In Silico Toxicology and Vertebrate-Free Safety Assessments for Plant Protection Products

Sarah Whalley and Richard Currie

In Silico Toxicology, 30th September 2020

Agenda

- Plant protection products vs pharmaceuticals
- Bringing products to the market
- Toxicology – past and present
- In silico – where does it fit? What are our challenges?
- DARTable genome
- Workflows to enable mechanistic toxicology
  - SYNTOX tool
  - Cell line selection
  - Quantitative AOP generation

What is a Plant Protection Product?

Herbicides

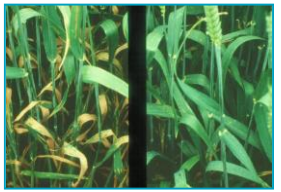
- control weeds that compete with crops for light and nutrients

Insecticides

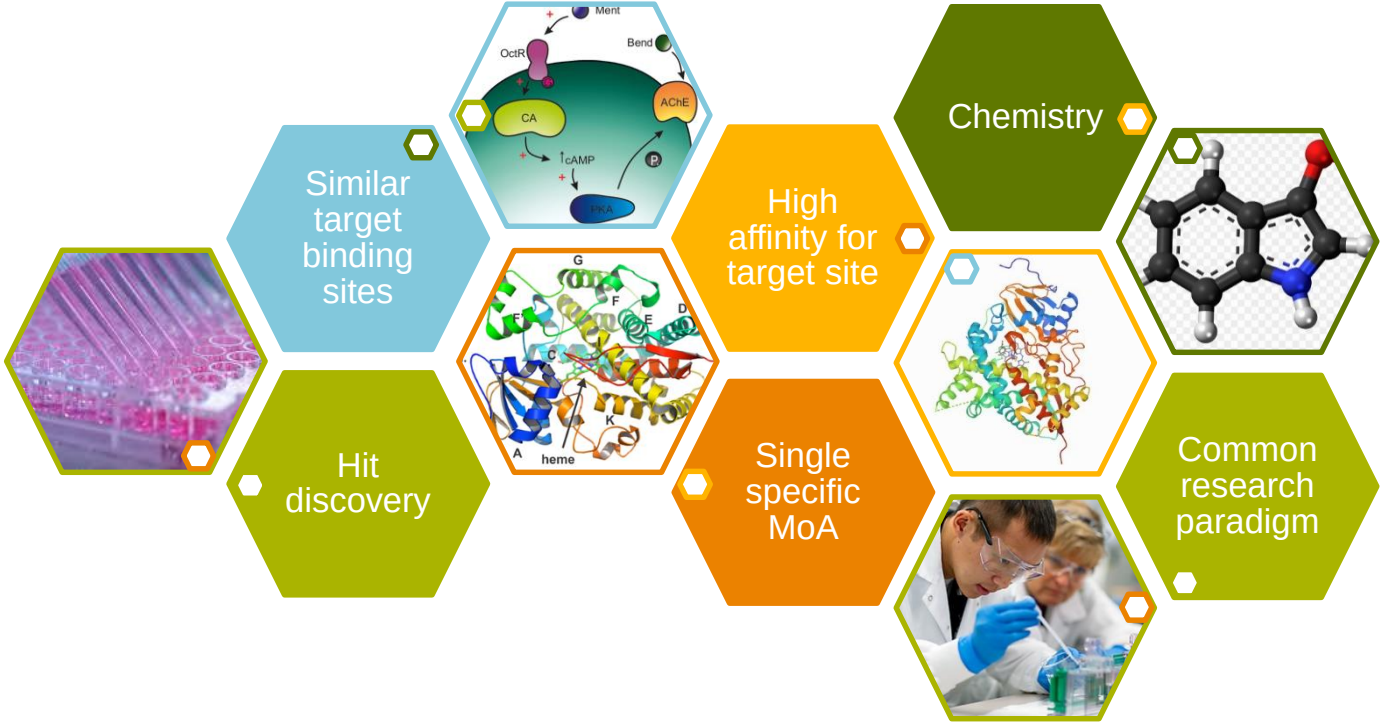
- control pests which reduce yields by damaging crops

Fungicides

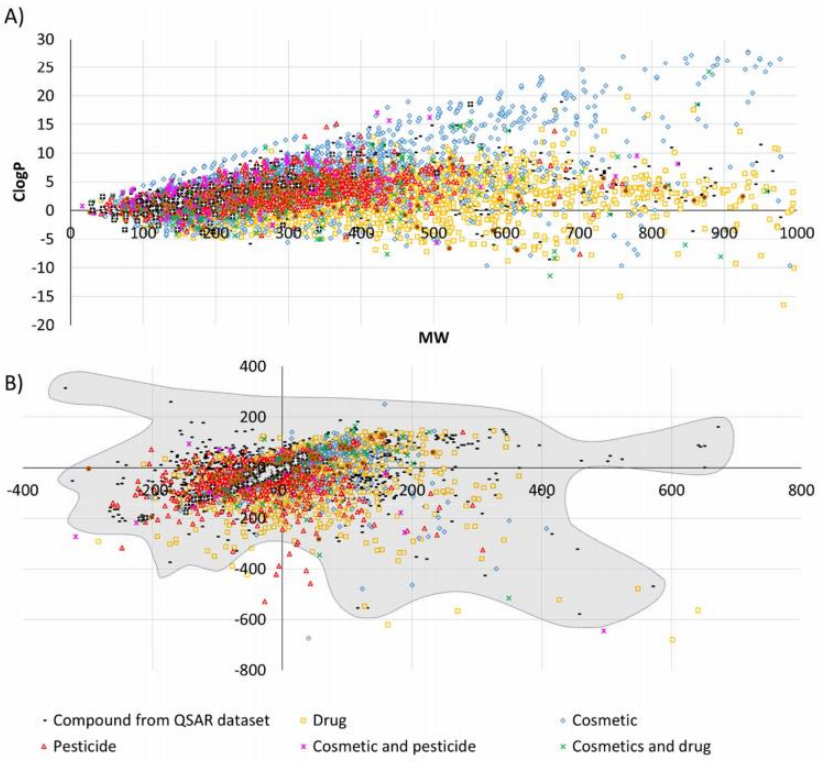
- prevent and cure fungal disease



Plant Protection Products (PPP) and Pharmaceuticals - similarities



Chemical Space – dispelling the myth?

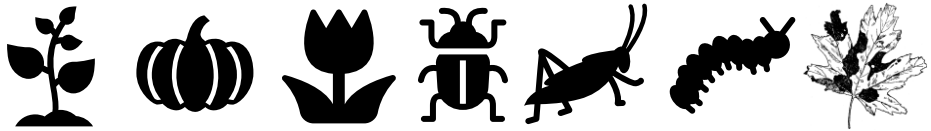
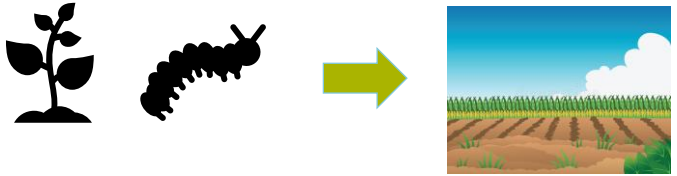
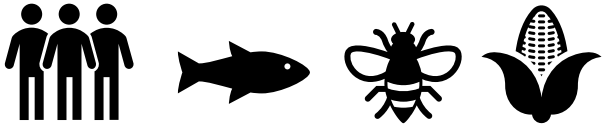


Chemical space of investigated compounds defined by ClogP and MW

Chemical space of investigated compounds in barycentric coordinates obtained from 2D DRAGON descriptors. Shadowed area represent the chemical space occupied by compounds from datasets used to generate current toxicity QSAR models

Alves et al, Food Chem Toxicol. 2018 February ; 112: 526–534. doi:10.1016/j.fct.2017.04.008

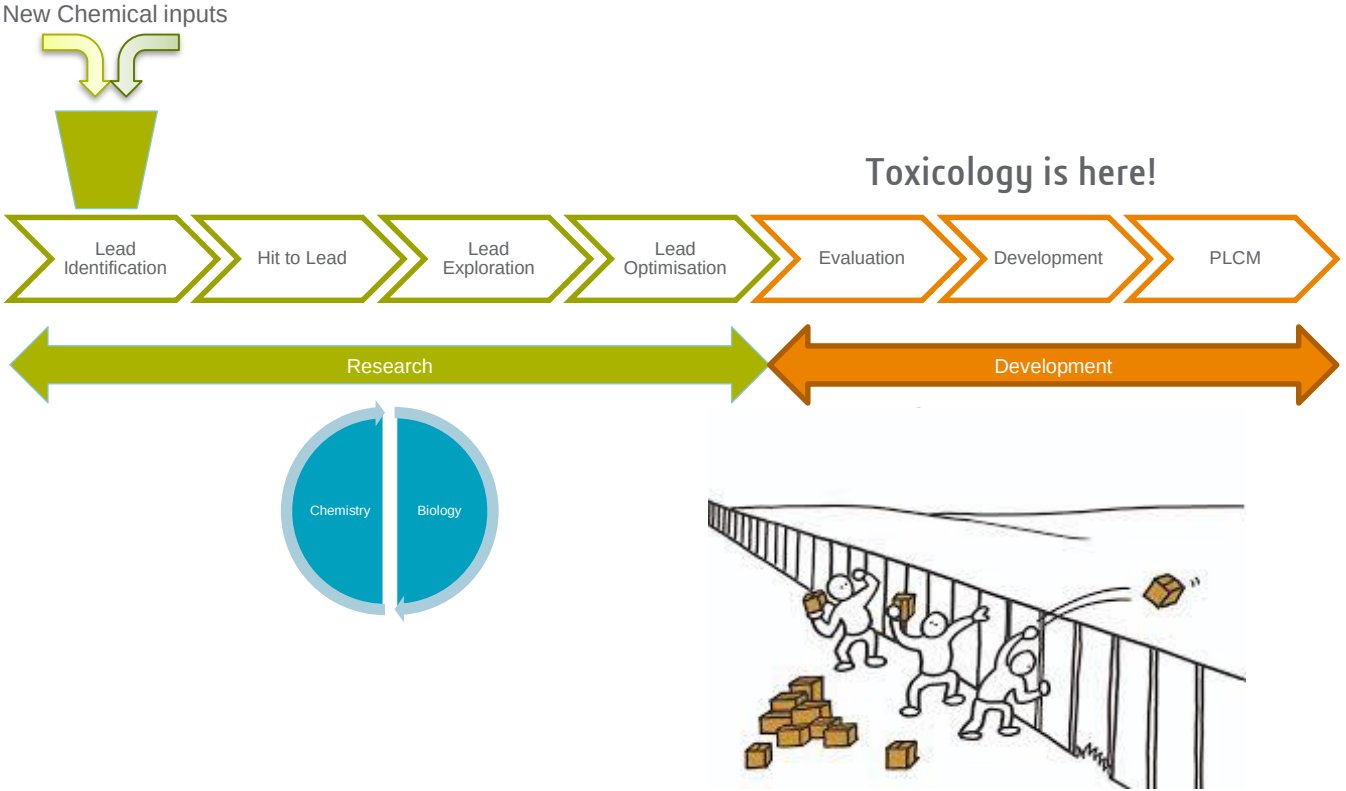
Plant Protection Products and Pharmaceuticals - differences

	Pharmaceutical	Plant Protection Product
Target Species		
Efficacy testing		
Species Selectivity	?	

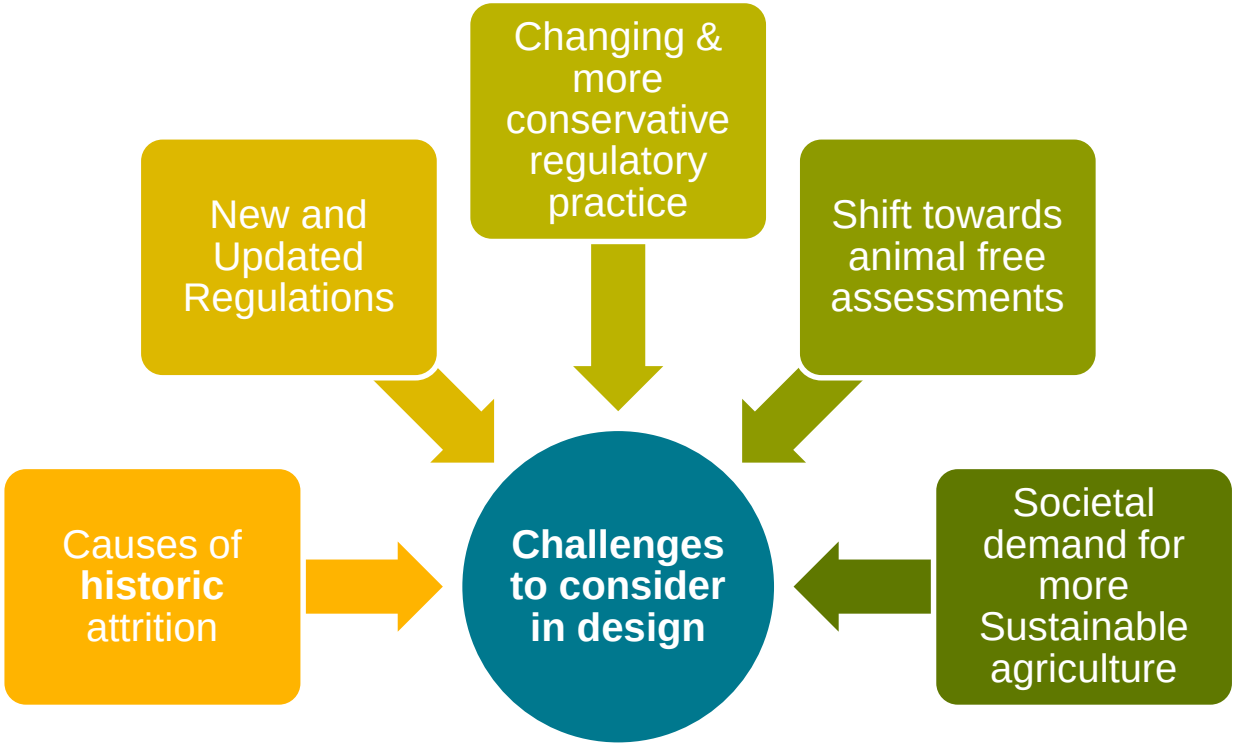
Bringing new products to the market has become a multi-faceted challenge…

Mammalian Toxicology Knowledge Required

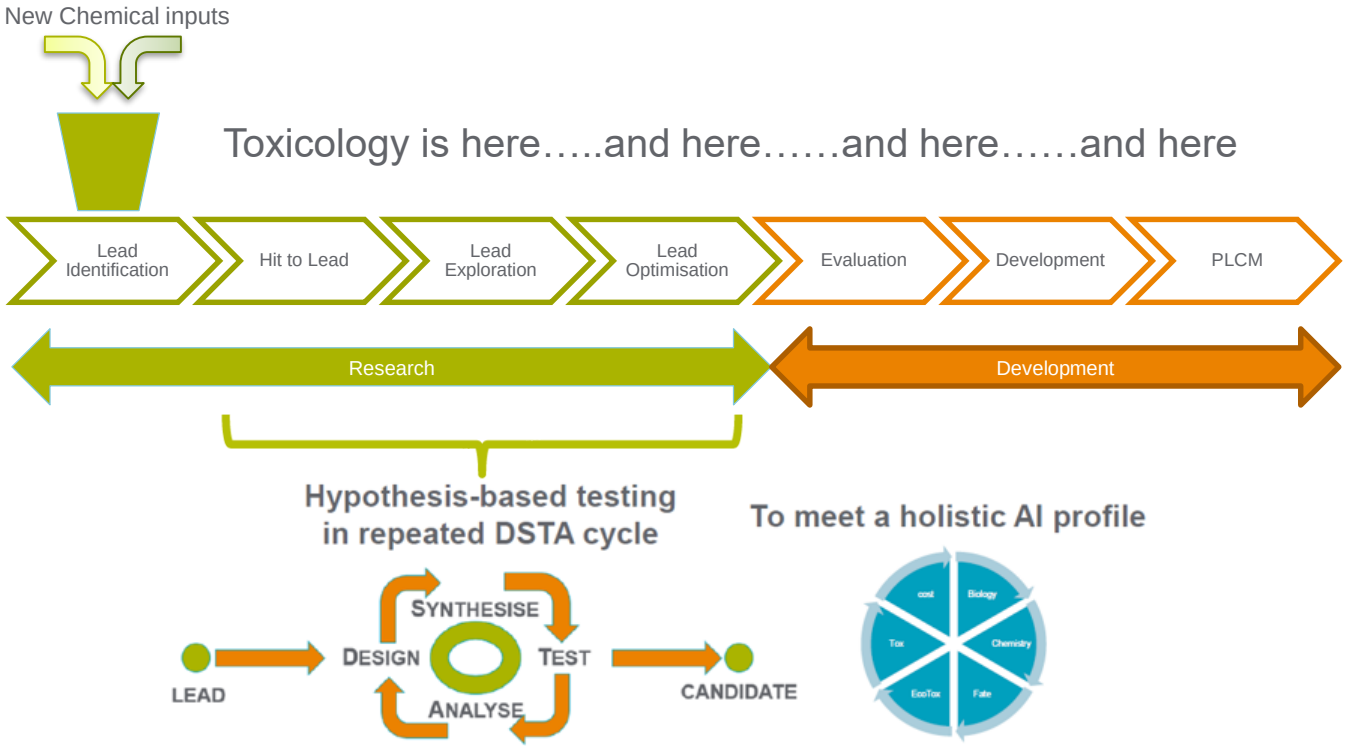
Plant Protection Product R&D: A Traditional View of the R&D Pipeline



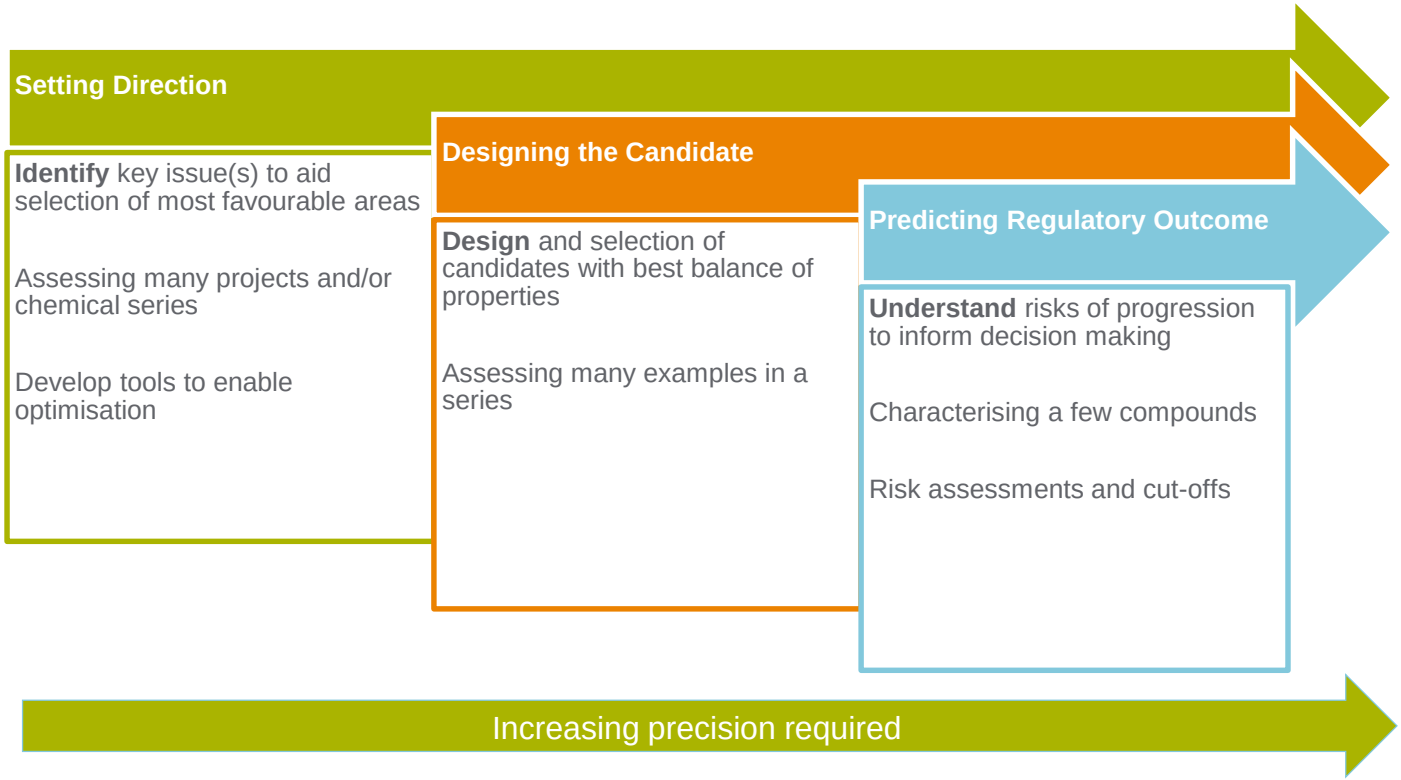
How has this traditional view changed in recent years?



Plant Protection Product R&D: The reality of a modern R&D Pipeline



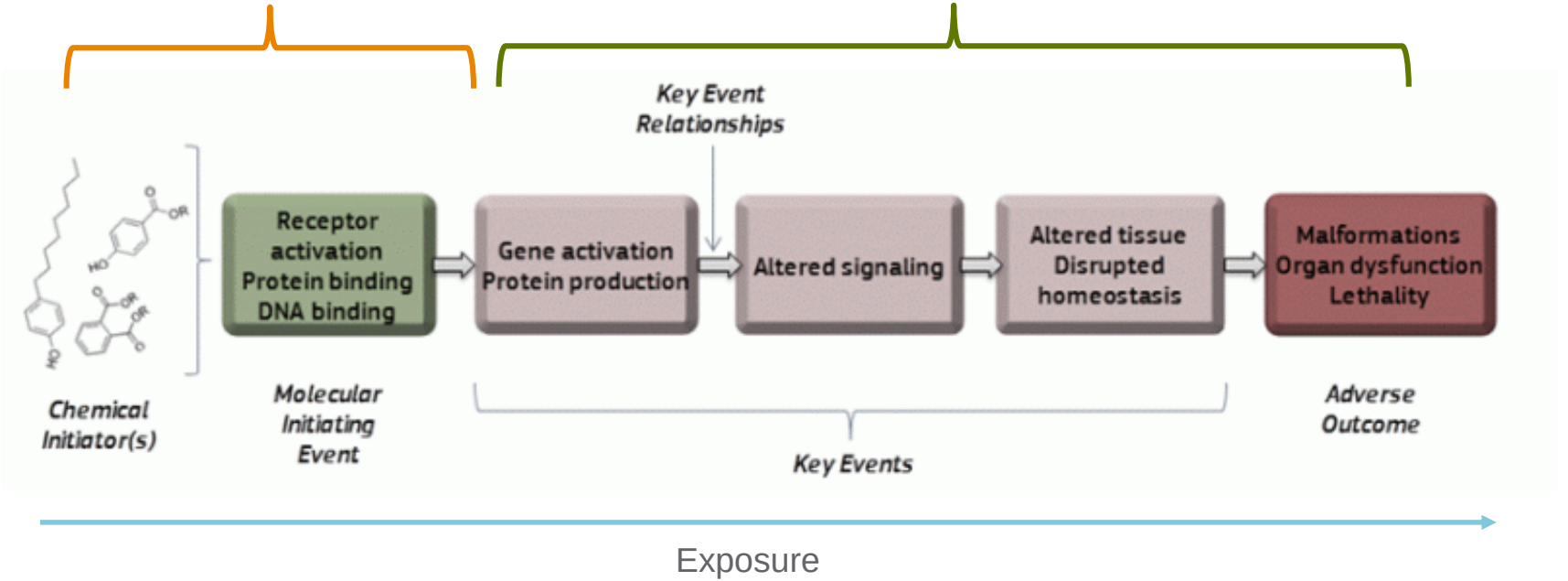
Three Phases Prior to Development



How can we predict toxicity earlier? Developing Adverse Outcome Pathways

Features that change with **chemistry**

Features that change with **biology**



syngenta

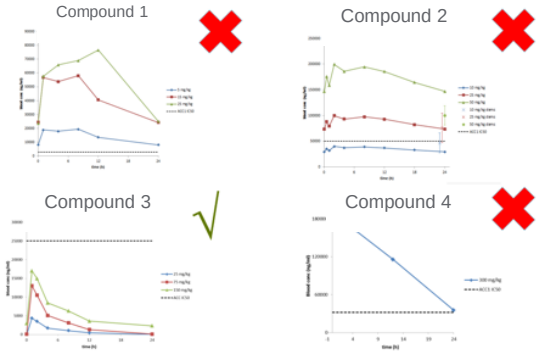
Setting Direction – Developmental Toxicity

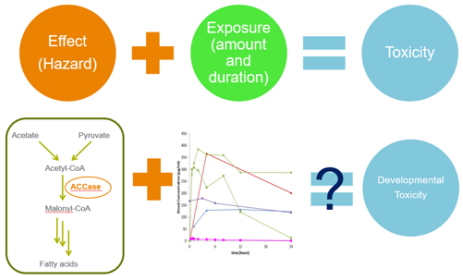
Amenable to DSTA cycles

In vivo Dev Tox Study

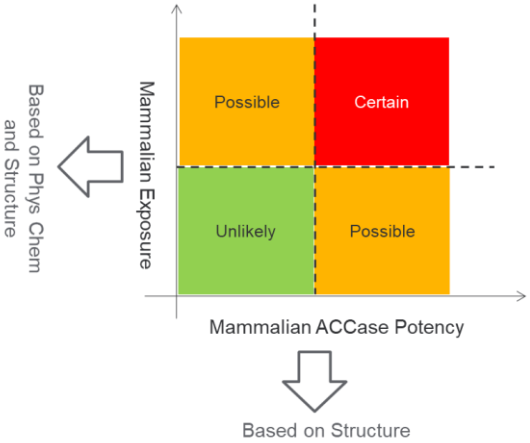
In vitro studies for exposure and potency

In silico predictions





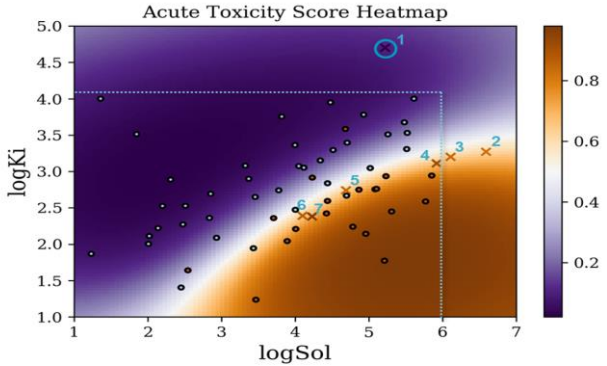
Hypothesis Generation



Setting Direction – Acute Toxicity

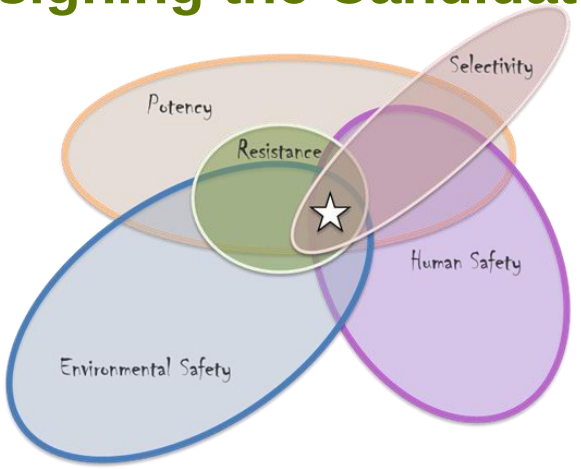
Amenable to DSTA cycles

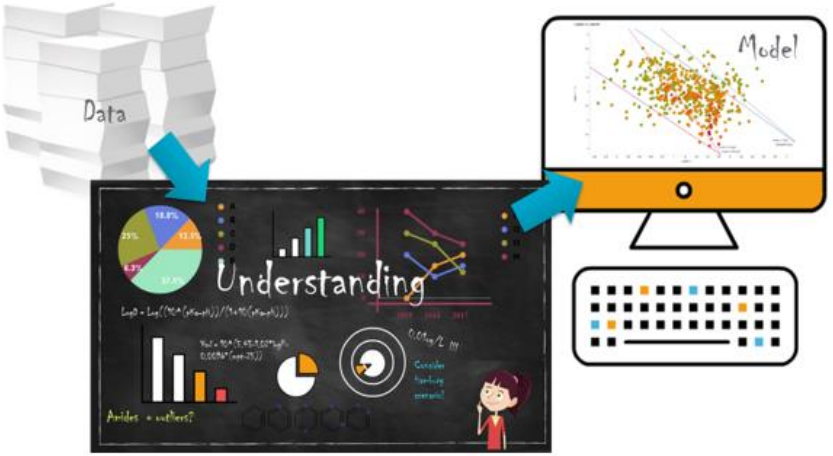
In vivo Acute tox Study → In vitro studies for exposure and potency → In silico predictions

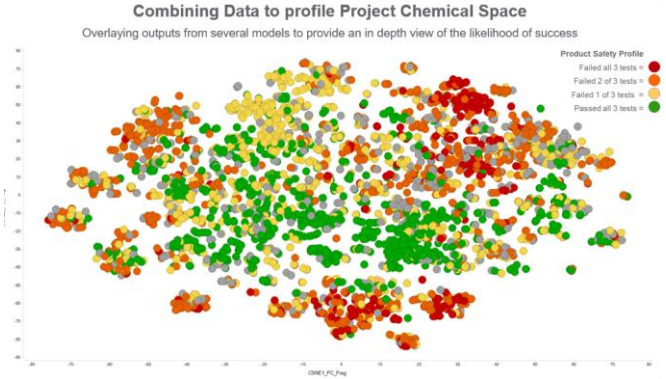


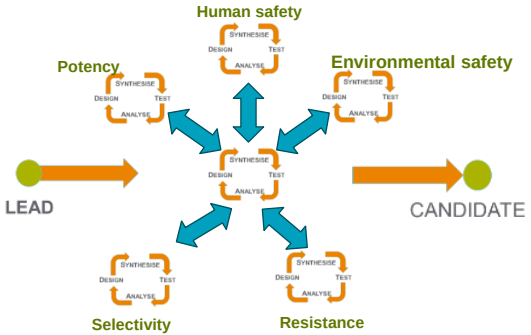
Inputs to model are either *in vitro* measurements or categorical predictions from QSAR models

Designing the Candidate

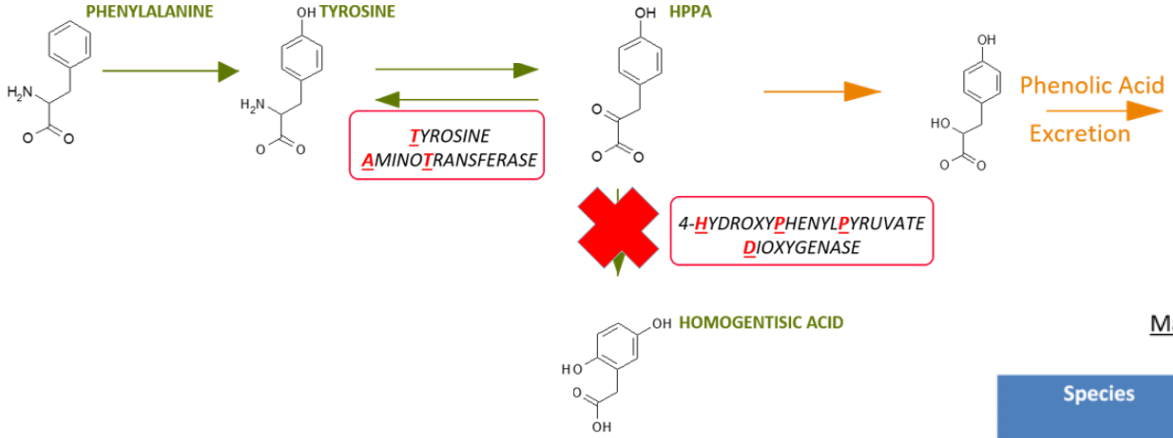








Predicting Regulatory Outcomes: 4-hydroxyphenylpyruvate dioxygenase (HPPD) Inhibitors

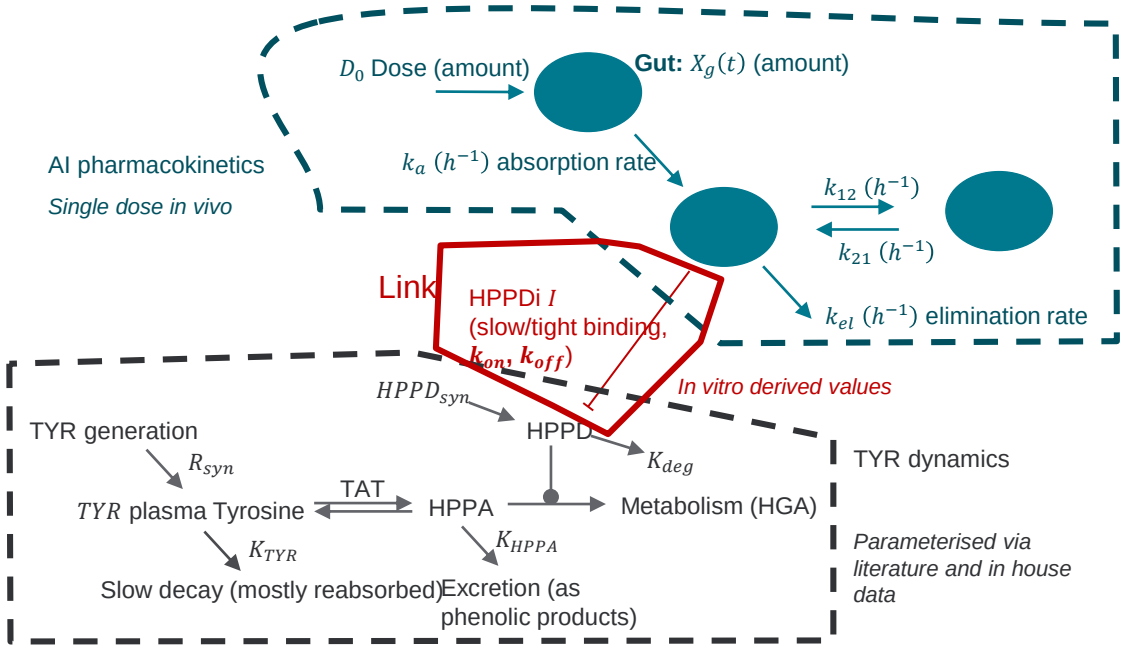


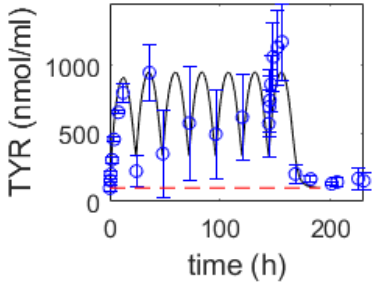
- Discovered for herbicidal use in ~1980 (13 in class; 1 pharma)
- HPPD inhibitors are capable of binding & inhibiting HPPD in rat, mouse and human (and plants!)
- Dose dependent increases in whole blood tyrosine are observed
- Tyrosine is the toxicophore – clear, consistent spectrum of toxicities associated with elevated tyrosine

Main information available (numbers of chemicals in red)

Species	*in silico*	*in vitro*	*in vivo*
Plants	Crystal structures HPPD protein sequences and homology modelling	HPPD inhibition assay data ('00s)	Herbicidal potency ('000s)
Mouse			Assay (kinetic and tyrosine data, '00s) Toxicity studies ('0s)
Rat		HPPD inhibition assay data ('0s)	Assay (kinetic and tyrosine data, '0s) Toxicity studies ('0s)
Human			Kinetic and tyrosine data and safety record (1)

The mechanistic PK-PD modelling approach for HPPD inhibitors

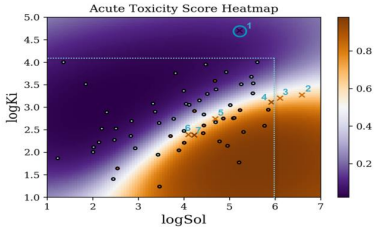




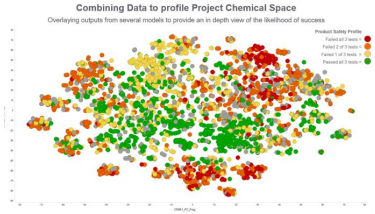
- Predict thresholds of concern on repeat dosing
- Explore species differences
- Extend PK to PBPK to allow
  - organ level exposure
  - species to species extrapolation
  - Develop in silico inputs for model (ADME and in vitro binding)

Summary so far……

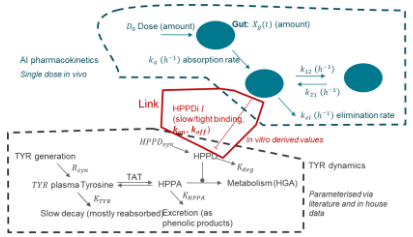
Setting Direction



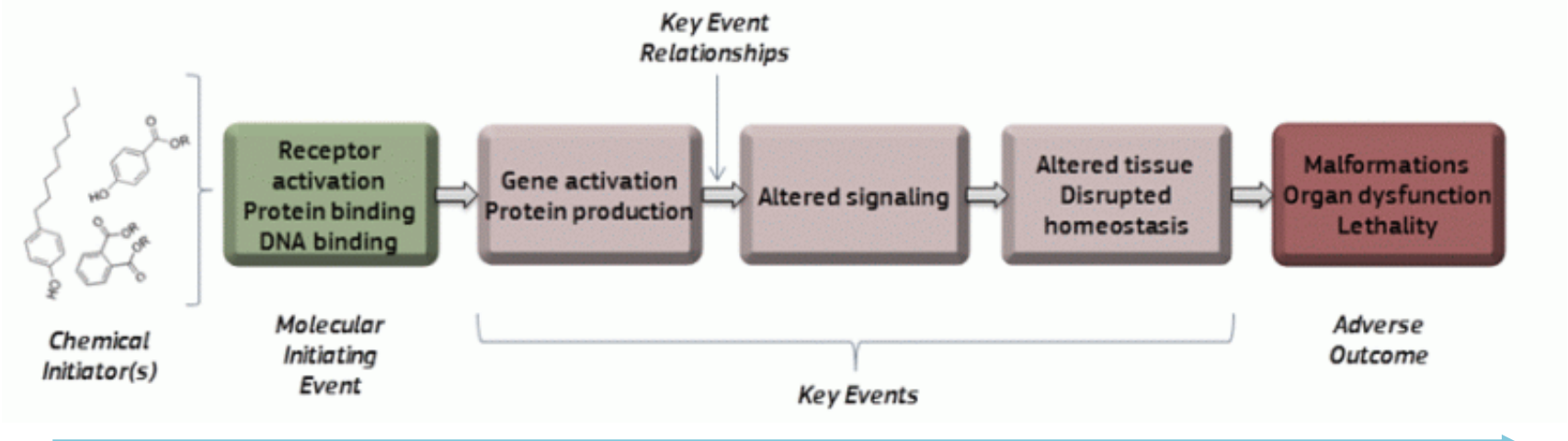
Designing the Candidate



Predicting Regulatory Outcome



How can we predict toxicity earlier? Developing Adverse Outcome Pathways



Pesticidal targets

What else?

NAMs that help de-risk critical design challenges during R&D

NAMs that meet future regulatory need

syngenta

EPA 2035 Challenge: iterative exploration of NAMs to meet their regulatory needs

September 2019 - a directive to prioritize EPA's efforts to **reduce animal testing including reducing mammal study requests and funding 30 percent by 2025 and eliminating them by 2035**. …….the Agency will continue to rely on the **development and application of new approach methodologies (NAMs)**, which refer to any technology, methodology, approach, or combination that can provide information on chemical hazard and risk assessment to avoid the use of animal testing.

Figure 1. Five work plan objectives towards achieving the EPA mammalian testing reduction goals.

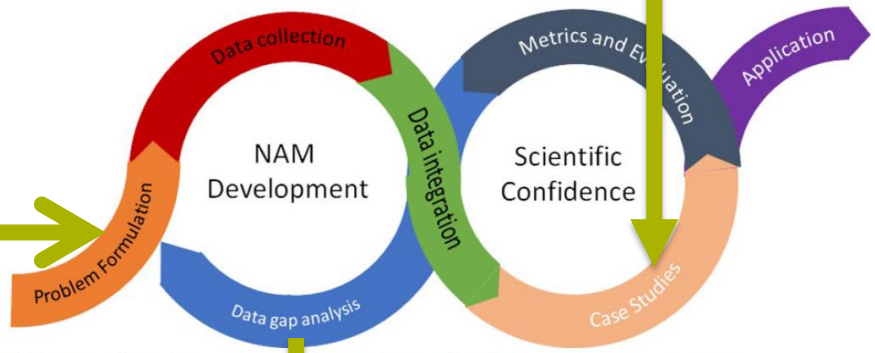


Figure 2. Problem-focused research planning and implementation process at EPA.

What are our biggest challenges in deploying these new approaches?

Access to data to form hypotheses, build models, identify gaps and develop alternative approaches

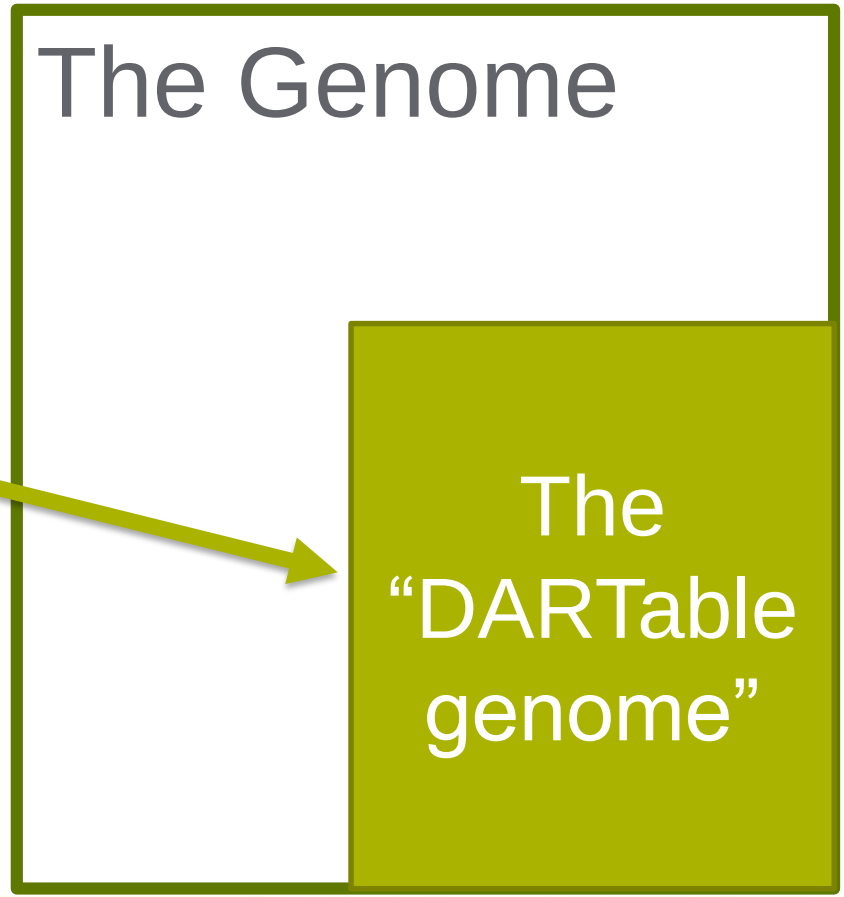
Time to find, collate and interpret data

Continuing to build our own confidence in these approaches

Acceptance of approaches (situational)

What is the DARTable genome?

The **subset** of the genome that, **when perturbed, results** in a **developmental** or **reproductive** effect.



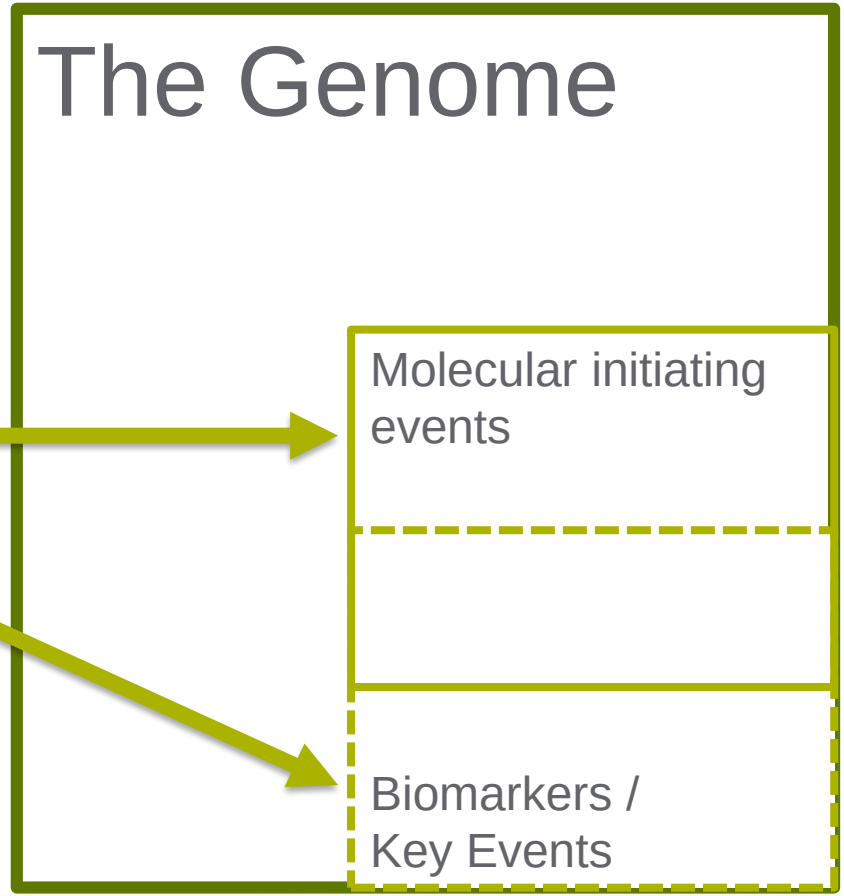
What is the DARTable genome?

DARTable gene products might be **direct targets of exogenous toxicants**

DARTable gene products might **have their abundance changed or activity altered**

- **in response** to a chemical treatment or genetic alteration that results in a DART effect

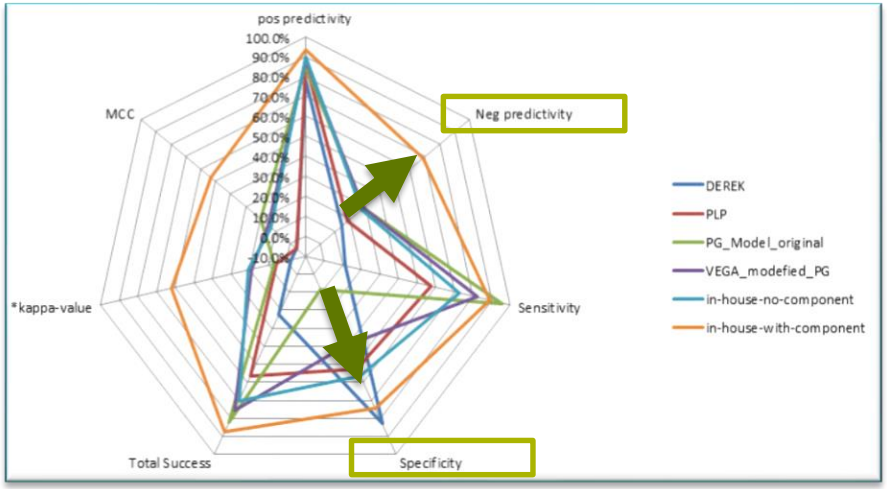
So the “DARTable genome hypothesis” represents **the comprehensive set of MIE and transcript/proteomic biomarkers for DART effects**



How to identify the DARTable genome?

	Known chemically-induced toxicity	Unknown chemically-induced toxicity
Known MIE	*The proteins that are **known** to mediate the adverse effects of known chemicals…* We **curated** a list of 58 MIEs (expands to 122 genes) based on our historic experiences	*The proteins we that **might** mediate the DART effects of chemicals **if** they interact with them…* Identify suitable proteins from **public databases using bio-informatics** approaches
Unknown MIE	*The proteins we are unaware of that mediate the adverse effects of known chemicals…* **Mine the historic data properly**! Perform investigative experiments.	

The “known knowns” DARTable genome-based models have good predictive value



- We used **published models** (Mervin *et al,* 2015) developed ChEMBL data and an FDA test set of dev tox results to compared with other modelling approaches (e.g. DEREK Nexus, Wu et al, 2013 decision tree),
- we found
  - **similar sensitivity** and
  - **improved negative predictivity** and **sensitivity**

Wu S, Fisher J, Naciff J, Laufersweiler M, Lester C, Daston G, Blackburn K. Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. Chem Res Toxicol. 2013 Dec 16;26(12):1840-61.
Mervin LH, Afzal AM, Drakakis G, Lewis R, Engkvist O, Bender A. Target prediction utilising negative bioactivity data covering large chemical space. J Cheminform. 2015 Oct 24;7:51

How to identify the DARTable genome? #2

Models built on the “known knowns” have good predictive value: **provides confidence** that the **properties of these genes** may be predictive of other “DARTable genes”

- protein is **associated with pathways** known to be **important in development**
- **expressed during development**
- **Knock out** (in rats, Zebrafish, C. elegans, Drosophila) or **mutations** (in humans) can **cause developmental defects**
- Chemicals that bind with high affinity to a target also **cause consistent developmental defects within a species** (in humans, rodents, rabbits, Zebrafish, C. elegans, Drosophila)

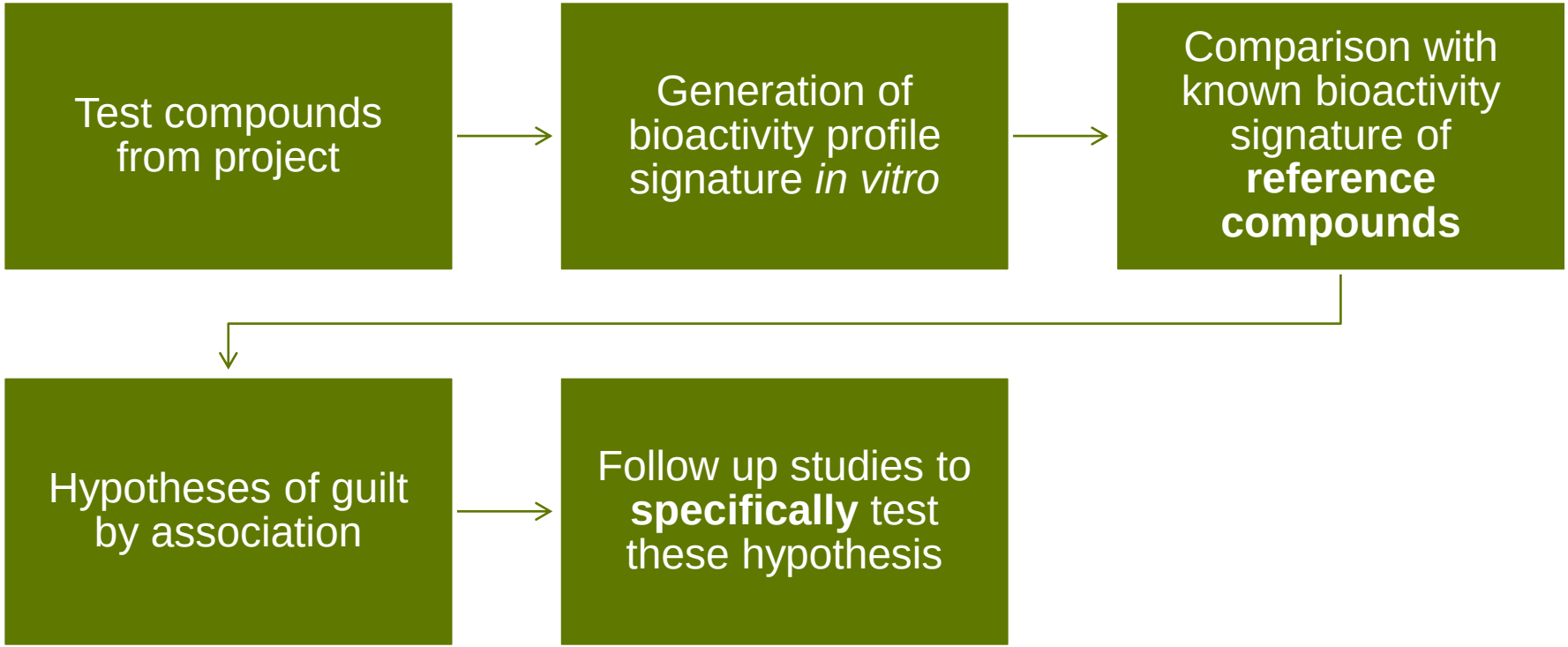
Initial exploration of public datasets: How big is the DARTable genome?

Analysis of connections between genes and developmental phenotypes in knock-out mice

- Can network analysis tools help us prioritise the key genes to include?

5402 genes

Can we identify DARTable gene responses from omics signatures?



Evaluation of in vitro high-throughput transcriptomics (L1000) data as an early screen for potential to perturb the DARTable genome

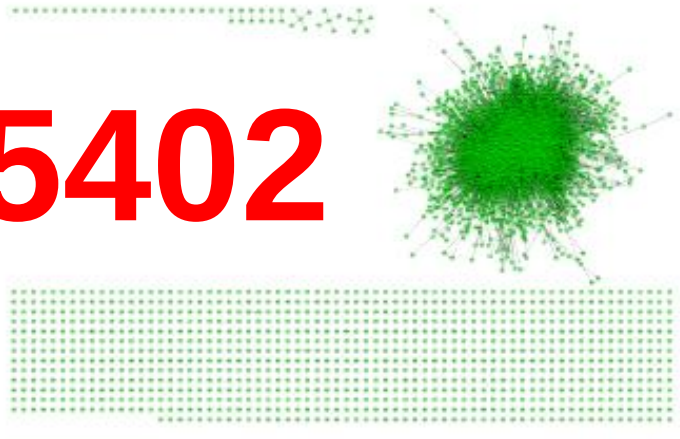
signature of CSAA535232

- Mycophenolic acid inhibits de novo purine biosynthesis via Inosine-5′-monophosphate dehydrogenase (IMPDH) inhibition
  - ✓ IMPDH2 KO mice are embryonic lethal
- CSAA535232 is leflunomide a Dihydroorotate dehydrogenase (DHOD) inhibitor, and rodent teratogen
  - ❑ DHOD KO mice are embryonic lethal
- Both **purine/pyrimidine synthesis gives similar response**, but anti-correlated with inhibition of proteostasis

Initial exploration of public datasets

How big is the DARTable genome?

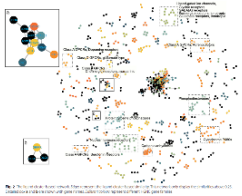
Analysis of connections between genes and developmental phenotypes in knock-out mice



5402

What fraction of the DARTable genome is known to be influenced by chemicals?

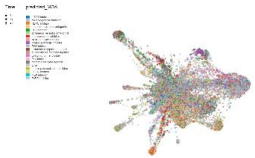
Polypharmacology in ESCAPE DB



<40%

What fraction of the DARTable genome can we measure *in vitro* responses for?

L1000 compound profile similarity by MOA



<15%

Learnings

High content *in vitro* data **can** provide information about mechanistic hypotheses based on similarity to other perturbations

- Provided there are **suitable example chemical or gene knockdown perturbations** in the reference database
- But some (most) mechanisms are **not available** at the moment!

Additional experimentation to identify MIEs is essential

- Bioactivity signatures are **not specific** to a particular MIE but represent the cellular response that may be **common to groups of MIEs**
- Bioactivity measurements are likely to generate **several MIE hypotheses**

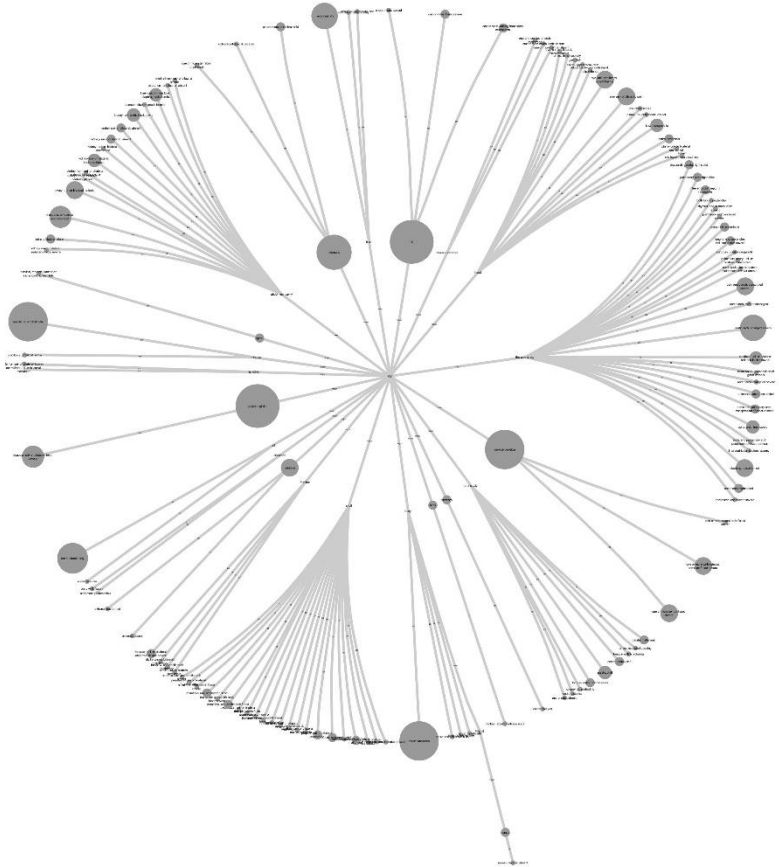
Questions

- What are the best technologies to use to generate bioactivity signatures?
- What are the best methods for generating bioactivity signatures?
- Can we infer mechanisms from a combination of *in silico* fingerprint and *in vitro* bioactivity data?
- Can we collectively enhance the collection of relevant signatures?

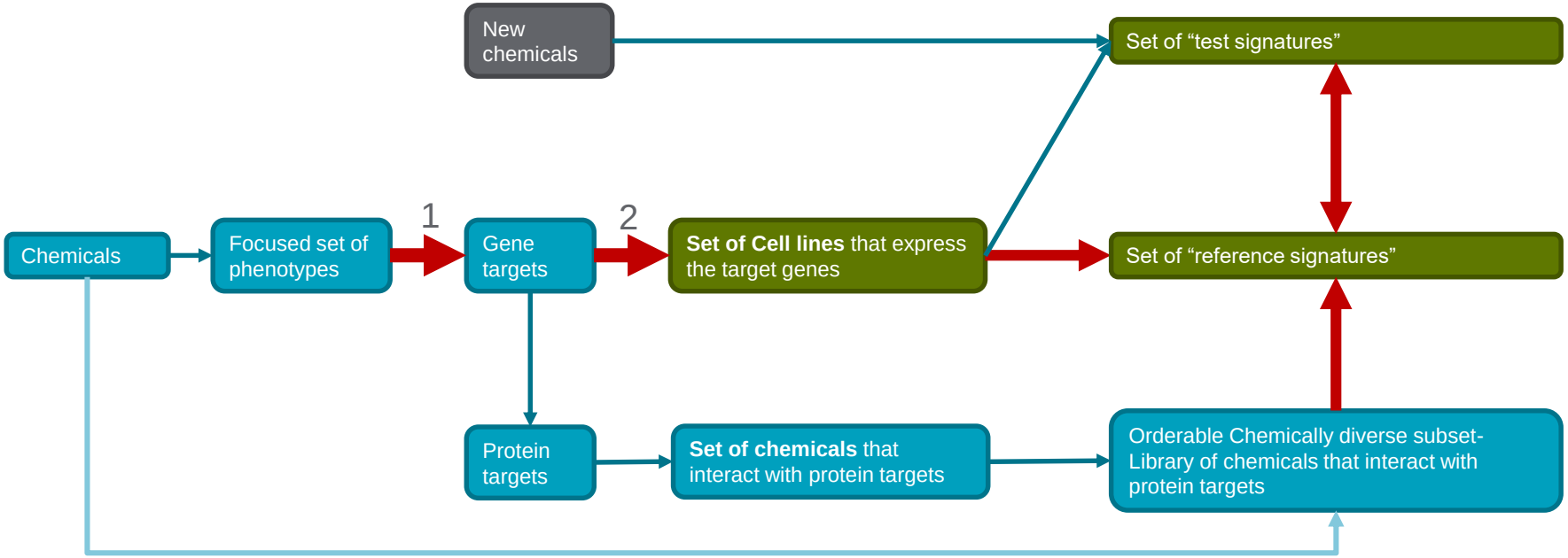
Do we need to predict everything?

A question of prioritisation:

- Analysis of our DART studies since 2007 demonstrates **typically 3-5 reproducible malformations** per MOA/chemical class causing defects
- Can we **share data** to compile a definitive list of defects caused by chemicals in rodent preclinical studies?
- Can we identify potential MIEs that might cause these effects?



Build the DARTable genome and tools to explore it from observed phenotypes

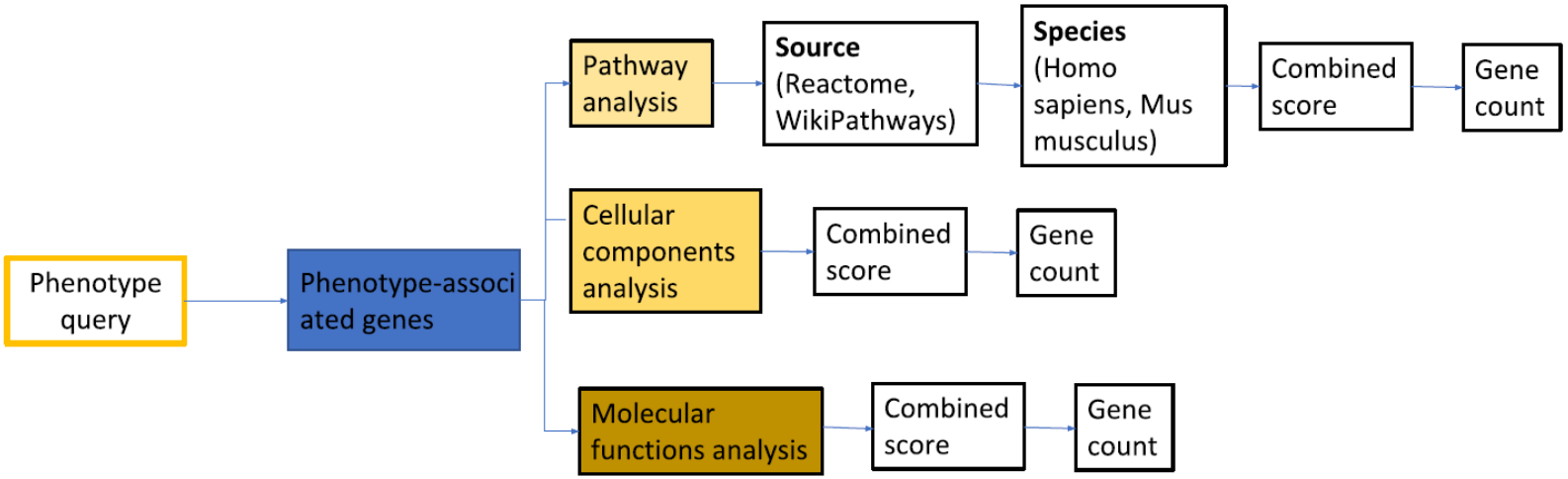


Collaboration with Edelweiss Connect to create the Syntox tool – hypothetical MIE identification from phenotypes

Syntox: the use case of omphalocele and gastroschisis

Tatyana Doktorova, PhD

Syntox Methodology – web-based tool leveraging public databases to explore potential MIEs for a user-defined specific (set of) phenotype(s)

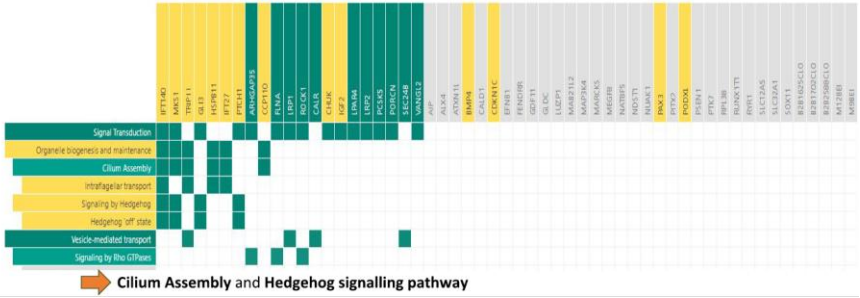


Example use case exploring drivers for the phenotype “omphalocele”: can explore the known biology to focus on specific targets of interest

Syntox: the use case of omphalocele

Pathway analysis:
- Source: Wiki pathways, Reactome
- Species: Homo sapiens
- Combines score: above 5
- Gene count: above 2
- Adjusted p-value: below 0.05

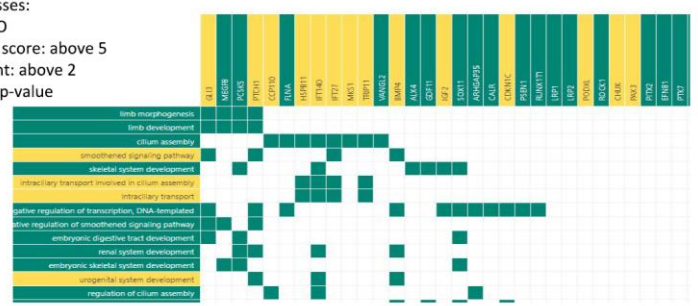
Selected genes (below in yellow):
BMP4, CCP110, CDKN1C, CHUK, GLI3, HSPB11, IFT140, IFT27, IGF2, MKS1, PAX3, PODXL, PTCH1, TRIP11



Cilium Assembly and **Hedgehog signalling pathway**

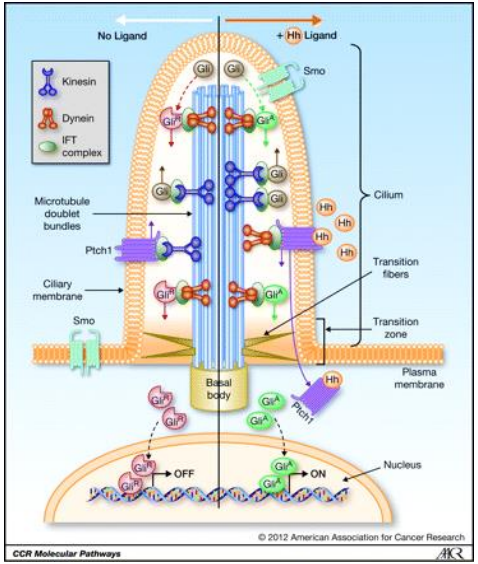
Syntox: the use case of omphalocele

Biological processes:
- Source: GO
- Combines score: above 5
- Gene count: above 2
- Sorted by p-value



Cilium assembly and intraciliary transport

In addition to data-driven Targets bioinformatic analysis can identify mechanistic hypothesis to enrich the Targets of interest lists.



Syntox: the use case of omphalocele

Hypothesis: Omphalocele is a disorder caused by dysfunction of primary **cilia** during early embryogenesis.

Evidences:

Evidences at a **PATHWAY** level:

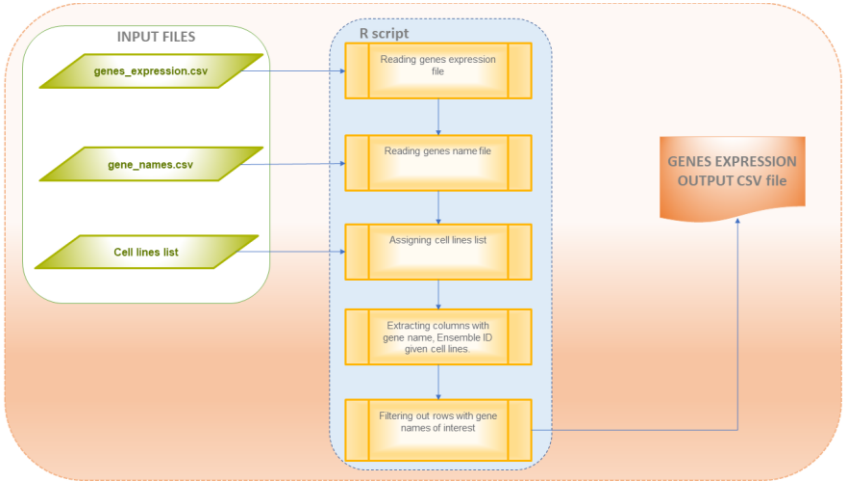
- **Hedgehog signaling** is coordinated by the differential localization of the receptors patched-1 **(PTCH1)** and Smoothened (**SMO**) in the primary cilium.
- **Cilia assembly** is mediated by **intraflagellar transport (IFT)**

Evidences at a **BIOLOGICAL FUNCTION** level:

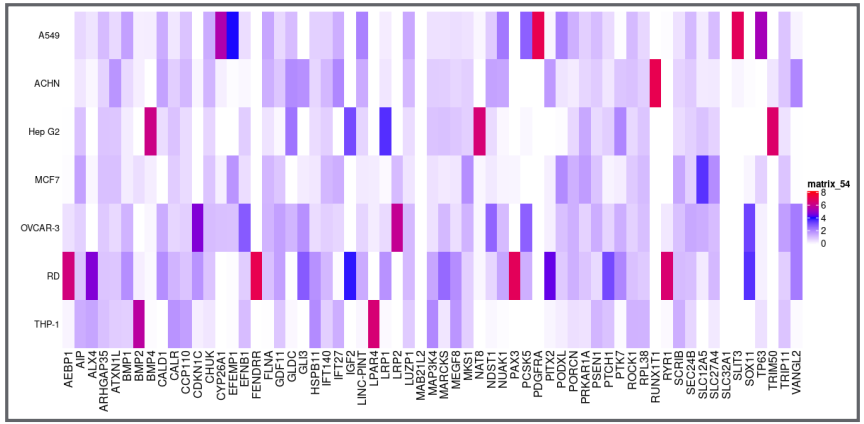
- Limb development
- **Smoothened** signalling pathway
- (Regulation of) **Cilium assembly**
- Embryonic **digestive tract development**
- **Intracellular** transport

Selection of suitable cell lines to test DARTable genome Targets of Interest identified via Syntox

Extracting genes and cell lines of interest from Human Expression Atlas Dataset

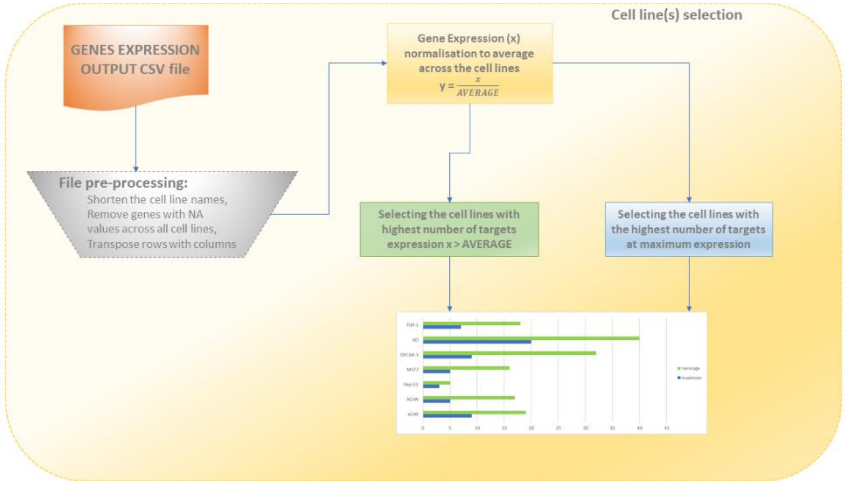


Targets of interest are in Expression Atlas

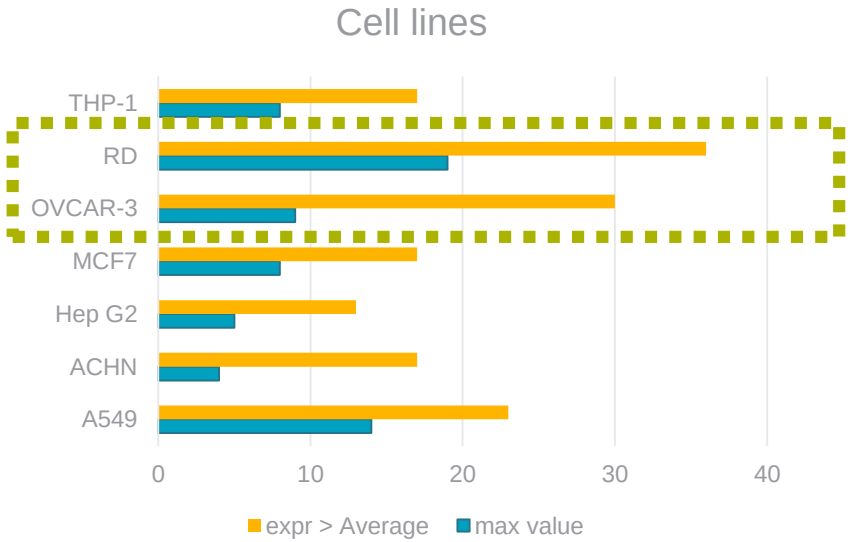


The workflow identifies RD and OVCAR-3 cells as the core cell lines that express targets of interest for the DARTable genome.

Pipeline to select the best set of cell lines for the target of interest list

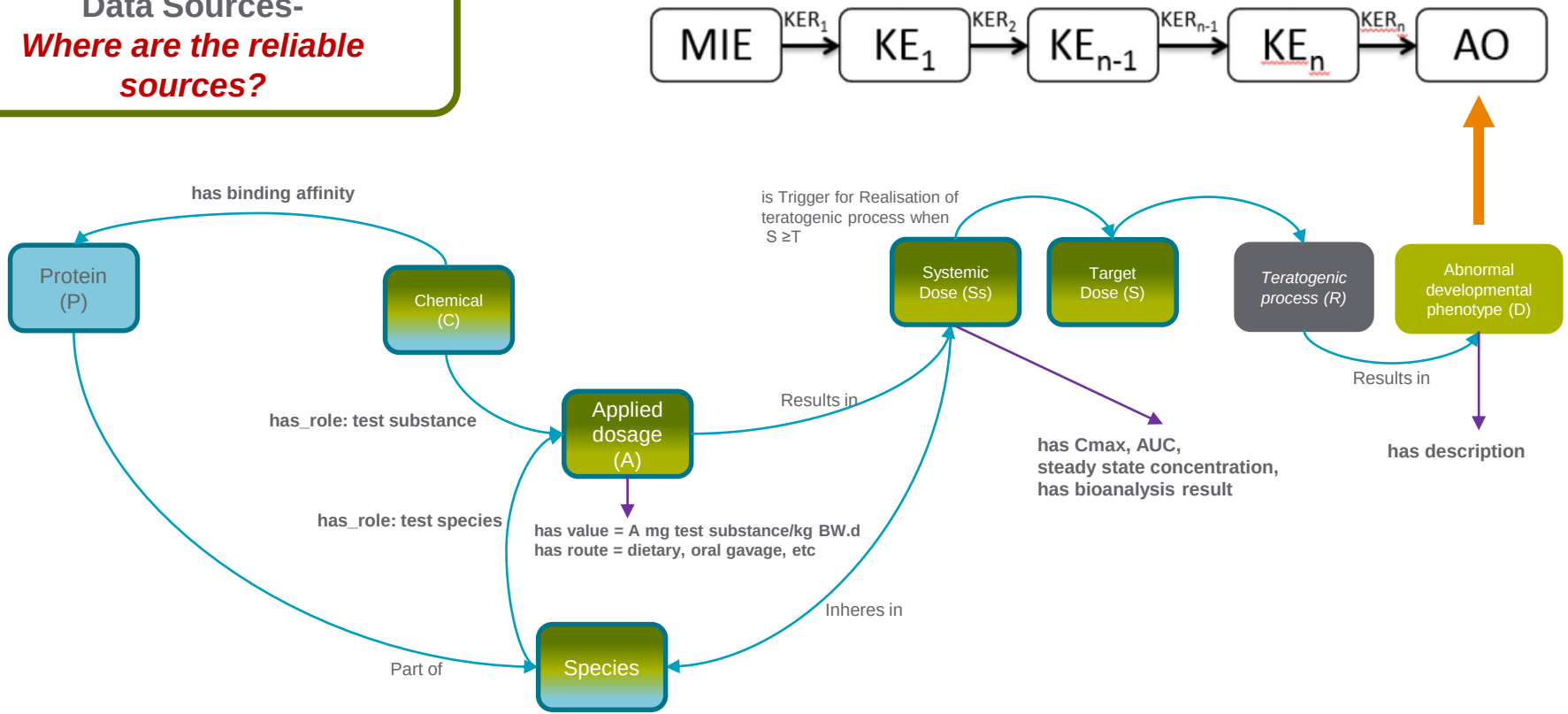


Cell line selection

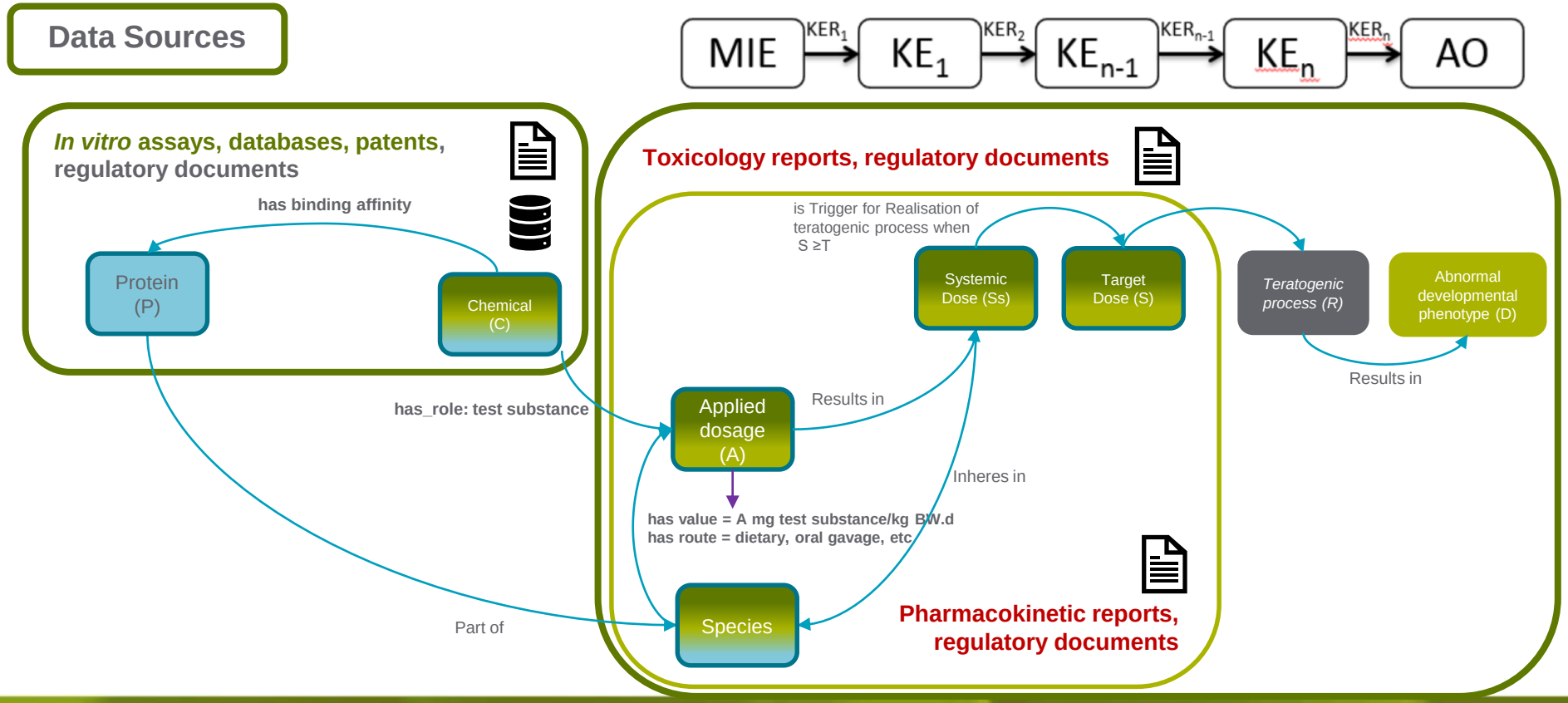


qAOPs as a data integration challenge: finding the MIE is only the start

Data Sources-
Where are the reliable sources?



Data extraction and integration: missing data from bioinformatics databases?



Bioinformatics tools can help identify the DARTable genome

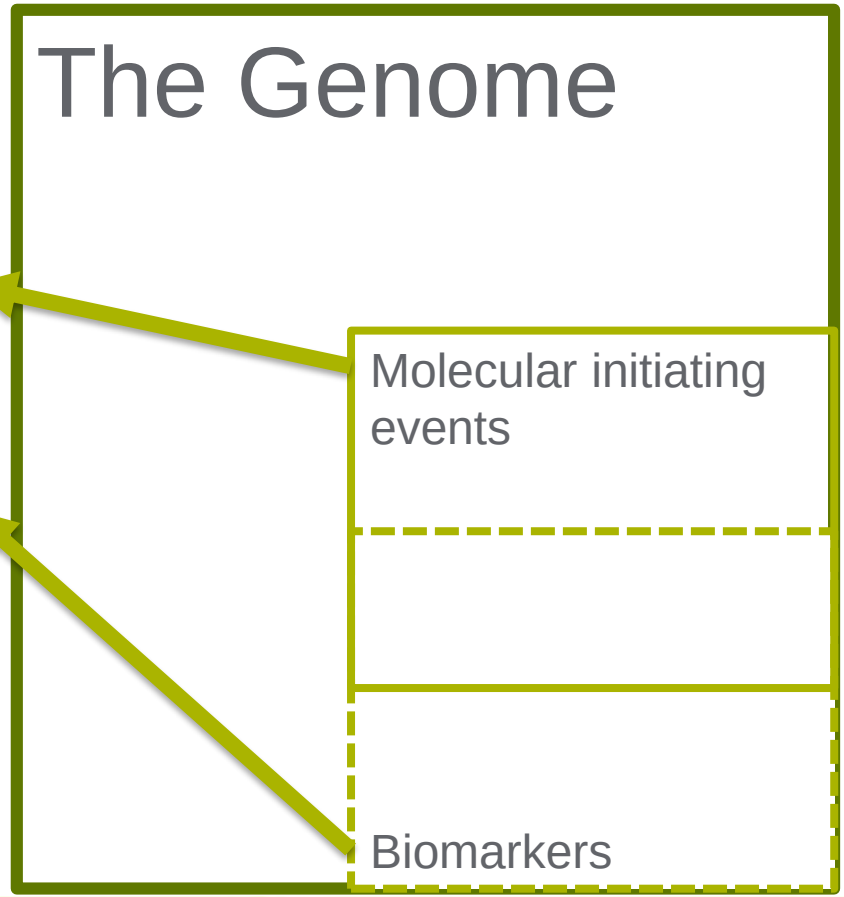
DARTable gene products might be **direct targets of exogenous toxicants**

- Bioinformatics tools help with **MIE hypothesis generation**

DARTable gene products might **have their abundance change in response** to a chemical or genetic treatment

- Bioinformatics tools hold **expression signatures of direct targets**
- Bioinformatics tools can aid **optimal cell line selection** to measure signatures

It is possible to identify all of these features **if the relevant data is available in the public domain**



Challenges

Availability of relevant data sources:

- **mostly unstructured**
- **hard to find**
- **not public,** or
- **not yet generated**

Survivor Bias

Prediction of the quantitative MIE thresholds in each species is **not yet possible**

- at the moment they are **empirically** derived

What are the **opportunities to** share more data to enable **collective** problem solving?

Acknowledgements

Syngenta

- Product Safety
- Computational Chemistry
- Physical Chemistry

Collaborators

- HESI DART technical committee
- Eidelweiss Connect
- Genometry
- NC3Rs DARTpaths
- IBC/STFC