korzekwa, K. R. (1992) Site-directed mutagenesis of cytochrome P450s CYP2A1 and CYP2A2: influence of the distal helix on the kinetics of testosterone hydroxylation. *Biochemistry 31*, 3364–3370.

(240) Williams, T. M., and Borghoff, S. J. (2000) Induction of Testosterone Biotransformation Enzymes following Oral Administration of Methyl tert-Butyl Ether to Male Sprague-Dawley Rats. *Toxicol. Sci. 57*, 147–155.

(241) Olokpa, E., Bolden, A., and Stewart, L. V. (2016) The Androgen Receptor Regulates PPARγ Expression and Activity in Human Prostate Cancer Cells. *J. Cell. Physiol. 231*, 2664–72.

(242) Olokpa, E., Moss, P. E., and Stewart, L. V. (2017) Crosstalk between the Androgen Receptor and PPAR Gamma Signaling Pathways in the Prostate. *PPAR Res. 2017*, 1–13.

(243) Lin, H. Y., Yu, I. C., Wang, R. S., Chen, Y. T., Liu, N. C., Altuwaijri, S., Hsu, C. L., Ma, W. L., Jokinen, J., Sparks, J. D., Yeh, S., and Chang, C. (2008) Increased hepatic steatosis and insulin resistance in mice lacking hepatic androgen receptor. *Hepatology 47*, 1924–1935.

(244) Daskalopoulos, E. P., Lang, M. A., Marselos, M., Malliou, F., and Konstandi, M. (2012) D2-Dopaminergic Receptor-Linked Pathways: Critical Regulators of CYP3A, CYP2C, and CYP2D. *Mol. Pharmacol. 82*, 668–678.

(245) Harkitis, P., Daskalopoulos, E. P., Malliou, F., Lang, M. A., Marselos, M., Fotopoulos, A., Albucharali, G., and Konstandi, M. (2015) Dopamine D2-receptor antagonists down-regulate CYP1A1/2 and CYP1B1 in the rat liver. *PLoS One 10*, e0128708.

(246) Schwabe, R. F., Schnabl, B., Kweon, Y. O., and Brenner, D. A. (2001) CD40 Activates NF- B and c-Jun N-Terminal Kinase and Enhances Chemokine Secretion on Activated Human Hepatic Stellate Cells. *J. Immunol. 166*, 6812–6819.

(247) Yamaguchi, J., Tanaka, T., Eto, N., and Nangaku, M. (2015) Inflammation and hypoxia linked to renal injury by CCAAT/enhancer-binding protein δ. *Kidney Int. 88*, 262–275.

(248) Anderson, E. R., Taylor, M., Xue, X., Martin, A., Moons, D. S., Omary, M. B., and Shah, Y. M. (2012) The hypoxia-inducible factor-C/EBPα axis controls ethanol-mediated hepcidin repression. *Mol. Cell. Biol. 32*, 4068–77.

(249) Jakobsen, J. S., Waage, J., Rapin, N., Bisgaard, H. C., Larsen, F. S., and Porse, B. T. (2013) Temporal mapping of CEBPA and CEBPB binding during liver regeneration reveals dynamic occupancy and specific regulatory codes for homeostatic and cell cycle gene batteries. *Genome Res. 23*, 592–603.

(250) Barchetta, I., Carotti, S., Labbadia, G., Gentilucci, U. V., Muda, A. O., Angelico, F., Silecchia, G., Leonetti, F., Fraioli, A., Picardi, A., Morini, S., and Cavallo, M. G. (2012) Liver vitamin D receptor, CYP2R1, and CYP27A1 expression: relationship with liver histology and vitamin D3 levels in patients with nonalcoholic steatohepatitis or hepatitis C virus. *Hepatology 56*, 2180–2187.

(251) Petta, S., Grimaudo, S., Tripodo, C., Cabibi, D., Calvaruso, M., Di Cristina, A., Guarnotta, C., Macaluso, F. S. alvatore, Minissale, M. G. iovanna, Marchesini, G., and Craxì, A. (2015) The hepatic expression of vitamin D receptor is inversely associated with the severity of liver damage in genotype 1 chronic hepatitis C patients. *J. Clin. Endocrinol. Metab. 100*, 193–200.

(252) Chen, Y., and Goldstein, J. A. (2009) The transcriptional regulation of the human CYP2C genes. *Curr. Drug Metab. 10*, 567–78.

(253) Cook, L. C., Hillhouse, A. E., Myles, M. H., Lubahn, D. B., Bryda, E. C., Davis, J. W., and Franklin, C. L. (2014) The role of estrogen signaling in a mouse model of inflammatory bowel disease: a Helicobacter hepaticus model. *PLoS One 9*, e94209.

(254) Zeremski, M., Dimova, R., Astemborski, J., Thomas, D. L., and Talal, A. H. (2011) CXCL9 and CXCL10 chemokines as predictors of liver fibrosis in a cohort of primarily African-American injection drug users with chronic hepatitis C. *J. Infect. Dis. 204*, 832–6.

(255) Jaeschke, H. (2007) Troglitazone hepatotoxicity: Are we getting closer to understanding idiosyncratic liver injury? *Toxicol. Sci. 97*, 1–3.

(256) Masubuchi, Y. (2006) Metabolic and Non-Metabolic Factors Determining Troglitazone Hepatotoxicity: A Review. *Drug Metab. Pharmacokinet. 21*, 347–356.

(257) Vansant, G., Pezzoli, P., Saiz, R., Birch, A., Duffy, C., Ferre, F., and Monforte, J. (2006) Gene Expression Analysis of Troglitazone Reveals Its Impact on Multiple Pathways in Cell Culture: A Case for In Vitro Platforms Combined with Gene Expression Analysis for Early (Idiosyncratic) Toxicity Screening. *Int. J. Toxicol. 25*, 85–94.

(258) Ong, M. M. K., Latchoumycandane, C., and Boelsterli, U. A. (2007) Troglitazone-Induced Hepatic Necrosis in an Animal Model of Silent Genetic Mitochondrial Abnormalities. *Toxicol. Sci. 97*, 205–213.

(259) Jaeschke, H. (2007) Troglitazone hepatotoxicity: Are we getting closer to understanding idiosyncratic liver injury? *Toxicol. Sci. 97*, 1–3.

(260) Bale, S., Vernetti, L., Senutovitch, N., Jindal, R., Hegde, M., Gough, A., Mccarty, W. J., Bakan, A., Bhushan, A., Shun, T. Y., Golberg, I., Debiasio, R., Osman Usta, B., Taylor, D. L., and Yarmush, M. L. (2014) In Vitro Platforms for Evaluating Liver Toxicity. *Exp Biol Med 239*, 1180–1191.

(261) Giuliano, K. A., Gough, A. H., Taylor, D. L., Vernetti, L. A., and Johnston, P. A. (2010) Early safety assessment using cellular systems biology yields insights into mechanisms of action. *J. Biomol. Screen. 15*, 783–797.

(262) Messner, S., Agarkova, I., Moritz, W., and Kelm, J. M. (2013) Multi-cell type human liver microtissues for hepatotoxicity testing. *Arch. Toxicol. 87*, 209–213.

(263) Khetani, S., and Bhatia, S. (2007) Microscale culture of human liver cells for drug development. *Nat. Biotechnol. 26*, 120–6.

(264) Hop, C. E. C. A. (2012) Role of ADME Studies in Selecting Drug Candidates : Dependence of ADME Parameters on Physicochemical Properties, in *Encyclopedia of Drug Metabolism and Interactions*. John Wiley & Sons, Inc., Hoboken, NJ, USA.

(265) Ritchie, T. J., and Macdonald, S. J. F. (2009) The impact of aromatic ring count on compound

developability - are too many aromatic rings a liability in drug design? *Drug Discov. Today 14*, 1011–1020.

(266) Trainor, G. L. (2007) The importance of plasma protein binding in drug discovery. *Expert Opin. Drug Discov. 2*, 51–64.

(267) Pellegatti, M., Pagliarusco, S., Solazzo, L., and Colato, D. (2011) Plasma protein binding and blood-free concentrations: which studies are needed to develop a drug? *Expert Opin. Drug Metab. Toxicol. 7*, 1009–1020.

(268) Faria, M., Garcia-Reyero, N., Padrós, F., Babin, P. J., Sebastián, D., Cachot, J., Prats, E., Arick, M., Rial, E., Knoll-Gellida, A., Mathieu, G., Le Bihanic, F., Escalon, B. L., Zorzano, A., Soares, A. M. V. M., and Raldúa, D. (2015) Zebrafish Models for Human Acute Organophosphorus Poisoning. *Sci. Rep. 5*, 15591.

(269) Auer, J., and Bajorath, J. (2006) Emerging chemical patterns: A new methodology for molecular classification and compound selection. *J. Chem. Inf. Model. 46*, 2502–2514.

(270) Nagata, K., Washio, T., Kawahara, Y., and Unami, A. (2014) Toxicity prediction from toxicogenomic data based on class association rule mining. *Toxicol. Reports 1*, 1133–1142.

(271) Li, J., and Wong, L. (2002) Identifying good diagnostic gene groups from gene expression profiles using the concept of emerging patterns. *Bioinformatics 18*, 725–34.

(272) Namasivayam, V., Hu, Y., Balfer, J., and Bajorath, J. (2013) Classification of compounds with distinct or overlapping multi-target activities and diverse molecular mechanisms using emerging chemical patterns. *J. Chem. Inf. Model. 53*, 1272–1281.

(273) Bolton, E. E., Wang, Y., Thiessen, P. A., and Bryant, S. H. (2008) Chapter 12 – PubChem: Integrated Platform of Small Molecules and Biological Activities, in *Annual Reports in Computational Chemistry*, pp 217–241.

(274) Mervin, L. H., Afzal, A. M., Drakakis, G., Lewis, R., Engkvist, O., and Bender, A. (2015) Target prediction utilising negative bioactivity data covering large chemical space. *J. Cheminform. 7*, 51.

(275) Borgelt, C., Meinl, T., and Berthold, M. (2005) MoSS: a program for molecular substructure mining, in *Proceedings of the 1st international workshop on open source data miningRegulation of human metabolism by hypoxia-inducible factor*, pp 6–15.

(276) Winder, C., Azzi, R., and Wagner, D. (2005) The development of the globally harmonized system (GHS) of classification and labelling of hazardous chemicals. *J. Hazard. Mater. 125*, 29–44.

(277) Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res. 13*, 2498–2504.

(278) Fang, G., Wang, W., Oatley, B., Van Ness, B., Steinbach, M., and Kumar, V. (2011) Characterizing Discriminative Patterns. *Comput. Res. Repos. abs/1102.4*.

(279) Cortina-Borja, M., Smith, A. D., Combarros, O., and Lehmann, D. J. (2009) The synergy factor: a statistic to measure interactions in complex diseases. *BMC Res. Notes 2*, 105.

(280) Xu, Y., Pei, J., and 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*, 2672–2685.

(281) Li, X., Chen, L., Cheng, F., Wu, Z., Bian, H., Xu, C., Li, W., Liu, G., Shen, X., and Tang, Y.

(2014) In silico prediction of chemical acute oral toxicity using multi-classification methods. *J. Chem. Inf. Model. 54*, 1061–9.

(282) Sun, L., Zhang, C., Chen, Y., Li, X., Zhuang, S., Li, W., Liu, G., Lee, P. W., and Tang, Y. (2015) In silico prediction of chemical aquatic toxicity with chemical category approaches and substructural alerts. *Toxicol. Res. (Camb). 4*, 452–463.

(283) Girvan, M., and Newman, M. E. J. (2002) Community structure in social and biological networks., in *Proceedings of the National Academy of Sciences of the United States of America*, pp 7821–6.

(284) Lau, A., and Tymianski, M. (2010) Glutamate receptors, neurotoxicity and neurodegeneration. *Pflugers Arch. Eur. J. Physiol. 460*, 525–542.

(285) Hwang, J. Y., Kim, Y. H., Ahn, Y. H., Wie, M. B., and Koh, J. Y. (1999) N-methyl-D-aspartate receptor blockade induces neuronal apoptosis in cortical culture. *Exp. Neurol. 159*, 124–130.

(286) Rahn, C. A., Bombick, D. W., and Doolittle, D. J. (1991) Assessment of mitochondrial membrane potential as an indicator of cytotoxicity. *Fundam. Appl. Toxicol. 16*, 435–448.

(287) Bhhatarai, B., Wilson, D. M., Bartels, M. J., Chaudhuri, S., Price, P. S., and Carney, E. W. (2015) Acute Toxicity Prediction in Multiple Species by Leveraging Mechanistic ToxCast Mitochondrial Inhibition Data and Simulation of Oral Bioavailability. *Toxicol. Sci. 147*, 386–96.

(288) Schep, L. J., Slaughter, R. J., and Beasley, D. M. G. (2009) Nicotinic plant poisoning. *Clin. Toxicol. 47*, 771–781.

(289) Wolkove, N., and Baltzan, M. Amiodarone pulmonary toxicity. *Can. Respir. J. 16*, 43–8.

(290) Ying, G. G. (2006) Fate, behavior and effects of surfactants and their degradation products in the environment. *Environ. Int. 32*, 417–431.

(291) Lémery, E., Briançon, S., Chevalier, Y., Bordes, C., Oddos, T., Gohier, A., and Bolzinger, M.-A. (2015) Skin toxicity of surfactants: Structure/toxicity relationships. *Colloids Surfaces A Physicochem. Eng. Asp. 469*, 166–179.

(292) Ullrich, V., King, L. J., Wolf, C. R., and Nastainczyk, W. (1979) Carbenes and Free Radicals of Haloalkanes as Toxic Intermediates. *Toxicology* 131–138.

(293) Weber, L. W. D., Boll, M., and Stampfl, A. (2003) Hepatotoxicity and mechanism of action of haloalkanes: Carbon tetrachloride as a toxicological model. *Crit. Rev. Toxicol. 33*, 105–136.

(294) Eisentraeger, A., Brinkmann, C., Hollert, H., Sagner, A., Tiehm, A., and Neuwoehner, J. (2008) Heterocyclic compounds: Toxic effects using algae, daphnids, and the Salmonella/microsome test taking methodical quantitative aspects into account. *Environ. Toxicol. Chem. 27*, 1590–1596.

(295) Pezdirc, M., Žegura, B., and Filipič, M. (2013) Genotoxicity and induction of DNA damage responsive genes by food-borne heterocyclic aromatic amines in human hepatoma HepG2 cells. *Food Chem. Toxicol. 59*, 386–394.

(296) Goswami, D. G., Kumar, D., Tewari-Singh, N., Orlicky, D. J., Jain, A. K., Kant, R., Rancourt, R. C., Dhar, D., Inturi, S., Agarwal, C., White, C. W., and Agarwal, R. (2015) Topical nitrogen mustard exposure causes systemic toxic effects in mice. *Exp. Toxicol. Pathol. 67*, 161–170.

(297) Colovic, M. B., Krstic, D. Z., Lazarevic-Pasti, T. D., Bondzic, A. M., and Vasic, V. M. (2013) Acetylcholinesterase Inhibitors: Pharmacology and Toxicology. *Curr. Neuropharmacol. 11*, 315–335.

(298) Mcaninch, E. A., and Bianco, A. C. (2014) Thyroid hormone signaling in energy homeostasis

and energy metabolism. *Ann. N. Y. Acad. Sci. 1311*, 77–87.

(299) Biondi, B., Palmieri, E. A., Lombardi, G., and Fazio, S. (2002) Effects of Thyroid Hormone on Cardiac Function - The Relative Importance of Heart Rate, Loading Conditions, and Myocardial Contractility in the Regulation of Cardiac Performance in Human Hyperthyroidism. *J. Clin. Endocrinol. Metab. 87*, 968–974.

(300) Mancini, A., Di Segni, C., Raimondo, S., Olivieri, G., Silvestrini, A., Meucci, E., and Currò, D. (2016) Thyroid Hormones, Oxidative Stress, and Inflammation. *Mediators Inflamm. 2016*, 6757154.

(301) Takahashi, K., Furuya, F., Shimura, H., Kaneshige, M., and Kobayashi, T. (2014) Impaired oxidative endoplasmic reticulum stress response caused by deficiency of thyroid hormone receptor α. *J. Biol. Chem. 289*, 12485–93.

(302) Dietrich, C. (2016) Antioxidant Functions of the Aryl Hydrocarbon Receptor. *Stem Cells Int. 2016*, 1–10.

(303) Uberti, F., Lattuada, D., Morsanuto, V., Nava, U., Bolis, G., Vacca, G., Squarzanti, D. F., Cisari, C., and Molinari, C. (2014) Vitamin D protects human endothelial cells from oxidative stress through the autophagic and survival pathways. *J. Clin. Endocrinol. Metab. 99*, 1367–1374.

(304) Bosland, M. C., and Mahmoud, A. M. (2011) Hormones and prostate carcinogenesis: Androgens and estrogens. *J. Carcinog. 10*, 33.

(305) Lee, C. (2017) Collaborative Power of Nrf2 and PPAR γ Activators against Metabolic and Drug-Induced Oxidative Injury. *Oxid. Med. Cell. Longev. 2017*, 1–14.

(306) Yang, B., and Papoian, T. (2018) Preclinical approaches to assess potential kinase inhibitor-induced cardiac toxicity: Past, present and future. *J. Appl. Toxicol. 38*, 790–800.

(307) Kong, D., and Yamori, T. (2008) Phosphatidylinositol 3-kinase inhibitors: Promising drug candidates for cancer therapy. *Cancer Sci. 99*, 1734–1740.

(308) Tokarska-Schlattner, M. (2005) Acute toxicity of doxorubicin on isolated perfused heart: response of kinases regulating energy supply. *AJP Hear. Circ. Physiol. 289*, H37–H47.

(309) Barquilla, A., and Pasquale, E. B. (2015) Eph receptors and ephrins: therapeutic opportunities. *Annu. Rev. Pharmacol. Toxicol. 55*, 465–87.

(310) Wu, B., Wang, S., De, S. K., Barile, E., Quinn, B. A., Zharkikh, I., Purves, A., Stebbins, J. L., Oshima, R. G., Fisher, P. B., and Pellecchia, M. (2015) Design and Characterization of Novel EphA2 Agonists for Targeted Delivery of Chemotherapy to Cancer Cells. *Chem. Biol. 22*, 876–887.

(311) Amato, K. R., Wang, S., Hastings, A. K., Youngblood, V. M., Santapuram, P. R., Chen, H., Cates, J. M., Colvin, D. C., Ye, F., Brantley-Sieders, D. M., Cook, R. S., Tan, L., Gray, N. S., and Chen, J. (2014) Genetic and pharmacologic inhibition of EPHA2 promotes apoptosis in NSCLC. *J. Clin. Invest. 124*, 2037–49.

(312) Shen, H. M., and Liu, Z. G. (2006) JNK signaling pathway is a key modulator in cell death mediated by reactive oxygen and nitrogen species. *Free Radic. Biol. Med. 40*, 928–939.

(313) Zhang, J., Min, R. W. M., Le, K., Zhou, S., Aghajan, M., Than, T. A., Win, S., and Kaplowitz, N. (2017) The role of MAP2 kinases and p38 kinase in acute murine liver injury models. *Cell Death Dis. 8*, e2903.

(314) Moslehi, J. J., and Deininger, M. (2015) Tyrosine Kinase Inhibitor-Associated Cardiovascular Toxicity in Chronic Myeloid Leukemia. *J. Clin. Oncol. 33*, 4210–8.

(315) Ashpole, N. M., Song, W., Brustovetsky, T., Engleman, E. A., Brustovetsky, N., Cummins, T. R., and Hudmon, A. (2012) Calcium/calmodulin-dependent protein kinase II (CaMKII) inhibition induces neurotoxicity via dysregulation of glutamate/calcium signaling and hyperexcitability. *J. Biol. Chem. 287*, 8495–8506.

(316) Swaminathan, P. D., Purohit, A., Hund, T. J., and Anderson, M. E. (2012) Calmodulin-dependent protein kinase II: Linking heart failure and arrhythmias. *Circ. Res. 110*, 1661–1677.

(317) Gonano, L. A., Sepulveda, M., Rico, Y., Kaetzel, M., Valverde, C. A., Dedman, J., Mattiazzi, A., and Vila Petroff, M. (2011) Calcium-Calmodulin Kinase II Mediates Digitalis-Induced Arrhythmias. *Circ. Arrhythmia Electrophysiol. 4*, 947–957.

(318) Chung, J. H., Biesiadecki, B. J., Ziolo, M. T., Davis, J. P., and Janssen, P. M. L. (2016) Myofilament calcium sensitivity: Role in regulation of in vivo cardiac contraction and relaxation. *Front. Physiol. 7*, 562.

(319) Pagel, P. S., Haikala, H., Pentikäinen, P. J., Toivonen, M.-L., Nieminen, M. S., Lehtonen, L., Papp, J. G., and Warltier, D. C. (1996) Pharmacology of Levosimendan: A New Myofilament Calcium Sensitizer. *Cardiovasc. Drug Rev. 14*, 286–316.

(320) Wilson, K., Guggilam, A., West, T. A., Zhang, X., Trask, A. J., Cismowski, M. J., de Tombe, P., Sadayappan, S., and Lucchesi, P. A. (2014) Effects of a myofilament calcium sensitizer on left ventricular systolic and diastolic function in rats with volume overload heart failure. *Am. J. Physiol. Circ. Physiol. 307*, H1605–H1617.

(321) Parvatiyar, M. S., Pinto, J. R., Dweck, D., and Potter, J. D. (2010) Cardiac troponin mutations and restrictive cardiomyopathy. *J. Biomed. Biotechnol. 2010*, 1–9.

(322) Chen, Z., Mou, R.-T., Feng, D.-X., Wang, Z., and Chen, G. (2017) The role of nitric oxide in stroke. *Med. Gas Res. 7*, 194–203.

(323) Deng, S., Kruger, A., Schmidt, A., Metzger, A., Yan, T., Gödtel-Armbrust, U., Hasenfuss, G., Brunner, F., and Wojnowski, L. (2009) Differential roles of nitric oxide synthase isozymes in cardiotoxicity and mortality following chronic doxorubicin treatment in mice. *Naunyn. Schmiedebergs. Arch. Pharmacol. 380*, 25–34.

(324) Liu, D., Zeng, X., Li, X., Mehta, J. L., and Wang, X. (2018) Role of NLRP3 inflammasome in the pathogenesis of cardiovascular diseases. *Basic Res. Cardiol. 113*, 5.

(325) Reese, I., Ballmer-Weber, B., Beyer, K., Fuchs, T., Kleine-Tebbe, J., Klimek, L., Lepp, U., Niggemann, B., Saloga, J., Schäfer, C., Werfel, T., Zuberbier, T., and Worm, M. (2017) German guideline for the management of adverse reactions to ingested histamine. *Allergo J. Int. 26*, 72–79.

(326) Taylor, S. L., Stratton, J. E., and Nordlee, J. A. (1989) Histamine poisoning (scombroid fish poisoning): an allergy-like intoxication. *J. Toxicol. Clin. Toxicol. 27*, 225–40.

(327) Villalba, J. M., and Alcaín, F. J. (2012) Sirtuin activators and inhibitors. *Biofactors 38*, 349–59.

(328) Kim, H.-S., Vassilopoulos, A., Wang, R.-H., Lahusen, T., Xiao, Z., Xu, X., Li, C., Veenstra, T. D., Li, B., Yu, H., Ji, J., Wang, X. W., Park, S.-H., Cha, Y. I., Gius, D., and Deng, C.-X. (2011) SIRT2 Maintains Genome Integrity and Suppresses Tumorigenesis through Regulating APC/C Activity. *Cancer Cell 20*, 487–499.

(329) Schetter, A. J., and Harris, C. C. (2012) Tumor suppressor p53 (TP53) at the crossroads of the exposome and the cancer genome, in *Proceedings of the National Academy of Sciences*, pp 7955–7956.

(330) Zhu, W., Soonpaa, M. H., Chen, H., Shen, W., Payne, R. M., Liechty, E. A., Caldwell, R. L.,

Shou, W., and Field, L. J. (2009) Acute doxorubicin cardiotoxicity is associated with p53-induced inhibition of the mammalian target of rapamycin pathway. *Circulation 119*, 99–106.

(331) Altucci, L., and Gronemeyer, H. (2001) Nuclear receptors in cell life and death. *Trends Endocrinol. Metab. 12*, 460–468.

(332) Neto, M., Naval-Sánchez, M., Potier, D., Pereira, P. S., Geerts, D., Aerts, S., and Casares, F. (2017) Nuclear receptors connect progenitor transcription factors to cell cycle control. *Sci. Rep. 7*, 4845.

(333) Abd Elhalim, M. M., Ismail, N. S. M., M Yahya, S. M., Omar, Y. Y., Abd Rabou, A. A., Lasheen, D. S., Zawrah, M. F., and Elmegeed, G. A. (2017) Synthesis, Characterization, and Evaluation of Cytotoxic Effects of Novel Hybrid Steroidal Heterocycles as PEG Based Nanoparticles. *Asian Pac. J. Cancer Prev. 18*, 1937–1946.

(334) Singh, H., Jindal, D. P., Yadav, M. R., and Kumar, M. (1991) Heterosteroids and drug research. *Prog. Med. Chem. 28*, 233–300.

(335) Drocourt, L., Ourlin, J. C., Pascussi, J. M., Maurel, P., and Vilarem, M. J. (2002) Expression of CYP3A4, CYP2B6, and CYP2C9 is regulated by the vitamin D receptor pathway in primary human hepatocytes. *J. Biol. Chem. 277*, 25125–25132.

(336) Matsunawa, M., Amano, Y., Endo, K., Uno, S., Sakaki, T., Yamada, S., and Makishima, M. (2009) The aryl hydrocarbon receptor activator benzo[a]pyrene enhances vitamin D3 catabolism in macrophages. *Toxicol. Sci. 109*, 50–58.

(337) Vogel, C. F. A., William Chang, W. L., Kado, S., McCulloh, K., Vogel, H., Wu, D., Haarmann-Stemmann, T., Yang, G. X., Leung, P. S. C., Matsumura, F., and Gershwin, M. E. (2016) Transgenic overexpression of aryl hydrocarbon receptor repressor (AhRR) and AhR-mediated induction of CYP1A1, cytokines, and acute toxicity. *Environ. Health Perspect. 124*, 1071–1083.

(338) Diry, M., Tomkiewicz, C., Koehle, C., Coumoul, X., Bock, K. W., Barouki, R., and Transy, C. (2006) Activation of the dioxin/aryl hydrocarbon receptor (AhR) modulates cell plasticity through a JNK-dependent mechanism. *Oncogene 25*, 5570–5574.

(339) Kamat, P. K., Kalani, A., Tyagi, S. C., and Tyagi, N. (2015) Hydrogen sulfide epigenetically attenuates homocysteine-induced mitochondrial toxicity mediated through NMDA receptor in mouse brain endothelial (bEnd3) cells. *J. Cell. Physiol. 230*, 378–394.

(340) Kysenius, K., Brunello, C. A., and Huttunen, H. J. (2014) Mitochondria and NMDA receptor-dependent toxicity of berberine sensitizes neurons to glutamate and rotenone injury. *PLoS One 9*, e107129.

(341) Duarte, A. I., Santos, M. S., Seiça, R., and Oliveira, C. R. (2004) Oxidative stress affects synaptosomal gamma-aminobutyric acid and glutamate transport in diabetic rats: the role of insulin. *Diabetes 53*, 2110–6.

(342) Duarte, A., Santos, M., Seiça, R., and Resende de Oliveira, C. (2000) Effect of Oxidative Stress on the Uptake of GABA and Glutamate in Synaptosomes Isolated from Diabetic Rat Brain. *Neuroendocrinology 72*, 179–186.

(343) Wiens, S. C., and Trudeau, V. L. (2006) Thyroid hormone and gamma-aminobutyric acid (GABA) interactions in neuroendocrine systems. *Comp. Biochem. Physiol. A. Mol. Integr. Physiol. 144*, 332–44.

(344) Féron, F., Burne, T. H. J., Brown, J., Smith, E., McGrath, J. J., Mackay-Sim, A., and Eyles, D. W. (2005) Developmental Vitamin D3deficiency alters the adult rat brain. *Brain Res. Bull. 65*, 141–148.

(345) Cho, H., Nam, G. B., Lee, S. H., Earm, Y. E., and Ho, W. K. (2001) Phosphatidylinositol 4,5-bisphosphate is acting as a signal molecule in alpha(1)-adrenergic pathway via the modulation of acetylcholine-activated K(+) channels in mouse atrial myocytes. *J. Biol. Chem. 276*, 159–64.

(346) Martínez-Miguel, P., Valdivielso, J. M., Medrano-Andrés, D., Román-García, P., Cano-Peñalver, J. L., Rodríguez-Puyol, M., Rodríguez-Puyol, D., and López-Ongil, S. (2014) The active form of vitamin D, calcitriol, induces a complex dual upregulation of endothelin and nitric oxide in cultured endothelial cells. *Am. J. Physiol. Metab. 307*, E1085–E1096.

(347) Bhatt, S., Qin, J., Bennett, C., Qian, S., Fung, J. J., Hamilton, T. A., and Lu, L. (2014) All-trans retinoic acid induces arginase-1 and inducible nitric oxide synthase-producing dendritic cells with T cell inhibitory function. *J. Immunol. 192*, 5098–108.

(348) Zhang, H., and Zhang, L. (2008) Role of protein kinase C isozymes in the regulation of alpha1-adrenergic receptor-mediated contractions in ovine uterine arteries. *Biol. Reprod. 78*, 35–42.

(349) Minneman, K. P. (1988) Alpha 1-adrenergic receptor subtypes, inositol phosphates, and sources of cell Ca2+. *Pharmacol. Rev. 40*, 87.

(350) Kalaria, R. N., and Prince, A. K. (1985) Effects of thyroid deficiency on the development of cholinergic, GABA, dopaminergic and glutamate neuron markers and DNA concentrations in the rat corpus striatum. *Int J Dev Neurosci. 3*, 655–666.

(351) Rowley, N. M., Madsen, K. K., Schousboe, A., and Steve White, H. (2012) Glutamate and GABA synthesis, release, transport and metabolism as targets for seizure control. *Neurochem. Int. 61*, 546–558.

(352) Seyfried, T. N., Glaser, G. H., and Yu, R. K. (1979) Thyroid hormone influence on the susceptibility of mice to audiogenic seizures. *Science. 205*, 598–600.

(353) Martin, J. V, Padron, J. M., Newman, M. A., Chapell, R., Leidenheimer, N. J., and Burke, L. A. (2004) Inhibition of the activity of the native γ-aminobutyric acidA receptor by metabolites of thyroid hormones: correlations with molecular modeling studies. *Brain Res. 1004*, 98–107.

(354) Accardi, M. V., Daniels, B. A., Brown, P. M. G. E., Fritschy, J. M., Tyagarajan, S. K., and Bowie, D. (2014) Mitochondrial reactive oxygen species regulate the strength of inhibitory GABA-mediated synaptic transmission. *Nat. Commun. 5*, 3168.

(355) Formenti, F., Constantin-Teodosiu, D., Emmanuel, Y., Cheeseman, J., Dorrington, K. L., Edwards, L. M., Humphreys, S. M., Lappin, T. R. J., McMullin, M. F., McNamara, C. J., Mills, W., Murphy, J. A., O'Connor, D. F., Percy, M. J., Ratcliffe, P. J., Smith, T. G., Treacy, M., Frayn, K. N., Greenhaff, P. L., Karpe, F., Clarke, K., and Robbins, P. A. (2010) Regulation of human metabolism by hypoxia-inducible factor, in *Proceedings of the National Academy of Sciences of the United States of America*, pp 12722–7. National Academy of Sciences.

(356) Nilsson, G. E., and Lutz, P. L. (1993) Role of GABA in hypoxia tolerance, metabolic depression and hibernation—Possible links to neurotransmitter evolution. *Comp. Biochem. Physiol. Part C Comp. Pharmacol. 105*, 329–336.

(357) Tannahill, G. M., Curtis, A. M., Adamik, J., Palsson-McDermott, E. M., McGettrick, A. F., Goel, G., Frezza, C., Bernard, N. J., Kelly, B., Foley, N. H., Zheng, L., Gardet, A., Tong, Z., Jany, S. S., Corr, S. C., Haneklaus, M., Caffrey, B. E., Pierce, K., Walmsley, S., Beasley, F. C., Cummins, E., Nizet, V., Whyte, M., Taylor, C. T., Lin, H., Masters, S. L., Gottlieb, E., Kelly, V. P., Clish, C., Auron, P. E., Xavier, R. J., and O'Neill, L. A. J. (2013) Succinate is an inflammatory signal that induces IL-1β through HIF-1α. *Nature 496*, 238–42.

(358) Hu, H., Takano, N., Xiang, L., Gilkes, D. M., Luo, W., and Semenza, G. L. (2014) Hypoxia-

inducible factors enhance glutamate signaling in cancer cells. *Oncotarget 5*, 8853–68.

(359) Binny, D. (2017) Biochemical abnormalities in OPC poisoning and its prognostic significance. *IOSR J. Dent. Med. Sci. 16*, 2279–861.

(360) Cha, Y. S., Kim, H., Go, J., Kim, T. H., Kim, O. H., Cha, K. C., Lee, K. H., and Hwang, S. O. (2014) Features of myocardial injury in severe organophosphate poisoning. *Clin. Toxicol. 52*, 873–879.

(361) Duffy, P. A., Betton, G., Horner, S., Horner, J., Cotton, P., McMahon, N., Lawrence, C., Prior, H., Armstrong, D., Philp, K., and Roberts, R. A. (2007) Biomarkers for safety and toxicology: Drug induced cardiac injury and dysfunction. *Eur. J. Cancer Suppl. 5*, 143–151.

(362) Roongsritong, C., Warraich, I., and Bradley, C. (2004) Common causes of troponin elevations in the absence of acute myocardial infarction: incidence and clinical significance. *Chest 125*, 1877–84.

(363) Repplinger, D., Su, M. K., and McKinnon, K. (2014) Troponin elevations and organophosphate poisoning: Direct cardiac injury or demand ischemia? *Clin. Toxicol. 52*, 1298–1298.

(364) Marchetti, C., Toldo, S., Chojnacki, J., Mezzaroma, E., Liu, K., Salloum, F. N., Nordio, A., Carbone, S., Mauro, A. G., Das, A., Zalavadia, A. A., Halquist, M. S., Federici, M., Van Tassell, B. W., Zhang, S., and Abbate, A. (2015) Pharmacologic Inhibition of the NLRP3 Inflammasome Preserves Cardiac Function After Ischemic and Nonischemic Injury in the Mouse. *J. Cardiovasc. Pharmacol. 66*, 1–8.

(365) El-Sheikh, A., Hashem, A., Elgohary, M., Elfadl, A., and Lashin, H. (2017) Evaluation of the potential cardiotoxic effects in acute organophosphate toxicity as a prognostic factor. *Tanta Med. J. 45*, 115.

(366) Rameh, L. E., Tolias, K. F., Duckworth, B. C., and Cantley, L. C. (1997) A new pathway for synthesis of phosphatidylinositol-4,5-bisphosphate. *Nature 390*, 192–196.

(367) Xie, H. Q., Xu, H. M., Fu, H. L., Hu, Q., Tian, W. J., Pei, X. H., and Zhao, B. (2013) AhR-mediated effects of dioxin on neuronal acetylcholinesterase expression in vitro. *Environ. Health Perspect. 121*, 613–618.

(368) Andrukhova, O., Slavic, S., Zeitz, U., Riesen, S. C., Heppelmann, M. S., Ambrisko, T. D., Markovic, M., Kuebler, W. M., and Erben, R. G. (2014) Vitamin D Is a Regulator of Endothelial Nitric Oxide Synthase and Arterial Stiffness in Mice. *Mol. Endocrinol. 28*, 53–64.

(369) Gouni-Berthold, I., Krone, W., and Berthold, H. K. (2009) Vitamin D and cardiovascular disease. *Curr. Vasc. Pharmacol. 7*, 414–22.

(370) Ghosh, R., Siddharth, M., Singh, N., Kare, P. K., Dev Banerjee, B., Wadhwa, N., and Tripathi, A. K. (2017) Organochlorine pesticide-mediated induction of NADPH oxidase and nitric-oxide synthase in endothelial cell. *J. Clin. Diagnostic Res. 11*, BC09-BC12.

(371) Sahu, B., and Maeda, A. (2016) Retinol dehydrogenases regulate vitamin A metabolism for visual function. *Nutrients 8*, 746.

(372) Marchitti, S. A., Brocker, C., Stagos, D., and Vasiliou, V. (2008) Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily. *Expert Opin. Drug Metab. Toxicol. 4*, 697–720.

(373) Sica, D. A. (2005) Alpha1-adrenergic blockers: current usage considerations. *J. Clin. Hypertens. 7*, 757–762.

(374) Slivka, S. R., Meier, K. E., and Insel, P. A. (1988) α1-Adrenergic receptors promote

phosphatidylcholine hydrolysis in MDCK-D1 cells. A mechanism for rapid activation of protein kinase C. *J. Biol. Chem. 263*, 12242–12246.

(375) Wynne, B. M., McCarthy, C. G., Szasz, T., Molina, P. A., Chapman, A. B., Webb, R. C., Klein, J. D., and Hoover, R. S. (2018) Protein kinase Cα deletion causes hypotension and decreased vascular contractility. *J. Hypertens. 36*, 510–519.

(376) Lei, B., Schwinn, D. A., and Morris, D. P. (2013) Stimulation of α1a Adrenergic Receptors Induces Cellular Proliferation or Antiproliferative Hypertrophy Dependent Solely on Agonist Concentration. *PLoS One 8*, e72430.

(377) Sun, J., Carlsson, L., Ahlberg, E., Norinder, U., Engkvist, O., and Chen, H. (2017) Applying Mondrian Cross-Conformal Prediction to Estimate Prediction Confidence on Large Imbalanced Bioactivity Data Sets. *J. Chem. Inf. Model. 57*, 1591–1598.

(378) Vovk, V., Gammerman, A., and Shafer, G. (2005) Algorithmic Learning in a Random World. Springer-Verlag Berlin.

(379) Lovering, F. (2013) Escape from Flatland 2: Complexity and promiscuity. *Medchemcomm 4*, 515–519.

(380) Méndez-Lucio, O., and Medina-Franco, J. L. (2017) The many roles of molecular complexity in drug discovery. *Drug Discov. Today 22*, 120–126.

(381) Hann, M. M., Leach, A. R., and Harper, G. (2001) Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *J. Chem. Inf. Comput. Sci. 41*, 856–864.

(382) Senger, M. R., Fraga, C. A. M., Dantas, R. F., and Silva, F. P. (2016) Filtering promiscuous compounds in early drug discovery: Is it a good idea? *Drug Discov. Today 21*, 868–872.

(383) Johnson, I., and Hahsler, M. arulesCBA: Classification for Factor and Transactional Data Sets Using Association Rules. R package version 1.1.4. https://CRAN.R-project.org/package=arulesCBA.
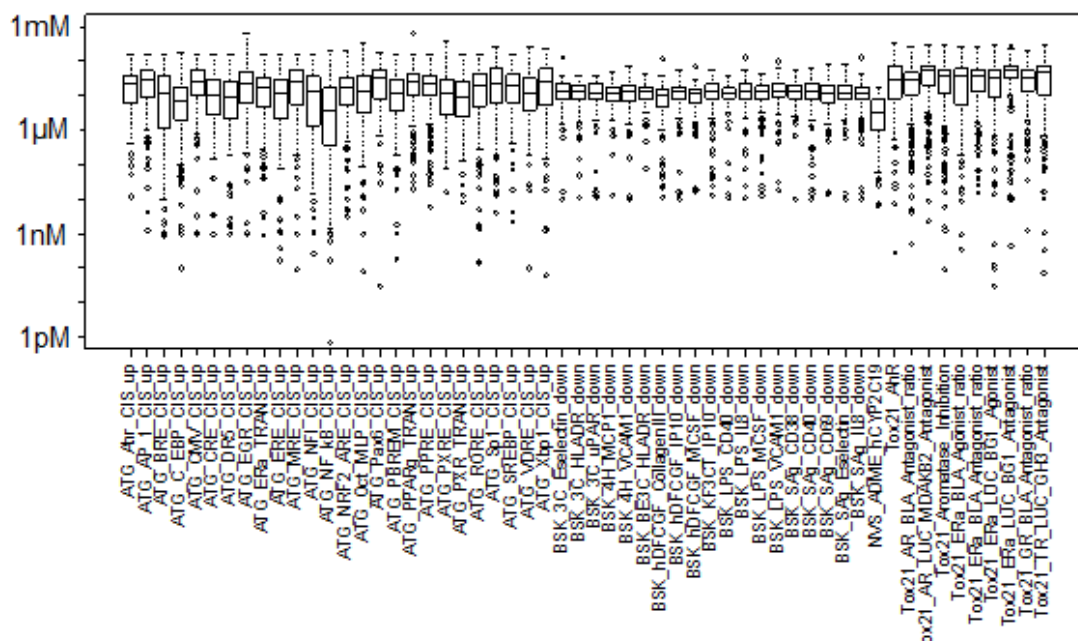
APPENDIX A

Figure (A-1) Distribution of AC$_{50}$ values in target-based assays
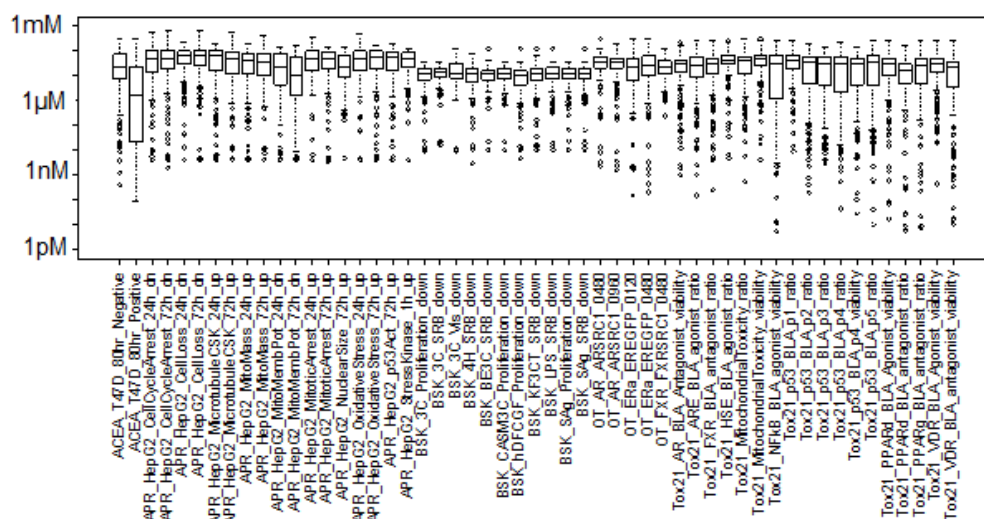
Figure (A-2) Distribution of AC$_{50}$ values in phenotypic-based assays

Table (A-1) Description of some assays, their source, class and target type. Further information on full assays annotations can be found in the ToxCast project website.

Assay abbreviation	Platform source	Assay class	Target type	Description	reference
PXR/PXRE	Attagene	Target-based	Nuclear receptor	Pregnane X receptor/ pregnane X receptor element: involved in the regulation and expression of detoxification proteins	1
DR5	Attagene	Target-based	Nuclear receptor	Response element of the retinoid acid receptor(ARE) that play an important role in the proliferation and differentiation of cells	2
PPARγ	Attagene	Target-based	Nuclear receptor	Peroxisome proliferator-activated receptor gamma is responsible for lipid metabolism/storage and adipocyte differentiation	3
NRF2/ARE	Attagene	Target-based	DNA binding	Nuclear factor erythroid 2-related factor 2 (Nrf2)-antioxidant response element (ARE): plays a key role in sensing and regulating oxidative stress by activating the expression of hundreds of antioxidant and detoxifying genes	4
AP-1	Attagene	Target-based	DNA binding	Activator protein-1 is a DNA binding protein that regulate the transcription of many proteins and it is involved in some disease pathways such as inflammatory responses	5
Eselectin	Bioseek	Target-based	Cell adhesion molecule	Selectin E protein plays a role in the adhesion of leukocytes at the site of injury as in inflammatory response	6
hLADR	Bioseek	Target-based	Cell adhesion molecule	Histocompatibility complex, class II, DR alpha protein involved in cell adhesion in immune response	7

Continue Table (A-1)

IL-8	Bioseek	Target-based	cytokine	Chemokine (C-X-C motif) ligand 8 protein that promotes angiogenesis	8
MCP-1	Bioseek	Target-based	cytokine	Chemokine (C-C motif) ligand 2 protein that is involved in neuroinflammatory response	9
uPAR	Bioseek	Target-based	cytokine	Plasminogen activator, urokinase receptor protein that regulate the activation of plasminogen system	10
VCAM-1	Bioseek	Target-based	Cell adhesion molecule	Vascular cell adhesion molecule 1 protein that induce endothelial adhesion	11
IP-10	Bioseek	Target-based	Cytokine	Chemokine (C-X-C motif) ligand 10 protein that activate T cell – endothelial adhesion and angiogenesis inhibition	12,13
CollagenIII	Bioseek	Target-based	Cell adhesion molecule	Collagen, type III, alpha 1 protein is a structural protein in connective tissues	14
SRB	Bioseek	Phenotypic-based	*Not applicable*	Quantification of the total protein levels in the system	15

References

(1) Kliewer, S. A., Goodwin, B., and Willson, T. M. (2002) The nuclear pregnane X receptor: a key regulator of xenobiotic metabolism. *Endocr. Rev. 23*, 687–702.

(2) Chambon, P. (1996) A decade of molecular biology of retinoic acid receptors. *FASEB J. 10*, 940–54.

(3) Jones, J. R., Barrick, C., Kim, K.-A., Lindner, J., Blondeau, B., Fujimoto, Y., Shiota, M., Kesterson, R. A., Kahn, B. B., and Magnuson, M. A. (2005) Deletion of PPARgamma in adipose tissues of mice protects against high fat diet-induced obesity and insulin resistance, in *Proceedings of the National Academy of Sciences of the United States of America*, pp 6207–12.

(4) Gan, L., and Johnson, J. A. (2014) Oxidative damage and the Nrf2-ARE pathway in neurodegenerative diseases. *Biochim. Biophys. Acta 1842*, 1208–18.

(5) Zenz, R., Eferl, R., Scheinecker, C., Redlich, K., Smolen, J., Schonthaler, H. B., Kenner, L., Tschachler, E., and Wagner, E. F. (2008) Activator protein 1 (Fos/Jun) functions in inflammatory bone and skin disease. *Arthritis Res. Ther. 10*, 201.

(6) Collins, T., Williams, A., Johnston, G. I., Kim, J., Eddy, R., Shows, T., Gimbrone, M. A., and Bevilacqua, M. P. (1991) Structure and chromosomal location of the gene for endothelial-leukocyte adhesion molecule 1. *J. Biol. Chem. 266*, 2466–73.

(7) Chinoy, H., Lamb, J. A., Ollier, W. E. R., and Cooper, R. G. (2009) An update on the immunogenetics of idiopathic inflammatory myopathies: major histocompatibility complex and beyond. *Curr. Opin. Rheumatol. 21*, 588–93.

(8) Strieter, R. M., Polverini, P. J., Kunkel, S. L., Arenberg, D. A., Burdick, M. D., Kasper, J., Dzuiba, J., Van Damme, J., Walz, A., Marriott, D., Chan, S.-Y., Roczniak, S., and Shanafelt, A. B. (1995) The Functional Role of the ELR Motif in CXC Chemokine-mediated Angiogenesis. *J. Biol. Chem. 270*, 27348–27357.

(9) Gerard, C., and Rollins, B. J. (2001) Chemokines and disease. *Nat. Immunol. 2*, 108–15.

(10) Behrendt, N. (2004) The urokinase receptor (uPAR) and the uPAR-associated protein (uPARAP/Endo180): membrane proteins engaged in matrix turnover during tissue remodeling. *Biol. Chem. 385*, 103–36.

(11) Lee, Y. W., Kühn, H., Hennig, B., Neish, A. S., and Toborek, M. (2001) IL-4-induced oxidative stress upregulates VCAM-1 gene expression in human endothelial cells. *J. Mol. Cell. Cardiol. 33*, 83–94.

(12) Dufour, J. H., Dziejman, M., Liu, M. T., Leung, J. H., Lane, T. E., and Luster, A. D. (2002) IFN-gamma-inducible protein 10 (IP-10; CXCL10)-deficient mice reveal a role for IP-10 in effector T cell generation and trafficking. *J. Immunol. 168*, 3195–204.

(13) Angiolillo, A. L., Sgadari, C., Taub, D. D., Liao, F., Farber, J. M., Maheshwari, S., Kleinman, H. K., Reaman, G. H., and Tosato, G. (1995) Human interferon-inducible protein 10 is a potent inhibitor of angiogenesis in vivo. *J. Exp. Med. 182*, 155–62.

(14) Di Lullo, G. A., Sweeney, S. M., Korkko, J., Ala-Kokko, L., and San Antonio, J. D. (2002) Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type I collagen. *J. Biol. Chem. 277*, 4223–31.

(15) Kavlock, R., Chandler, K., Houck, K., Hunter, S., Judson, R., Kleinstreuer, N., Knudsen, T., Martin, M., Padilla, S., Reif, D., Richard, A., Rotroff, D., Sipes, N., and Dix, D. (2012) Update on EPA's ToxCast program: Providing high throughput decision support tools for chemical risk management. *Chem. Res. Toxicol. 25*, 1287–1302.

APPENDIX B
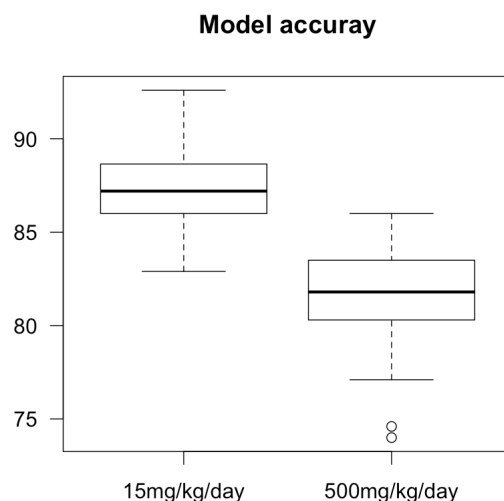
Model accuray

Figure (B-1) Distribution of rule model accuracy based on 5-fold repeated cross-validation performed in 100 trials. The average accuracy of models at 15mg/kg/day and 500mg/kg/day are 87% and 82%, respectively. Potent toxic compounds were detected at higher accuracy than less potent toxic compounds.
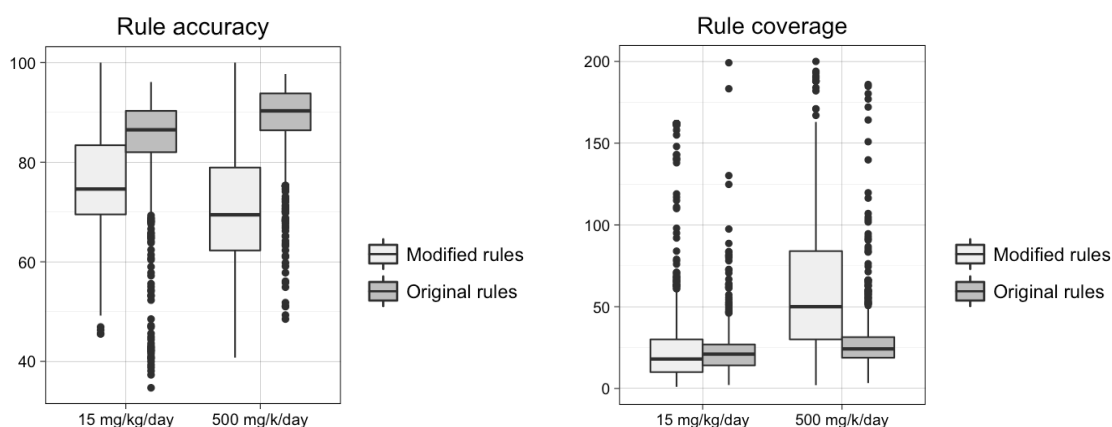
Figure (B-2) Distribution of rule accuracy (left) and coverage (right). The average accuracy of original models is higher than the corresponding modified rule set at both thresholds. Modified rules set at toxicity threshold of 15mg/kg/day have higher accuracy (~75%). Coverage here represents the number of compounds comply with the rule and correctly classified as toxic. At toxicity threshold of 500mg/kg/day, the overall coverage increased in the modified rule to around two folds in comparison to the original rule set. The average of coverage at 15mg/kg/day is not significantly different from the original set, however, it is more dispersed.

Bioactivity cutoffs of assays in selected rules

Bioactivity cutoffs are points in the continuous variables of assay measurements were split during the rule search process. These splitting points reflects the potency level in an assay that can best discriminate between toxic and non-toxic compounds. As the cutoff level varied by assay type, it can be inappropriate to set an arbitrary bioactivity cutoff to discretize assay measurements when analyzing *in vitro–in vivo* associations. For example, according to Figure B-3, compounds that are active in Cytochrome P assays at AC_{50} values of 50μM can be non-toxic, whereas, compounds possessing the same potency level against phenotypic or nuclear receptor assays can be toxic.

Additionally, potency cutoffs in assays differ with the potency of *in vivo* toxicity. Overall, the potent toxicants, *i.e.* at 15mg/kg/day, required higher potencies *in vitro*. The difference was most significant for phenotypic assays, as low potent toxicants exerted as double overall AC_{50} cutoffs as potent toxins.
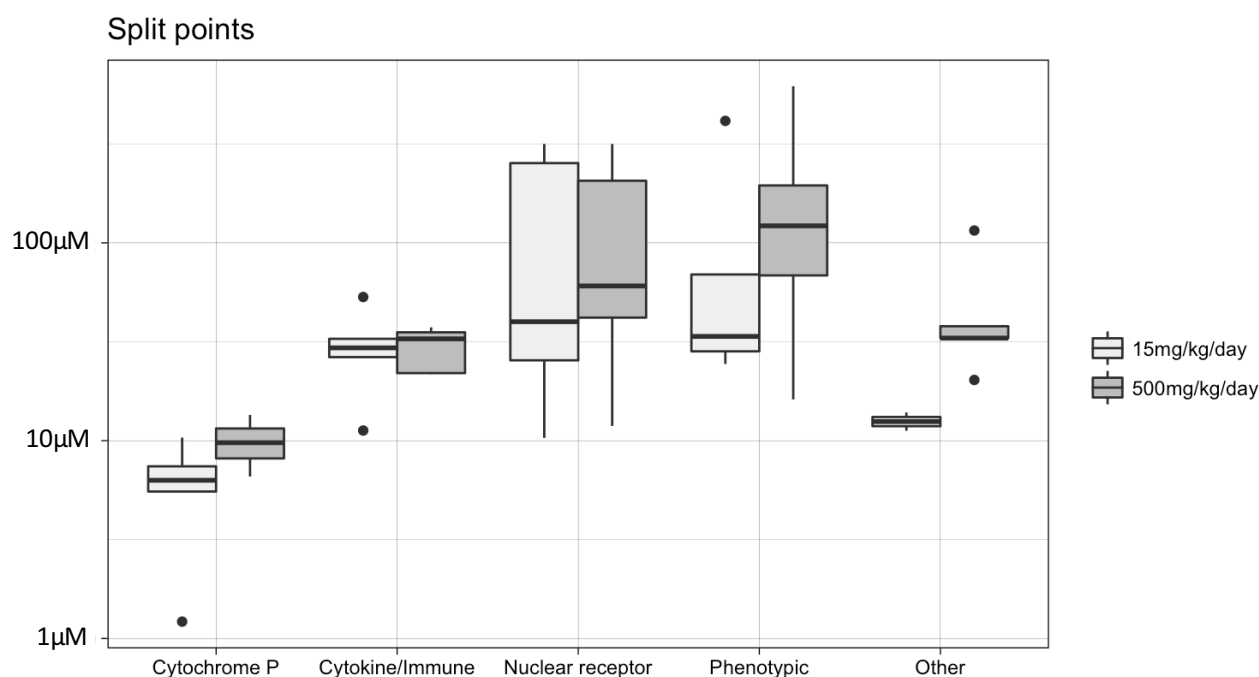
Figure (B-3) The distribution of split points of bioactivity conditions in prioritized rules. As the in vitro measurements were taken from half maximal activity concentrations (AC50) of dose-response curves, the split point of a bioactivity condition represents the potency cutoff that discriminates between toxic and non-toxic compounds. This bioactivity cutoff is lowest for cytochrome activity and highest for nuclear receptors and phenotypic activities. Consequently, hepatotoxic compounds exert high potency against Cytochrome P enzymes and moderate activities in nuclear receptor and phenotypic bioassays. This cutoff is also slightly lower among potent toxicants (LEL level of 15mg/kg/day).
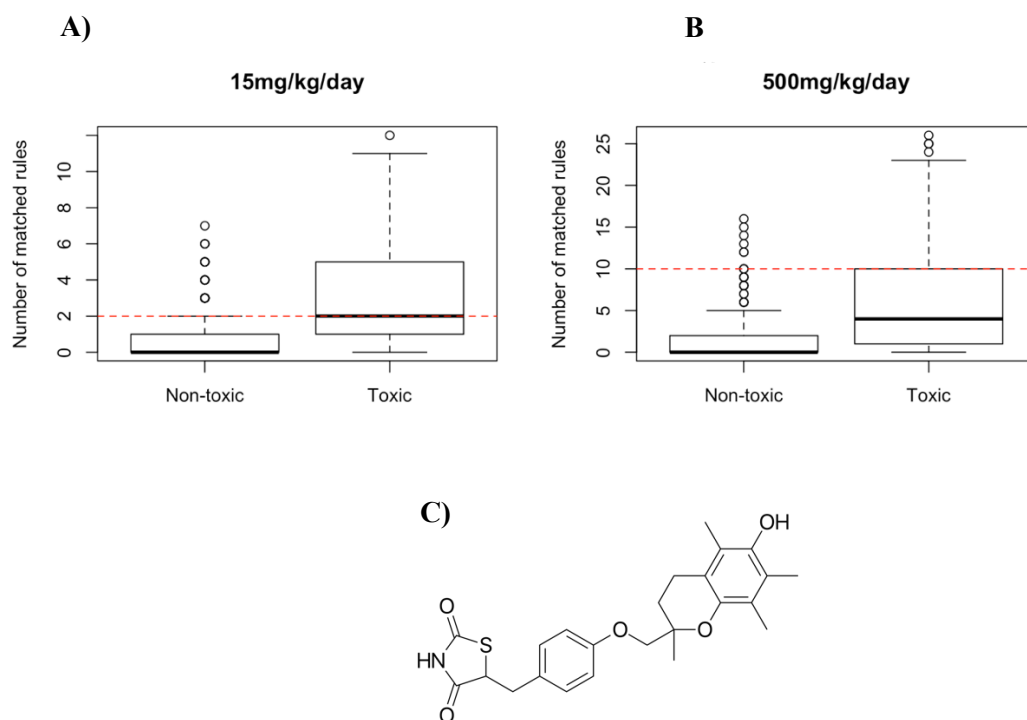
Figure (B-4) Number of rules matched by toxic and non-toxic compounds at levels of 15mg/kg/day (A) and 500mg/kg/day (B). Toxic compounds match significantly more rules than non-toxic compounds at both levels. Troglitazone (C), an antidiabetic drug withdrawn from the market, had matched two rules at 15mg/kg/day and 10 rules at 500mg/kg/day (Red dashed line). This is equivalent to average liability of toxic compounds at 15mg/kg/day level and higher than average at 500mg/kg/day, which indicate the likelihood of troglitazone to be hepatotoxic (see Table S9 for detailed rules).

Figure (B-5) Change in error rates (accuracy deterioration) in percentage as a result of removing physicochemical properties from prioritized rules, at each toxicity threshold. Rules at threshold of 15mg/kg/day had a greater overall deterioration in accuracy from removing physicochemical properties than rules at 500mg/kg/day.

Table (B-1) Histopathology endpoints on liver from ToxRefDB

(CHR=chronic/cancer; MGR=multigenerational reproductive; DEV=Prenatal developmental; SUB=Subchronic; SAC=Subacute; REP=reproductive fertility)

CHR_rat_SystemicCarcinogenic_adult_PathologyNonProliferative_AccessoryDigestive_Liver
CHR_rat_SystemicCarcinogenic_adult_PathologyProliferative_AccessoryDigestive_Liver
CHR_rat_SystemicCarcinogenic_adult_PathologyGross_AccessoryDigestive_Liver
CHR_rat_SystemicCarcinogenic_adult_PathologyNeoplastic_AccessoryDigestive_Liver
SUB_rat_SystemicCarcinogenic_adult_PathologyNonProliferative_AccessoryDigestive_Liver
SUB_rat_SystemicCarcinogenic_adult_PathologyGross_AccessoryDigestive_Liver
SUB_rat_SystemicCarcinogenic_adult_OrganWeight_AccessoryDigestive_Liver
SUB_rat_SystemicCarcinogenic_adult_PathologyProliferative_AccessoryDigestive_Liver
DEV_rat_SystemicCarcinogenic_adult_PathologyGross_AccessoryDigestive_Liver
MGR_rat_SystemicCarcinogenic_adult_PathologyNonProliferative_AccessoryDigestive_Liver
MGR_rat_SystemicCarcinogenic_adult_PathologyGross_AccessoryDigestive_Liver
MGR_rat_SystemicCarcinogenic_juvenile_PathologyNonProliferative_AccessoryDigestive_Liver
CHR_rat_SystemicCarcinogenic_adult_AccessoryDigestive_Liver
SUB_rat_SystemicCarcinogenic_adult_AccessoryDigestive_Liver
DEV_rat_SystemicCarcinogenic_adult_AccessoryDigestive_Liver
MGR_rat_SystemicCarcinogenic_adult_AccessoryDigestive_Liver
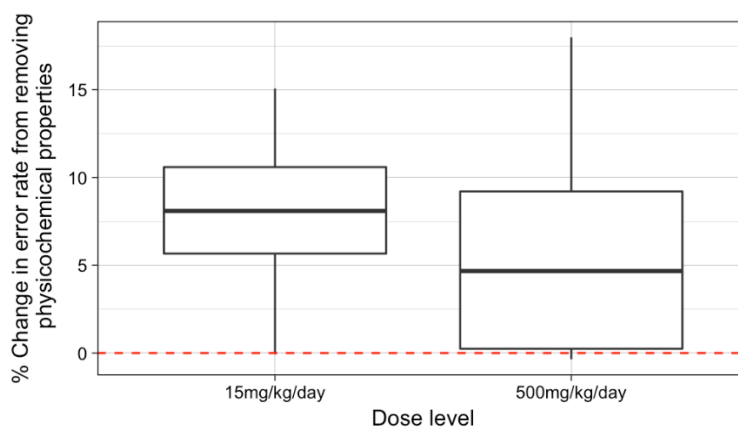MGR_rat_SystemicCarcinogenic_juvenile_AccessoryDigestive_Liver

Table (B-2) List of assays used in analysis

ACEA_T47D_80hr_Negative	BSK_LPS_PGE2_up
ACEA_T47D_80hr_Positive	BSK_LPS_SRB_down
APR_HepG2_CellCycleArrest_1h_up	BSK_LPS_TissueFactor_down
APR_HepG2_CellCycleArrest_24h_dn	BSK_LPS_TNFa_down
APR_HepG2_CellCycleArrest_24h_up	BSK_LPS_VCAM1_down
APR_HepG2_CellCycleArrest_72h_dn	BSK_SAg_CD38_down
APR_HepG2_CellCycleArrest_72h_up	BSK_SAg_CD40_down
APR_HepG2_CellLoss_24h_dn	BSK_SAg_CD69_down
APR_HepG2_CellLoss_72h_dn	BSK_SAg_Eselectin_down
APR_HepG2_MicrotubuleCSK_1h_dn	BSK_SAg_IL8_down
APR_HepG2_MicrotubuleCSK_24h_dn	BSK_SAg_MCP1_down
APR_HepG2_MicrotubuleCSK_24h_up	BSK_SAg_MIG_down
APR_HepG2_MicrotubuleCSK_72h_dn	BSK_SAg_PBMCCytotoxicity_down
APR_HepG2_MicrotubuleCSK_72h_up	BSK_SAg_PBMCCytotoxicity_up
APR_HepG2_MitoMass_1h_dn	BSK_SAg_Proliferation_down
APR_HepG2_MitoMass_1h_up	BSK_SAg_SRB_down
APR_HepG2_MitoMass_24h_dn	NVS_ADME_hCYP19A1

APR_HepG2_MitoMass_24h_up	NVS_ADME_hCYP1A1
APR_HepG2_MitoMass_72h_dn	NVS_ADME_hCYP1A2
APR_HepG2_MitoMass_72h_up	NVS_ADME_hCYP2B6
APR_HepG2_MitoMembPot_1h_dn	NVS_ADME_hCYP2C18
APR_HepG2_MitoMembPot_24h_dn	NVS_ADME_hCYP2C19
APR_HepG2_MitoMembPot_72h_dn	NVS_ADME_hCYP2C9
APR_HepG2_MitoMembPot_72h_up	NVS_ADME_hCYP2D6
APR_HepG2_MitoticArrest_1h_up	NVS_ADME_hCYP2J2
APR_HepG2_MitoticArrest_24h_up	NVS_ADME_hCYP3A4
APR_HepG2_MitoticArrest_72h_up	NVS_ADME_rCYP2A1
APR_HepG2_NuclearSize_1h_dn	NVS_ADME_rCYP2A2
APR_HepG2_NuclearSize_24h_dn	NVS_ADME_rCYP2B1
APR_HepG2_NuclearSize_24h_up	NVS_ADME_rCYP2C11
APR_HepG2_NuclearSize_72h_dn	NVS_ADME_rCYP2C12
APR_HepG2_NuclearSize_72h_up	NVS_ADME_rCYP2C13
APR_HepG2_OxidativeStress_1h_up	NVS_ADME_rCYP2C6
APR_HepG2_OxidativeStress_24h_up	NVS_ADME_rCYP2D1
APR_HepG2_OxidativeStress_72h_up	NVS_ADME_rCYP2D2
APR_HepG2_p53Act_24h_up	NVS_ADME_rCYP3A1
APR_HepG2_p53Act_72h_up	NVS_ADME_rCYP3A2
APR_HepG2_StressKinase_1h_up	NVS_ENZ_hAChE
APR_HepG2_StressKinase_24h_dn	NVS_ENZ_hBACE
APR_HepG2_StressKinase_24h_up	NVS_ENZ_hDUSP3
APR_HepG2_StressKinase_72h_dn	NVS_ENZ_hES
APR_HepG2_StressKinase_72h_up	NVS_ENZ_hGSK3b
ATG_Ahr_CIS_up	NVS_ENZ_hMMP7
ATG_AP_1_CIS_up	NVS_ENZ_hPDE4A1
ATG_BRE_CIS_up	NVS_ENZ_hPDE5
ATG_CAR_TRANS_up	NVS_ENZ_oCOX1
ATG_C_EBP_CIS_up	NVS_ENZ_oCOX2
ATG_CMV_CIS_up	NVS_ENZ_rabI2C
ATG_CRE_CIS_up	NVS_ENZ_rAChE
ATG_DR4_LXR_CIS_up	NVS_ENZ_rMAOAC
ATG_DR5_CIS_up	NVS_ENZ_rMAOAP
ATG_E2F_CIS_up	NVS_ENZ_rMAOBC
ATG_E_Box_CIS_up	NVS_ENZ_rMAOBP
ATG_EGR_CIS_up	NVS_GPCR_g5HT4
ATG_ERa_TRANS_up	NVS_GPCR_gH2
ATG_ERE_CIS_up	NVS_GPCR_gMPeripheral_NonSelective
ATG_FoxA2_CIS_up	NVS_GPCR_gOpiateK
ATG_FXR_TRANS_up	NVS_GPCR_h5HT5A
ATG_GLI_CIS_up	NVS_GPCR_h5HT6

ATG_HIF1a_CIS_up	NVS_GPCR_h5HT7
ATG_HSE_CIS_up	NVS_GPCR_hAdoRA1
ATG_IR1_CIS_up	NVS_GPCR_hAdoRA2a
ATG_LXRa_TRANS_up	NVS_GPCR_hAdra2C
ATG_LXRb_TRANS_up	NVS_GPCR_hDRD1
ATG_M_06_CIS_up	NVS_GPCR_hDRD2s
ATG_M_19_CIS_up	NVS_GPCR_hM1
ATG_M_19_TRANS_up	NVS_GPCR_hM2
ATG_M_32_CIS_up	NVS_GPCR_hM3
ATG_M_32_TRANS_up	NVS_GPCR_hM4
ATG_M_61_CIS_up	NVS_GPCR_hNK2
ATG_MRE_CIS_up	NVS_GPCR_hOpiate_mu
ATG_Myc_CIS_up	NVS_GPCR_p5HT2C
ATG_NFI_CIS_up	NVS_GPCR_rabPAF
ATG_NF_kB_CIS_up	NVS_GPCR_rmAdra2B
ATG_NRF1_CIS_up	NVS_GPCR_rOpiate_NonSelective
ATG_NRF2_ARE_CIS_up	NVS_IC_rCaBTZCHL
ATG_NURR1_TRANS_up	NVS_IC_rCaDHPRCh_L
ATG_Oct_MLP_CIS_up	NVS_IC_rNaCh_site2
ATG_p53_CIS_up	NVS_MP_hPBR
ATG_Pax6_CIS_up	NVS_MP_rPBR
ATG_PBREM_CIS_up	NVS_NR_bER
ATG_PPARa_TRANS_up	NVS_NR_bPR
ATG_PPARg_TRANS_up	NVS_NR_cAR
ATG_PPRE_CIS_up	NVS_NR_hAR
ATG_PXRE_CIS_up	NVS_NR_hCAR_Antagonist
ATG_PXR_TRANS_up	NVS_NR_hER
ATG_RARa_TRANS_up	NVS_NR_hFXR_Antagonist
ATG_RARb_TRANS_up	NVS_NR_hGR
ATG_RARg_TRANS_up	NVS_NR_hPPARa
ATG_RORE_CIS_up	NVS_NR_hPPARg
ATG_RXRb_TRANS_up	NVS_NR_hPR
ATG_Sox_CIS_up	NVS_NR_hPXR
ATG_Sp1_CIS_up	NVS_NR_hTRa
ATG_SREBP_CIS_up	NVS_NR_mERa
ATG_STAT3_CIS_up	NVS_NR_rAR
ATG_TA_CIS_up	NVS_OR_gSIGMA_NonSelective
ATG_TAL_CIS_up	NVS_TR_gDAT
ATG_TGFb_CIS_up	NVS_TR_hDAT
ATG_THRa1_TRANS_up	NVS_TR_hNET
ATG_VDRE_CIS_up	NVS_TR_hSERT
ATG_Xbp1_CIS_up	NVS_TR_rSERT

BSK_3C_Eselectin_down	NVS_TR_rVMAT2
BSK_3C_HLADR_down	OT_AR_ARELUC_AG_1440
BSK_3C_ICAM1_down	OT_AR_ARSRC1_0480
BSK_3C_IL8_down	OT_AR_ARSRC1_0960
BSK_3C_MCP1_down	OT_ERa_EREGFP_0120
BSK_3C_MIG_down	OT_ERa_EREGFP_0480
BSK_3C_Proliferation_down	OT_ER_ERaERa_0480
BSK_3C_SRB_down	OT_ER_ERaERa_1440
BSK_3C_Thrombomodulin_up	OT_ER_ERaERb_0480
BSK_3C_TissueFactor_down	OT_ER_ERaERb_1440
BSK_3C_uPAR_down	OT_ER_ERbERb_0480
BSK_3C_VCAM1_down	OT_ER_ERbERb_1440
BSK_3C_Vis_down	OT_FXR_FXRSRC1_0480
BSK_4H_Eotaxin3_down	OT_FXR_FXRSRC1_1440
BSK_4H_MCP1_down	OT_NURR1_NURR1RXRa_0480
BSK_4H_Pselectin_down	OT_PPARg_PPARgSRC1_1440
BSK_4H_SRB_down	Tox21_AhR
BSK_4H_uPAR_down	Tox21_AR_BLA_Agonist_ratio
BSK_4H_VCAM1_down	Tox21_AR_BLA_Antagonist_ratio
BSK_4H_VEGFRII_down	Tox21_AR_BLA_Antagonist_viability
BSK_BE3C_HLADR_down	Tox21_ARE_BLA_agonist_ratio
BSK_BE3C_IL1a_down	Tox21_AR_LUC_MDAKB2_Agonist
BSK_BE3C_IP10_down	Tox21_AR_LUC_MDAKB2_Antagonist
BSK_BE3C_MIG_down	Tox21_Aromatase_Inhibition
BSK_BE3C_MMP1_down	Tox21_AutoFluor_HEK293_Cell_blue
BSK_BE3C_MMP1_up	Tox21_AutoFluor_HEK293_Cell_green
BSK_BE3C_PAI1_down	Tox21_AutoFluor_HEK293_Cell_red
BSK_BE3C_SRB_down	Tox21_AutoFluor_HEK293_Media_blue
BSK_BE3C_SRB_up	Tox21_AutoFluor_HEK293_Media_green
BSK_BE3C_TGFb1_down	Tox21_AutoFluor_HEK293_Media_red
BSK_BE3C_tPA_down	Tox21_AutoFluor_HEPG2_Cell_blue
BSK_BE3C_uPA_down	Tox21_AutoFluor_HEPG2_Cell_green
BSK_BE3C_uPAR_down	Tox21_AutoFluor_HEPG2_Media_blue
BSK_BE3C_uPAR_up	Tox21_AutoFluor_HEPG2_Media_green
BSK_CASM3C_HLADR_down	Tox21_AutoFluor_HEPG2_Media_red
BSK_CASM3C_IL6_down	Tox21_ELG1_LUC_Agonist
BSK_CASM3C_IL6_up	Tox21_ERa_BLA_Agonist_ratio
BSK_CASM3C_IL8_down	Tox21_ERa_BLA_Antagonist_ratio
BSK_CASM3C_IL8_up	Tox21_ERa_BLA_Antagonist_viability
BSK_CASM3C_LDLR_down	Tox21_ERa_LUC_BG1_Agonist
BSK_CASM3C_LDLR_up	Tox21_ERa_LUC_BG1_Antagonist
BSK_CASM3C_MCP1_down	Tox21_ESRE_BLA_ratio

BSK_CASM3C_MCSF_down	Tox21_ESRE_BLA_viability
BSK_CASM3C_MIG_down	Tox21_FXR_BLA_agonist_ratio
BSK_CASM3C_Proliferation_down	Tox21_FXR_BLA_agonist_viability
BSK_CASM3C_SAA_down	Tox21_FXR_BLA_antagonist_ratio
BSK_CASM3C_SRB_down	Tox21_FXR_BLA_antagonist_viability
BSK_CASM3C_Thrombomodulin_up	Tox21_GR_BLA_Agonist_ratio
BSK_CASM3C_uPAR_down	Tox21_GR_BLA_Antagonist_ratio
BSK_CASM3C_VCAM1_down	Tox21_GR_BLA_Antagonist_viability
BSK_hDFCGF_CollagenIII_down	Tox21_HSE_BLA_agonist_ratio
BSK_hDFCGF_EGFR_down	Tox21_HSE_BLA_agonist_viability
BSK_hDFCGF_IL8_down	Tox21_MitochondrialToxicity_ratio
BSK_hDFCGF_IP10_down	Tox21_MitochondrialToxicity_viability
BSK_hDFCGF_MCSF_down	Tox21_NFkB_BLA_agonist_ratio
BSK_hDFCGF_MIG_down	Tox21_NFkB_BLA_agonist_viability
BSK_hDFCGF_MMP1_down	Tox21_p53_BLA_p1_ratio
BSK_hDFCGF_MMP1_up	Tox21_p53_BLA_p1_viability
BSK_hDFCGF_PAI1_down	Tox21_p53_BLA_p2_ratio
BSK_hDFCGF_Proliferation_down	Tox21_p53_BLA_p2_viability
BSK_hDFCGF_SRB_down	Tox21_p53_BLA_p3_ratio
BSK_hDFCGF_TIMP1_down	Tox21_p53_BLA_p3_viability
BSK_hDFCGF_VCAM1_down	Tox21_p53_BLA_p4_ratio
BSK_KF3CT_ICAM1_down	Tox21_p53_BLA_p4_viability
BSK_KF3CT_IL1a_down	Tox21_p53_BLA_p5_ratio
BSK_KF3CT_IP10_down	Tox21_p53_BLA_p5_viability
BSK_KF3CT_MCP1_down	Tox21_PPARd_BLA_agonist_ratio
BSK_KF3CT_MMP9_down	Tox21_PPARd_BLA_Agonist_viability
BSK_KF3CT_SRB_down	Tox21_PPARd_BLA_antagonist_ratio
BSK_KF3CT_SRB_up	Tox21_PPARd_BLA_antagonist_viability
BSK_KF3CT_TGFb1_down	Tox21_PPARg_BLA_Agonist_ratio
BSK_KF3CT_TIMP2_down	Tox21_PPARg_BLA_antagonist_ratio
BSK_KF3CT_uPA_down	Tox21_PPARg_BLA_antagonist_viability
BSK_LPS_CD40_down	Tox21_TR_LUC_GH3_Agonist
BSK_LPS_Eselectin_down	Tox21_TR_LUC_GH3_Antagonist
BSK_LPS_IL1a_down	Tox21_VDR_BLA_agonist_ratio
BSK_LPS_IL8_down	Tox21_VDR_BLA_Agonist_viability
BSK_LPS_MCP1_down	Tox21_VDR_BLA_antagonist_ratio
BSK_LPS_MCSF_down	Tox21_VDR_BLA_antagonist_viability
BSK_LPS_PGE2_down	

Table (B-3) List of physicochemical used in analysis

Physicochemical property	Source
SlogP	RDKit
SMR	RDKit
LabuteASA	RDKit
TPSA	RDKit
AMW	RDKit
ExactMW	RDKit
NumLipinskiHBA	RDKit
NumLipinskiHBD	RDKit
NumRotatableBonds	RDKit
NumHBD	RDKit
NumHBA	RDKit
NumAmideBonds	RDKit
NumHeteroAtoms	RDKit
NumHeavyAtoms	RDKit
NumAtoms	RDKit
NumRings	RDKit
NumAromaticRings	RDKit
NumSaturatedRings	RDKit
NumAliphaticRings	RDKit
NumAromaticHeterocycles	RDKit
NumSaturatedHeterocycles	RDKit
NumAliphaticHeterocycles	RDKit
NumAromaticCarbocycles	RDKit
NumSaturatedCarbocycles	RDKit
NumAliphaticCarbocycles	RDKit
FractionCSP3	RDKit
HallKierAlpha	RDKit
Strongest.acidic.pKa	ChemAxon
Strongest.basic.pKa	ChemAxon

Table (B-4) Rules matching Troglitazone describing toxicity at level of 500mg/kg/day

Rules	Condition 1	Condition 2	Condition 3	Rule accuracy	Compound coverage
1	ATG_p53_CIS_up <= 2.122144	Tox21_MitochondrialToxicity_viability <= 2.5	-	0.95	51
2	APR_HepG2_MitoMass_24h_up <= 2.319333	OT_AR_ARSRC1_0480 <= 2.5	NumLipinskiHBA > 1	0.92	66
3	ATG_PPARg_TRANS_up <= 1.694441	OT_AR_ARSRC1_0480 <= 2.5	NumRings > 0	0.87	78
4	OT_AR_ARSRC1_0480 <= 1.806514	NumHeavyAtoms <= 33	NumAromaticCarbocycles > 0	0.86	83
5	BSK_SAg_SRB_down <= 2.10206	NumHeavyAtoms <= 33	NumAromaticCarbocycles > 0	0.81	100
6	BSK_BE3C_uPA_down <= 2.06236	NumHeavyAtoms <= 33	NumAromaticHeterocycles <= 0	0.79	53
7	BSK_SAg_CD40_down <= 1.341604	NumHeteroAtoms > 1	NumHeavyAtoms <= 32	0.77	90
8	BSK_3C_IL8_down <= 1.336627	BSK_LPS_CD40_down <= 1.572855	-	0.75	56
9	APR_HepG2_CellCycleArrest_72h_dn <= 2.791415	Tox21_FXR_BLA_antagonist_ratio <= 2.40309	-	0.75	96
10	OT_FXR_FXRSRC1_0480 <= 2.073822	NumHeavyAtoms <= 33	NumAromaticCarbocycles > 0	0.75	113

Table (B-5) Rules matching Troglitazone describing toxicity at level of 15mg/kg/day

Rules	Condition 1	Condition 2	Condition 3	Rule accuracy	Compound coverage
1	BSK_4H_MCP1_down <= 1.422321	NumRotatableBonds <= 6	NumAliphaticCarbocycles <=0	0.81	40
2	BSK_hDFCGF_IP10_down <= 1.726166	Tox21_MitochondrialToxicity_viability <= 1.581255		0.89	25

APPENDIX C

Table (C-1) Tox21 assays used in emerging patterns in Chapter 5 and 6.

Assay description	Uniprot	Pubchem Assay ID
1. qHTS assay for small molecule agonists of the p53 signaling pathway	P04637	651631
2. qHTS assay for small molecule agonists of the p53 signaling pathway - cell viability	NA	651633
3. qHTS assay for small molecule agonists of the p53 signaling pathway: Summary	P04637	720552
4. qHTS assay for small molecules that induce genotoxicity in human embryonic kidney cells expressing luciferase-tagged ATAD5	Q96QE3	651632
5. qHTS assay for small molecules that induce genotoxicity in human embryonic kidney cells expressing luciferase-tagged ATAD5 - cell viability	NA	651634
6. qHTS assay for small molecules that induce genotoxicity in human embryonic kidney cells expressing luciferase-tagged ATAD5: Summary	Q96QE3	720516
7. qHTS assay for small molecule disruptors of the mitochondrial membrane potential - cell viability	NA	720634
8. qHTS assay for small molecule disruptors of the mitochondrial membrane potential	NA	720635
9. qHTS assay for small molecule disruptors of the mitochondrial membrane potential: Summary	NA	720637
10. qHTS assay to test for compound auto fluorescence at 460 nm (blue) in HEK293 cells	NA	720678
11. qHTS assay to test for compound auto fluorescence at 460 nm (blue) in HEK293 cell free culture	NA	720681
12. qHTS assay to test for compound auto fluorescence at 460 nm (blue) in HepG2 cell free culture	NA	720685
13. qHTS assay to test for compound auto fluorescence at 460 nm (blue) in HepG2 cells	NA	720687
14. qHTS assay to identify small molecule agonists of the glucocorticoid receptor (GR) signaling pathway	P04150	720691
15. qHTS assay to identify small molecule antagonists of the glucocorticoid receptor (GR) signaling pathway	P04150	720692
16. qHTS assay to identify small molecule antagonists of the glucocorticoid receptor (GR) signaling pathway - cell viability counter screen	NA	720693
17. qHTS assay to identify small molecule agonists of the glucocorticoid receptor (GR) signaling pathway: Summary	P04150	720719
18. qHTS assay to identify small molecule antagonists of the glucocorticoid receptor (GR) signaling pathway: Summary	P04150	720725
19. qHTS assay to identify small molecule antagonists of the androgen receptor (AR) signaling pathway - cell viability counter screen	NA	743033
20. qHTS assay to identify small molecule antagonists of the androgen receptor (AR) signaling pathway	P10275	743035
21. qHTS assay to identify small molecule agonists of the androgen receptor (AR) signaling pathway	P10275	743036

22. qHTS assay to identify small molecule antagonists of the androgen receptor (AR) signaling pathway using the MDA cell line - cell viability counter screen	NA	743041
23. qHTS assay to identify small molecule antagonists of the androgen receptor (AR) signaling pathway using the MDA cell line	P10275	743042
24. qHTS assay to identify small molecule agonists of the androgen receptor (AR) signaling pathway: Summary	P10275	743053
25. qHTS assay to identify small molecule antagonists of the androgen receptor (AR) signaling pathway using the MDA cell line: Summary	P10275	743054
26. qHTS assay to identify small molecule antagonists of the androgen receptor (AR) signaling pathway: Summary	P10275	743063
27. qHTS assay to identify small molecule antagonists of the thyroid receptor (TR) signaling pathway - cell viability counter screen	NA	743064
28. qHTS assay to identify small molecule antagonists of the thyroid receptor (TR) signaling pathway	NA	743065
29. qHTS assay to identify small molecule antagonists of the thyroid receptor (TR) signaling pathway: Summary	NA	743067
30. qHTS assay to identify small molecule antagonists of the estrogen receptor alpha (ER-alpha) signaling pathway	P03372	743069
31. qHTS assay to identify small molecule antagonists of the estrogen receptor alpha (ER-alpha) signaling pathway - cell viability counter screen	NA	743074
32. qHTS assay to identify small molecule agonists of the estrogen receptor alpha (ER-alpha) signaling pathway	P03372	743075
33. qHTS assay to identify small molecule agonists of the estrogen receptor alpha (ER-alpha) signaling pathway: Summary	P03372	743077
34. qHTS assay to identify small molecule antagonists of the estrogen receptor alpha (ER-alpha) signaling pathway: Summary	P03372	743078
35. qHTS assay to identify small molecule antagonists of the estrogen receptor alpha (ER-alpha) signaling pathway using the BG1 cell line	P03372	743080
36. qHTS assay to identify small molecule antagonists of the estrogen receptor alpha (ER-alpha) signaling pathway using the BG1 cell line - cell viability counter screen	NA	743081
37. qHTS assay to identify aromatase inhibitors	P11511	743083
38. qHTS assay to identify aromatase inhibitors - cell viability counter screen	NA	743084
39. qHTS assay to identify small molecule that activate the aryl hydrocarbon receptor (AhR) signaling pathway	P35869	743085
40. qHTS assay to identify small molecule that activate the aryl hydrocarbon receptor (AhR) signaling pathway - cell viability counter screen	NA	743086
41. qHTS assay to identify small molecule antagonists of the estrogen receptor alpha (ER-alpha) signaling pathway using the BG1 cell line: Summary	P03372	743091
42. qHTS assay to identify small molecule agonists of the peroxisome proliferator-activated receptor gamma (PPARg) signaling pathway	P37231	743094
43. qHTS assay to identify small molecule that activate the aryl hydrocarbon receptor (AhR) signaling pathway: Summary	P35869	743122
44. qHTS assay to identify aromatase inhibitors: Summary	P11511	743139
45. qHTS assay to identify small molecule agonists of the peroxisome proliferator-activated receptor gamma (PPARg) signaling pathway: Summary	P37231	743140

46. qHTS assay to identify small molecule antagonists of the retinoid-related orphan receptor gamma (ROR-gamma) signaling pathway - cell viability counter screen	NA	1159520
47. qHTS assay to identify small molecule antagonists of the retinoid-related orphan receptor gamma (ROR-gamma) signaling pathway	P51450	1159521
48. qHTS assay to identify small molecule antagonists of the retinoid-related orphan receptor gamma (ROR-gamma) signaling pathway: Summary	P51450	1159523
49. qHTS assay to identify small molecule agonists of the AP-1 signaling pathway - cell viability counter screen	NA	1159525
50. qHTS assay to identify small molecule agonists of the AP-1 signaling pathway	P05412	1159526
51. qHTS assay to identify small molecule agonists of the RXR signaling pathway	P19793	1159527
52. qHTS assay to identify small molecule agonists of the AP-1 signaling pathway: Summary	P05412	1159528
53. qHTS assay to identify small molecule agonists of the RXR signaling pathway - cell viability counter screen	NA	1159529
54. qHTS assay to identify small molecule agonists of the RXR signaling pathway: Summary	P19793	1159531
55. qHTS assay to identify small molecule antagonists of the retinoid acid receptor (RAR) signaling pathway - cell viability counter screen	NA	1159551
56. qHTS assay to identify small molecule antagonists of the retinoic acid receptor (RAR) signaling pathway	P10276	1159552
57. qHTS assay to identify small molecule antagonists of the retinoic acid receptor (RAR) signaling pathway: Summary	P10276	1159555
58. qHTS assay to identify small molecule agonists of the constitutive androstane receptor (CAR) signaling pathway - cell viability counter screen	NA	1224836
59. qHTS assay to identify small molecule antagonists of the constitutive androstane receptor (CAR) signaling pathway - cell viability counter screen	NA	1224837
60. qHTS assay to identify small molecule antagonists of the constitutive androstane receptor (CAR) signaling pathway	Q14994	1224838
61. qHTS assay to identify small molecule agonists of the constitutive androstane receptor (CAR) signaling pathway	Q14994	1224839
62. qHTS assay to identify small molecule agonists of the thyroid stimulating hormone receptor (TSHR) signaling pathway	P16473	1224843
63. qHTS assay to identify small molecule agonists of the hypoxia (HIF-1) signaling pathway - cell viability counter screen	NA	1224844
64. qHTS assay to identify small molecule agonists of H2AX	NA	1224845
65. qHTS assay to identify small molecule agonists of the hypoxia (HIF-1) signaling pathway	Q16665	1224846
66. qHTS assay to identify small molecule agonists of H2AX - cell viability counter screen	NA	1224847
67. qHTS assay to identify small molecule agonists of the constitutive androstane receptor (CAR) signaling pathway: Summary	Q14994	1224892
68. qHTS assay to identify small molecule antagonists of the constitutive androstane receptor (CAR) signaling pathway: Summary	Q14994	1224893
69. qHTS assay to identify small molecule agonists of the hypoxia (HIF-1) signaling pathway: Summary	Q16665	1224894
70. qHTS assay to identify small molecule agonists of the thyroid stimulating hormone receptor (TSHR) signaling pathway: Summary	P16473	1224895

71. qHTS assay to identify small molecule agonists of H2AX: Summary	NA	1224896
72. qHTS assay to identify small molecule antagonists of the estrogen receptor alpha (ER-alpha) signaling pathway using the BG1 cell line in the presence of 0.1 nM 17-beta-estradiol - cell viability counter screen	NA	1259241
73. qHTS assay to identify small molecule antagonists of the androgen receptor (AR) signaling pathway using the MDA cell line in the presence of 0.5 nM R1881 - cell viability counter screen	NA	1259242
74. qHTS assay to identify small molecule antagonists of the androgen receptor (AR) signaling pathway using the MDA cell line in the presence of 0.5 nM R1881	P10275	1259243
75. qHTS assay to identify small molecule antagonists of the estrogen receptor alpha (ER-alpha) signaling pathway using the BG1 cell line in the presence of 0.1 nM 17-beta-estradiol	P03372	1259244
76. qHTS assay to identify small molecule antagonists of the androgen receptor (AR) signaling pathway using the MDA cell line in the presence of 0.5 nM R1881: Summary	P10275	1259247
77. qHTS assay to identify small molecule antagonists of the estrogen receptor alpha (ER-alpha) signaling pathway using the BG1 cell line in the presence of 0.1 nM 17-beta-estradiol: Summary	P03372	1259248
78. qHTS assay to identify small molecule agonists of the NFkB signaling pathway	P19838	1159509
79. qHTS assay to identify small molecule agonists of the NFkB signaling pathway - cell viability counter screen	NA	1159515
80. qHTS assay to identify small molecule agonists of the endoplasmic reticulum stress response signaling pathway	P18850	1159516
81. qHTS assay to identify small molecule agonists of the endoplasmic reticulum stress response signaling pathway - cell viability counter screen	NA	1159517
82. qHTS assay to identify small molecule agonists of the NFkB signaling pathway: Summary	P19838	1159518
83. qHTS assay to identify small molecule agonists of the endoplasmic reticulum stress response signaling pathway: Summary	P18850	1159519
84. qHTS assay to identify small molecule antagonists of the peroxisome proliferator-activated receptor gamma (PPARg) signaling pathway	P37231	743191
85. qHTS assay to identify small molecule antagonists of the peroxisome proliferator-activated receptor gamma (PPARg) signaling pathway - cell viability counter screen	NA	743194
86. qHTS assay to identify small molecule antagonists of the peroxisome proliferator-activated receptor gamma (PPARg) signaling pathway: Summary	P37231	743199
87. qHTS assay for small molecule agonists of the antioxidant response element (ARE) signaling pathway	Q16236	743202
88. qHTS assay for small molecule agonists of the antioxidant response element (ARE) signaling pathway - cell viability counter screen	NA	743203
89. qHTS assay for small molecule activators of the heat shock response signaling pathway - cell viability counter screen	NA	743209
90. qHTS assay for small molecule activators of the heat shock response signaling pathway	P04792	743210
91. qHTS assay to identify small molecule agonists of the peroxisome proliferator-activated receptor delta (PPARd) signaling pathway - cell viability counter screen	NA	743211
92. qHTS assay to identify small molecule agonists of the peroxisome proliferator-activated receptor delta (PPARd) signaling pathway	Q03181	743212

|		
93. qHTS assay to identify small molecule antagonists of the peroxisome proliferator-activated receptor delta (PPARd) signaling pathway - cell viability counter screen	NA	743213
94. qHTS assay to identify small molecule antagonists of the peroxisome proliferator-activated receptor delta (PPARd) signaling pathway	Q03181	743215
95. qHTS assay to identify small molecule antagonists of the farnesoid-X-receptor (FXR) signaling pathway	Q96RI1	743217
96. qHTS assay to identify small molecule agonists of the farnesoid-X-receptor (FXR) signaling pathway - cell viability counter screen	NA	743218
97. qHTS assay for small molecule agonists of the antioxidant response element (ARE) signaling pathway: Summary	Q16236	743219
98. qHTS assay to identify small molecule agonists of the farnesoid-X-receptor (FXR) signaling pathway	Q96RI1	743220
99. qHTS assay to identify small molecule antagonists of the farnesoid-X-receptor (FXR) signaling pathway - cell viability counter screen	NA	743221
100. qHTS assay to identify small molecule agonists of the vitamin D receptor (VDR) signaling pathway	P11473	743222
101. qHTS assay to identify small molecule antagonists of the vitamin D receptor (VDR) signaling pathway	P11473	743223
102. qHTS assay to identify small molecule agonists of the vitamin D receptor (VDR) signaling pathway - cell viability counter screen	NA	743224
103. qHTS assay to identify small molecule antagonists of the vitamin D receptor (VDR) signaling pathway - cell viability counter screen	NA	743225
104. qHTS assay to identify small molecule antagonists of the peroxisome proliferator-activated receptor delta (PPARd) signaling pathway: Summary	Q03181	743226
105. qHTS assay to identify small molecule agonists of the peroxisome proliferator-activated receptor delta (PPARd) signaling pathway: Summary	Q03181	743227
106. qHTS assay for small molecule activators of the heat shock response signaling pathway: Summary	P04792	743228
107. qHTS assay to identify small molecule agonists of the farnesoid-X-receptor (FXR) signaling pathway: Summary	Q96RI1	743239
108. qHTS assay to identify small molecule antagonists of the farnesoid-X-receptor (FXR) signaling pathway: Summary	Q96RI1	743240
109. qHTS assay to identify small molecule agonists of the vitamin D receptor (VDR) signaling pathway: Summary	P11473	743241
110. qHTS assay to identify small molecule antagonists of the vitamin D receptor (VDR) signaling pathway: Summary	P11473	743242
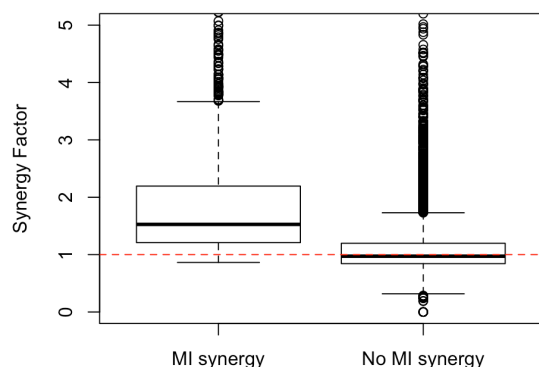
Figure (C-1) The correlation between the two synergy measures used in the study, mutual information (MI) and synergy factor. Threshold for MI-based synergy was set to 0.001 value. Pairs which did not exert MI synergy has shown synergy factor values less than 1.5.
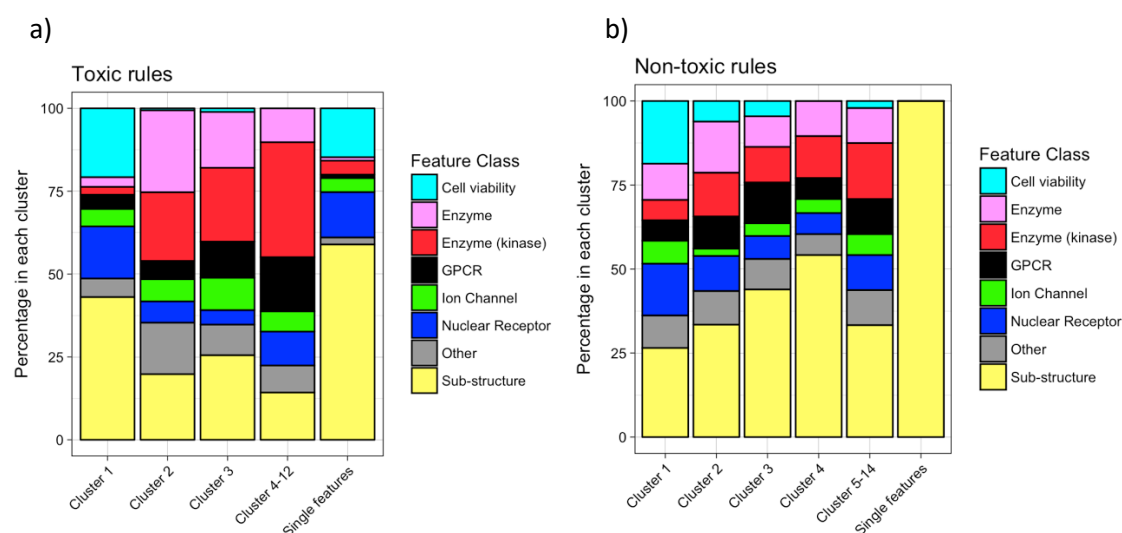

Figure (C-2) Percentage distribution of feature classes in each cluster as well as single condition feature rules in toxic (a) and non-toxic (b) networks. In toxic network, single feature rules and cluster 1, were more occupied by cell viability, nuclear receptor activity and structural features. The remaining clusters had mainly target-specific features such as enzymes, kinases, ion channels and GPCR. Clusters of non-toxic network had more diverse distribution of classes compared to toxic network.
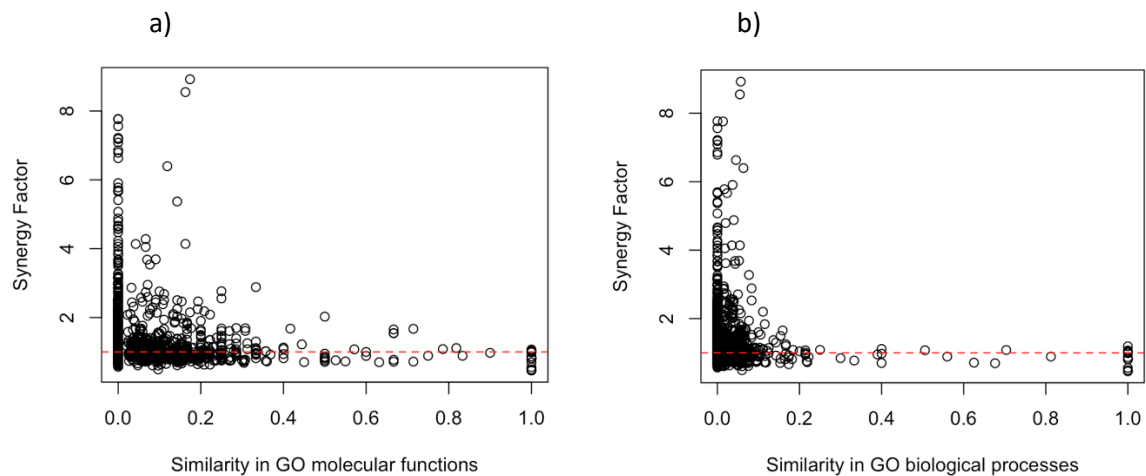
Figure (C-3) Similarity between target-specific pairs using Gene Ontology (GO) (a) molecular functions and (b) biological processes with respect to the synergy of the pair. Synergistic pairs show almost zero similarity in function and biological processes. However, not all dissimilar pair are synergistic.