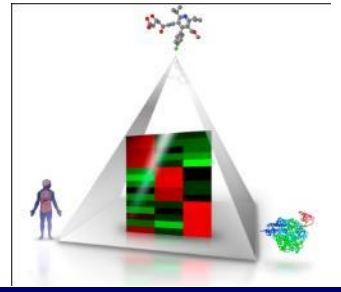
Towards Integrated Compound Safety Assessment, In Particular the Use of ’Omics Data and Pharmacokinetics Information, In Toxicity and Safety Prediction

Andreas Bender, PhD
Natural Philosopher for Molecular Informatics
Department of Chemistry, University of Cambridge
Fellow of King’s College, Cambridge
Director of Digital Life Sciences
Nuvisan, Berlin

UNIVERSITY OF CAMBRIDGE

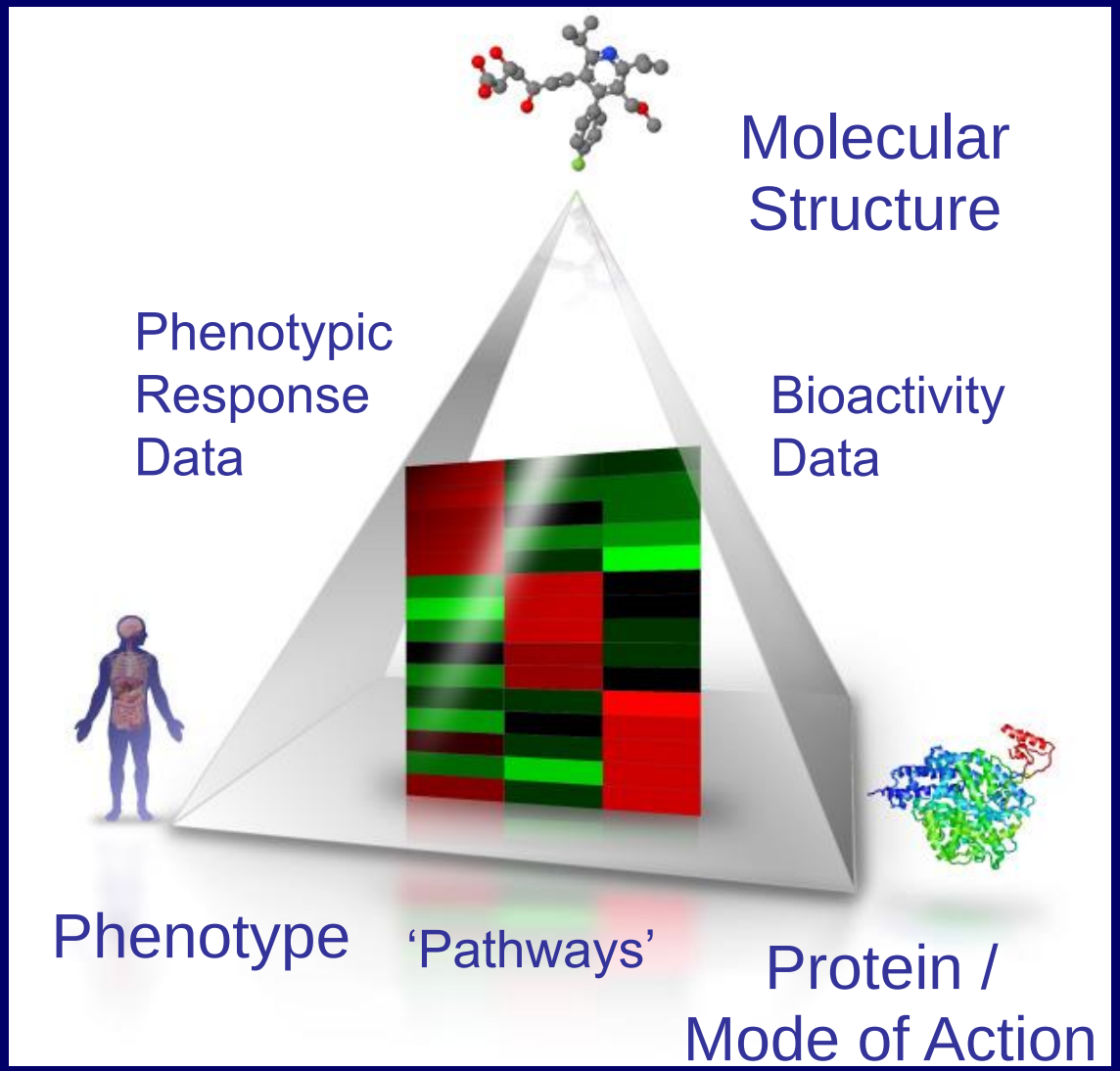


Any statements made during this talk are in my capacity as an academic

Outline: Chemical and biological data, its complexities, and applications to safety

- Chemical and biological data: The flat-earth view
  - And where a flat earth is great!
- Chemical and biological data: The round-earth view
  - Drug discovery data and its complexity
- Using ‘omics data and analytical methods, vs single-endpoint data and synthetic methods, for predictive safety
  - Using ‘omics data in DIVI, time-resolved gene expression data for AOP derivation in DILI
  - Anticipating DILI using assay-based information plus PK approximations
  - Machine learning for PK

A simple view on the world: Linking Chemistry, Phenotype, Targets / Mode of Action (myself, until *ca.* 2010)

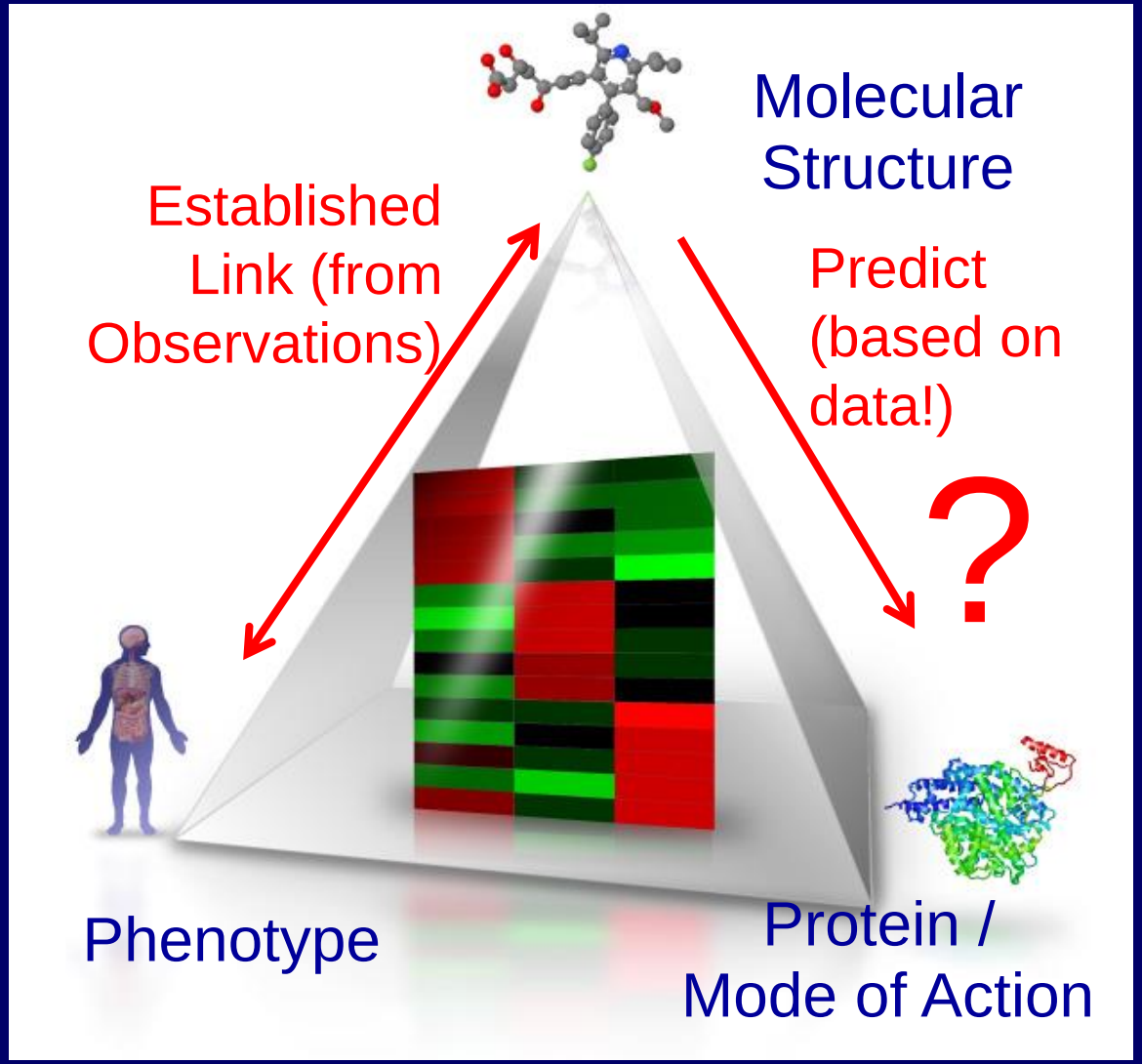


a.k.a.
"The world is flat"

= "We believe our labels"

(which are often insufficiently quantified, not directed, unconditional, don't have time/ concentration/biological setup dependence, *etc.*)

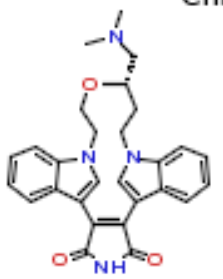
Starting from *in vivo* efficacy we can hypothesize the MoA, based on ligand chemistry

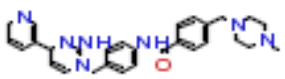


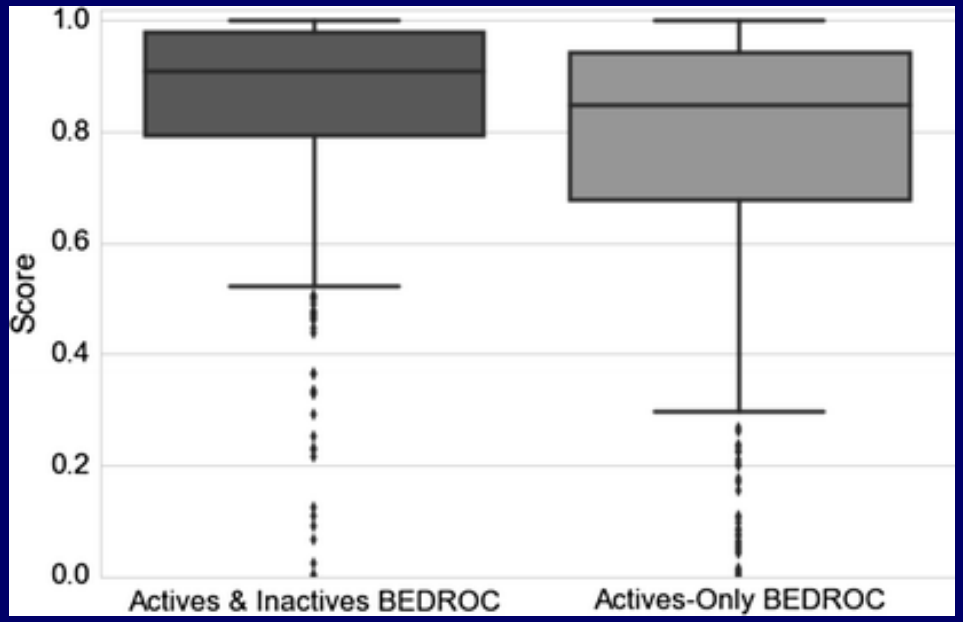
A. Koutsoukas *et* al., J Proteomics 2011 (74) 2554 – 2574.

The ‘flat earth’ view can *still* help! Eg Public target prediction model, based on ~200 mio data points

- E.g. work of Lewis Mervin, with AstraZeneca
- 2015, *J. Cheminformatics* (7) 51
- ChEMBL actives (~300k), PubChem inactives (~200m); 1,080 targets
- Can be retrained on in-house data
- https://github.com/lhm30/PIDGIN

Molecule	Targets	Scores
Chiral 	PRKCB1	95.81
	CAMK2G	87.48
	PRKCG	66.35
	PRKCA	56.99
	PRKCD	52.44
	PRKCH	51.41
	PRKCE	50.42
	PRKCZ	42.48

Molecule	Targets	Scores
	ABL1	46.50
	PDGFRB	28.99
	KIT	22.02
	CDK9	21.30
	BRAF	16.13
	FLT1	13.09
	PLK1	8.05
	BTK	5.44



Also data publicly available

So: Using bioactivity data for ligand-protein activity modelling '*is relatively possible*'

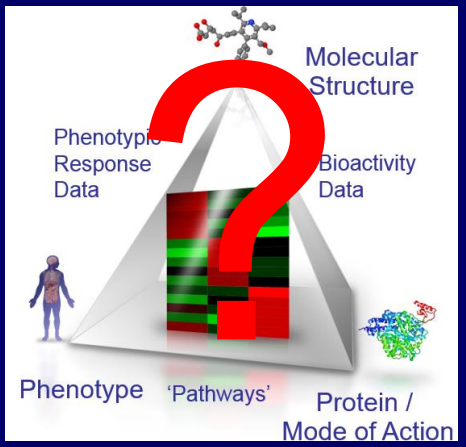
- We make use of existing data (millions of data points!)
- On-target bioactivities (links between chemical structure and protein targets) are *relatively large-scale*, and *relatively homogenous*
- Hence, generating models for bioactivities is 'possible'
- Can also be used for design (eg multi-target ligands)

BUT:

- Only covers known chemical space
- Labels are still heterogenous
- *In vivo* relevance of predictions needs to be established

(PK, target engagement *in vivo*, etc.)

BUT…The world is not flat. What now?



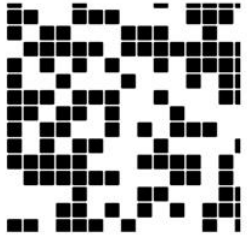
- Links between drugs/targets/diseases are quantitative, incompletely characterized
- Subtle differences in eg compound effects (partial vs full agonists, off-targets, residence times, biased signalling, etc.)
- 'Pathways' from very heterogenous underlying information; dynamic elements not captured etc.
- Effects are state-dependent (variation between individuals, age, sex, co-medication…) – PK is often rather neglected in AI approaches
- Phenotyping is sparse, subjective (deep phenotyping?)
- We don't understand biology ('the system'), we don't know what we *should* label, and measure, hence …
- We label what we *can* measure: 'T*echnology push*' vs '*science pull*' (!)
- **Are our labels – 'drug treats disease X', 'ligand is active against target Y', … - meaningful?**
- **Conditionality: Causality, confidence, quantification, ….?**
- **Computer science is tremendously powerful… but is our data?**

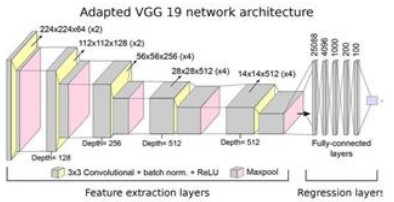
Example of difficulties with ‘labels’: adverse reactions

- **“*Does* drug Y cause adverse reaction Z? Yes, or no?”**
- Pharmacovigilance Department: Yes, *if* we have…
  - A patient with this *genotype* (which is generally unknown)
  - Who has this *disease endotype* (which is often insufficiently defined)
  - Who takes *dose X* of *drug Y* (but sometimes also forgets to take it)
  - With known targets 1...n, but also unknown targets (n+1…z)
  - Then we see *adverse reaction (effect) Z* …
  - But only in x*% of all cases* and
  - With *different severity* and
  - *Mostly if co-administered with a drug from class C*, and then
  - More frequently in *males* and
  - Only *long-term*
  - (Etc.)
- **So – does drug Y cause adverse event Z?**

Object | Representation | Model | Object Label

Image Domain





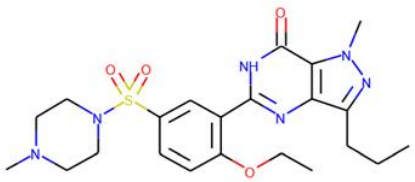
ResNet?
AlexNet?
CapsuleNet?

Cat

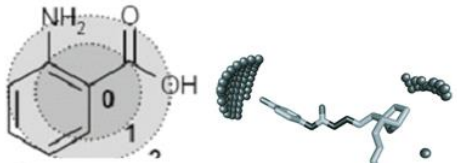
Largely *Unconditional* labels

Representation and model are *intrinsically linked* (ie, model uses native object representation by pixels)

Drug Discovery: Chemical Domain



?



logP = ...
Molecular Weight = ...
Molar Refractivity = ...

?

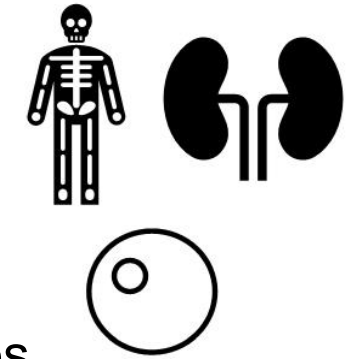
Artificial Neural Network/DNN? Support Vector Machine? Random Forest? Bayesian Classifier?...

Property A

Conditional labels (eg dependent on assay system, genotype, dose, endotype, sex, age, comedications, lifestyle, ...) ?

Both representation and modelling approach are *largely trial and error* (and *not* intrinsic to the chemical domain)

Drug Discovery: Biological Domain



?

Transcript-/proteomics? High-content imaging? Epigenetics? Histopathology? ....

?

Artificial Neural Network? Support Vector Machine? Random Forest?

State/Effect B

Heavily conditional labels (eg dependent on genotype, dose, endotype, sex, age, comedications, lifestyle, ...) ?

Both representation and modelling approach are *largely trial and error* (in particular the information content of biological readouts has only been established for particular cases)

Bender & Cortes, Drug Discovery Today, 2021

Data/'AI' in early discovery vs efficacy/safety

Early discovery/proxy space (usually *in vitro*)

- Often 'simple' readouts (eg protein activity), hence…
- Large number of data points for training models

- *Models have clear labels* (within limits of model system, eg 'ligand is active against protein at IC50<10uM', or solubilities, logP, or the like)
- Good for model generation: *Many*, *clearly categorized* data points

Efficacy/safety (usually *in vivo*)

- Quantitative data (dose, exposure, …)
- More complex models (to generate data), *fuzzy labels* (classes 'depend', on exposure, multiple eg histopathological endpoints) – hence…

- *Less, and less clearly labelled data*: Difficult from machine learning angle
- Data: *Recording* vs data *suitable for mining* – eg animal data tricky, even within single company

Problem setting in early discovery vs safety

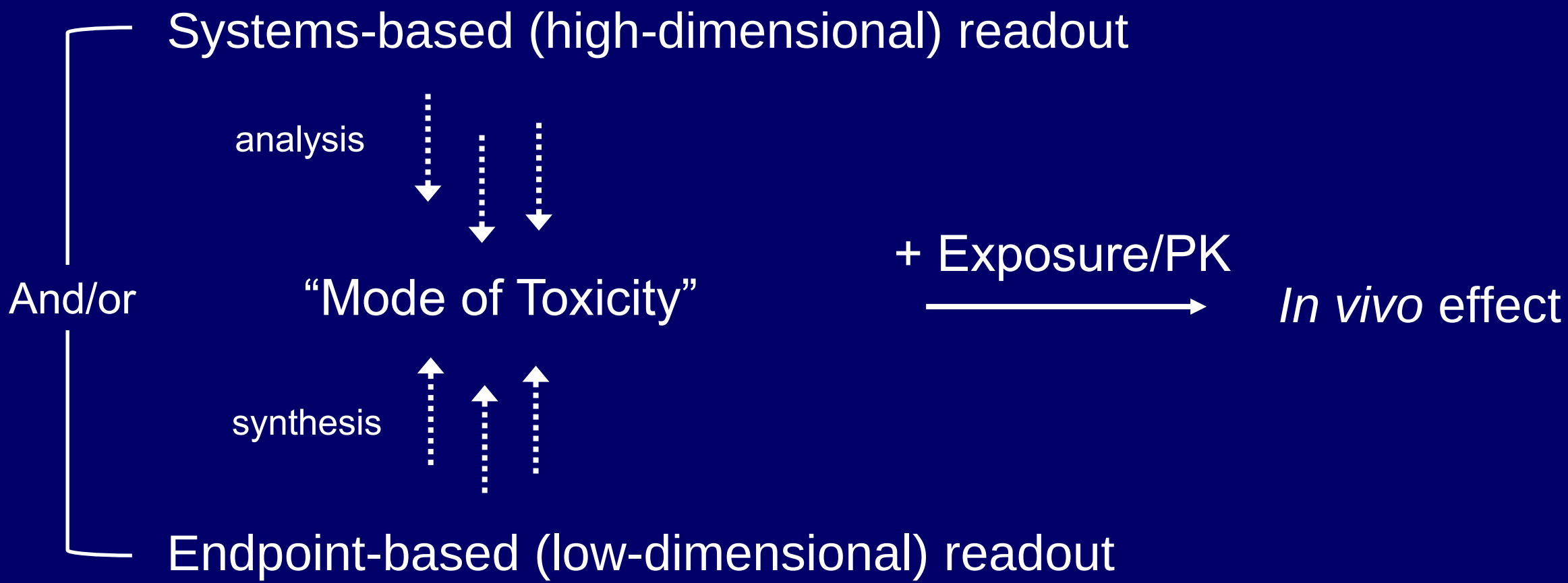
Early discovery/proxy space

- Discovery setting – ‘find me suitable 100s or 1000s out of a million’ (eg screening)
- Anything fulfilling (limited) set of criteria will do ‘for now’, predicting *presence of something*
- Computationally *generative* models often fine

Efficacy/safety

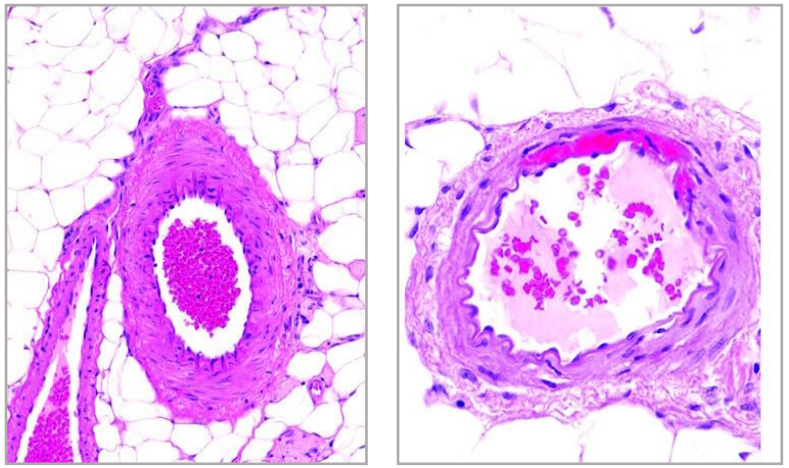
- Need to predict for *this particular data point, quantitatively!*
- *Long list of criteria to rule out*, *based on limited data*… predicting *absence of ‘everything’* (eg different modes of toxicity)
- *Predictive* models (more tricky than generative!)

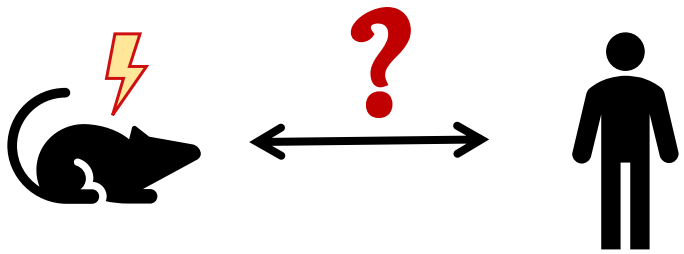
‘Omics vs endpoint-based safety models: Conceptual differences and DIVI, DILI as case studies



Drug-Induced Vascular Injury (DIVI): Work of Anika Liu, with GSK

- Biomechanical stress and/or direct action on the vascular cells can initiate DIVI which is characterized by morphological vascular changes, in particular medial arterial necrosis (MAN)
- Pathogenesis and translation to humans remain largely unclear, also because DIVI can often not be monitored clinically and only detected by histopathology.
- Despite small evidence for translation to humans, pre-clinical DIVI leads to delays in compound development and/or and termination





Goal: Identify transcriptomic biomarkers for MAN in rats which can help to understand and monitor pre-clinical DIVI.

UNIVERSITY OF CAMBRIDGE

gsk

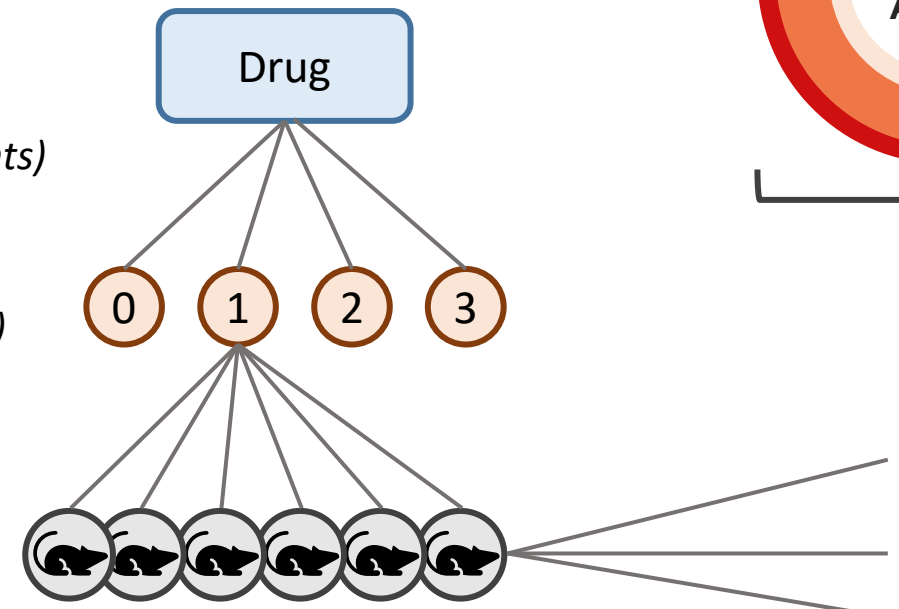
Images from Dalmas, D. A., et al. Transcriptional Profiling of Laser Capture Microdissected Rat Arterial Elements: Fenoldopam-Induced Vascular Toxicity as a Model System. *Toxicol. Pathol.* **2008**, *36* (3), 496–519.

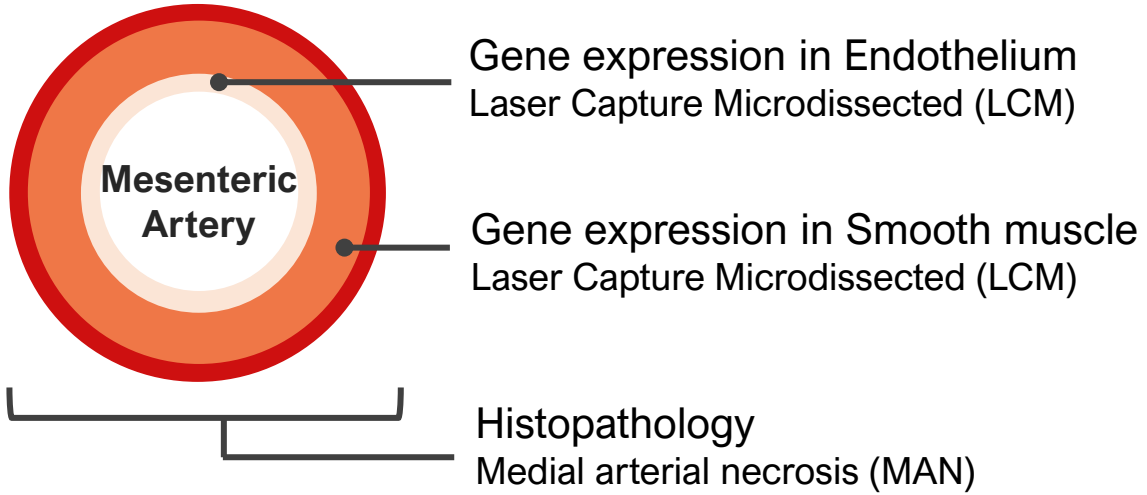
Data generation by Dalmas et al. [1]

14 Experiments
(12 compounds total,
2/12 compounds at 2 timepoints)

4 Compound doses
(e.g. 0/1/30/300 mg/kg/day)

~5 Biological replicates





Data	n(features)	n(animals)
Gene expression (*Tunica media*)	15240	304
Gene expression (*Tunica intima*)	15240	300
Histopathology	34	328

[1] Dalmas et al. (2011), Toxicology and Applied Pharmacology, 257(2), 284–300.

All studies were conducted in accordance with the GSK Policy on the Care, Welfare and Treatment of Laboratory Animals and were reviewed the Institutional Animal Care and Use Committee either at GSK or by the ethical review process at the institution where the work was performed.

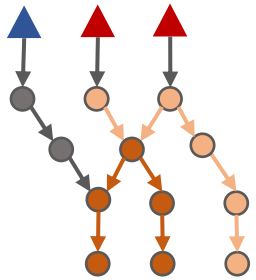
Criteria to identify potential transcriptomic biomarkers in DIVI

1) **Consistency** across conditions showing DIVI
2) **Specificity** for conditions showing DIVI
3) **Dose-dependency** of expression change for compounds showing DIVI
4) **Large (measurable) effect** across conditions showing DIVI

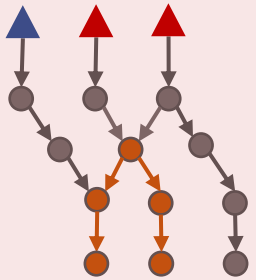
→ Identify few most promising genes as potential biomarkers
(at the risk of losing many other relevant ones)

What is a “mechanism of toxicity”?

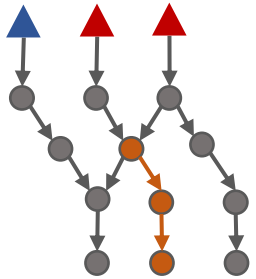
Associated



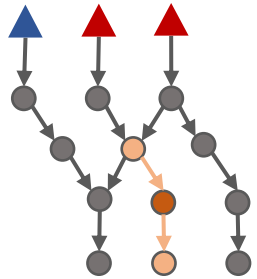
Conserved



Conserved and specific



(Mechanistic) biomarker



• Everything associated might be relevant	• Most likely downstream processes where AOP converges	• Distinguishes phenotype	• Distinguishes phenotype + large effect size
• Many potential covariates	• Misses upstream regulation which is likely compound-specific	• Might miss key parts of AOP • Importance != Specificity	• Little insight into mechanism • Importance != Effect size

Mechanism ← **Question** → **Biomarker**

High coverage ← **Detected biological entity** → **Strong evidence**

Compound

- No adverse event
- Adverse event

Biological entity (Protein/pathway/..)

- “Mechanism”
- Maybe “mechanism” (depending on evidence)
- Not “mechanism”

UNIVERSITY OF CAMBRIDGE

gsk

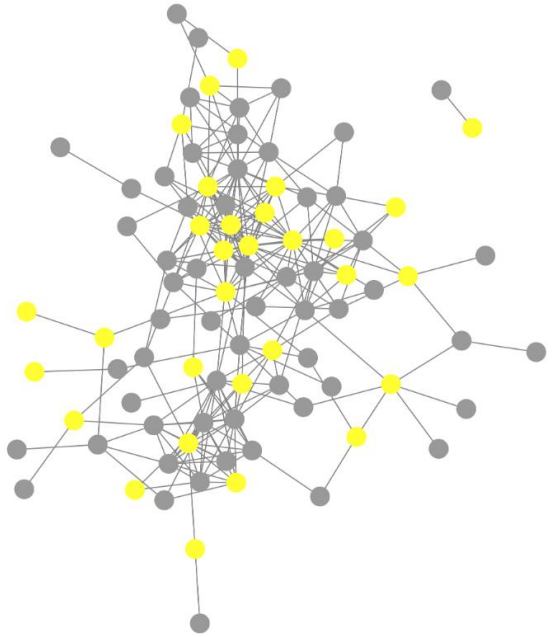
Conserved genes across DIVI conditions are highly interlinked on protein level

1) Identify conserved genes

2) Identify protein-protein interactions (PPI) between proteins encoded by conserved genes (STRING[1])

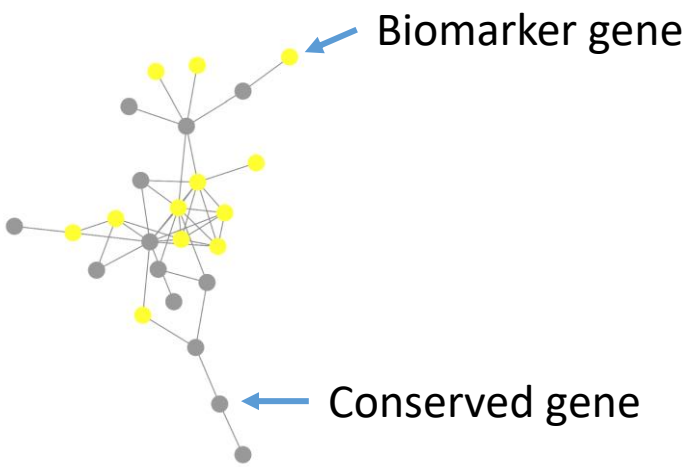
3) Is the number of protein-protein associations higher than expected at random? (PPI enrichment)

Smooth muscle
(135 conserved genes)



p-value: < 1.0e-16

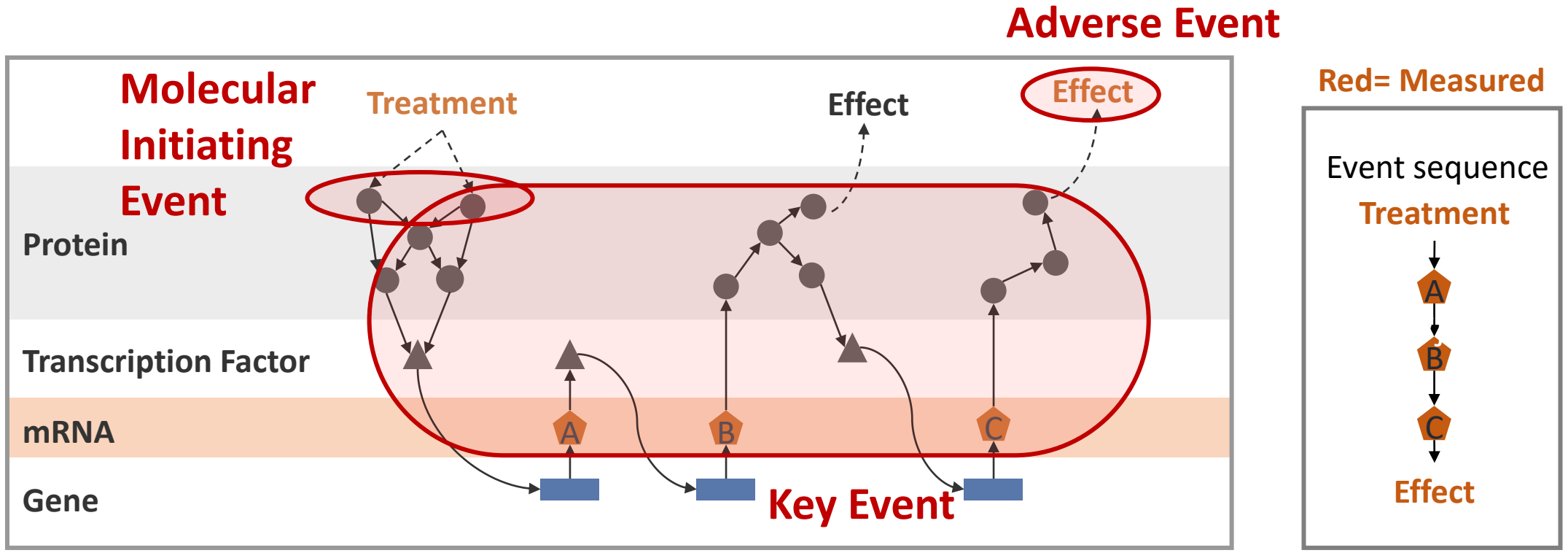
Endothelium
(50 conserved genes)



p-value: 5.6e-12

[1] Szklarczyk, et al. *Nucleic Acids Res.* **2019**, *47* (D1), D607–D613.

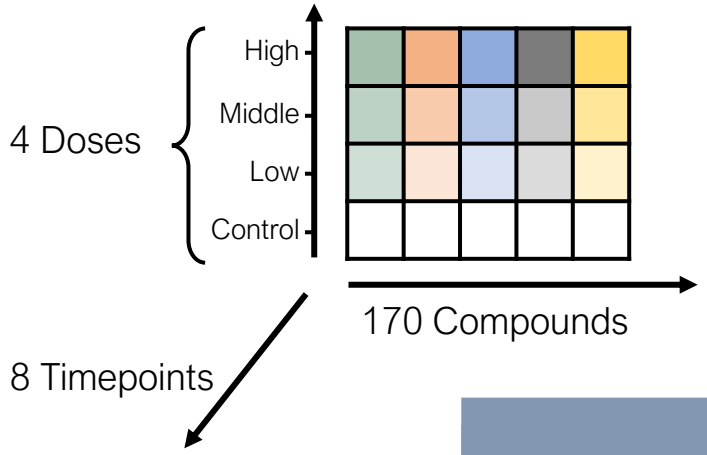
Ordered responses in DILI pathogenesis (work of Anika Liu)



- Extend AOP: If event A is observed, **how likely** and **when** will event B observed?
- Mechanistically: Link **proteins** expressed **early** with **genes** expressed **late**

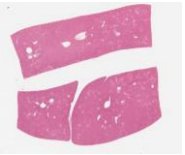
Open TG-GATEs

Toxicogenomics Project-Genomics Assisted Toxicity Evaluation System



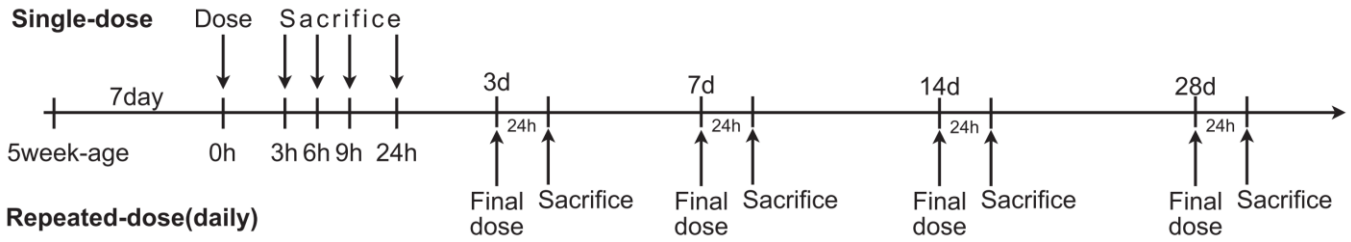
Data types

Transcriptomics
(**Liver**/Kidney)



Histopathology
(**Liver**/Kidney)

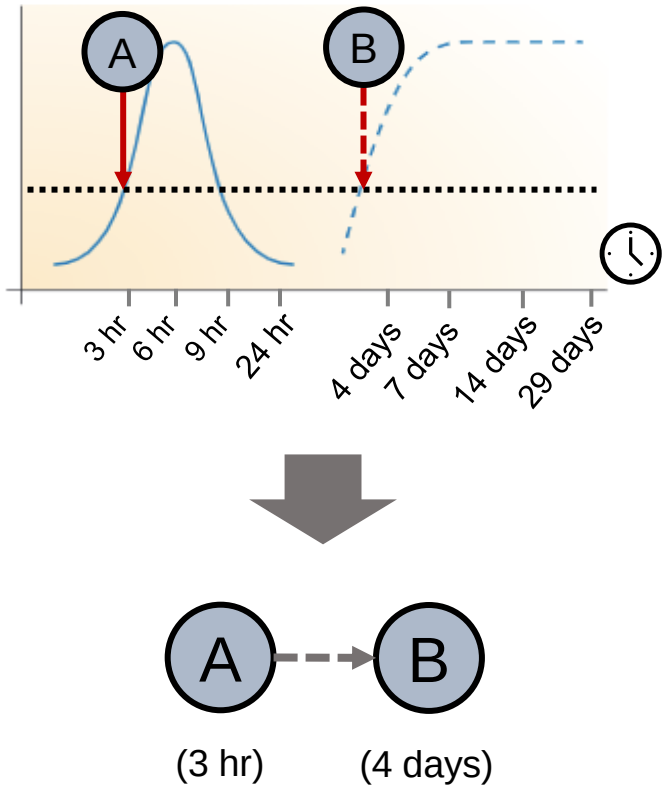
Experimental design



UNIVERSITY OF CAMBRIDGE

[1] Y. Igarashi et al., Nucleic Acids Res., 2015.

Deriving 1st activation per timeseries



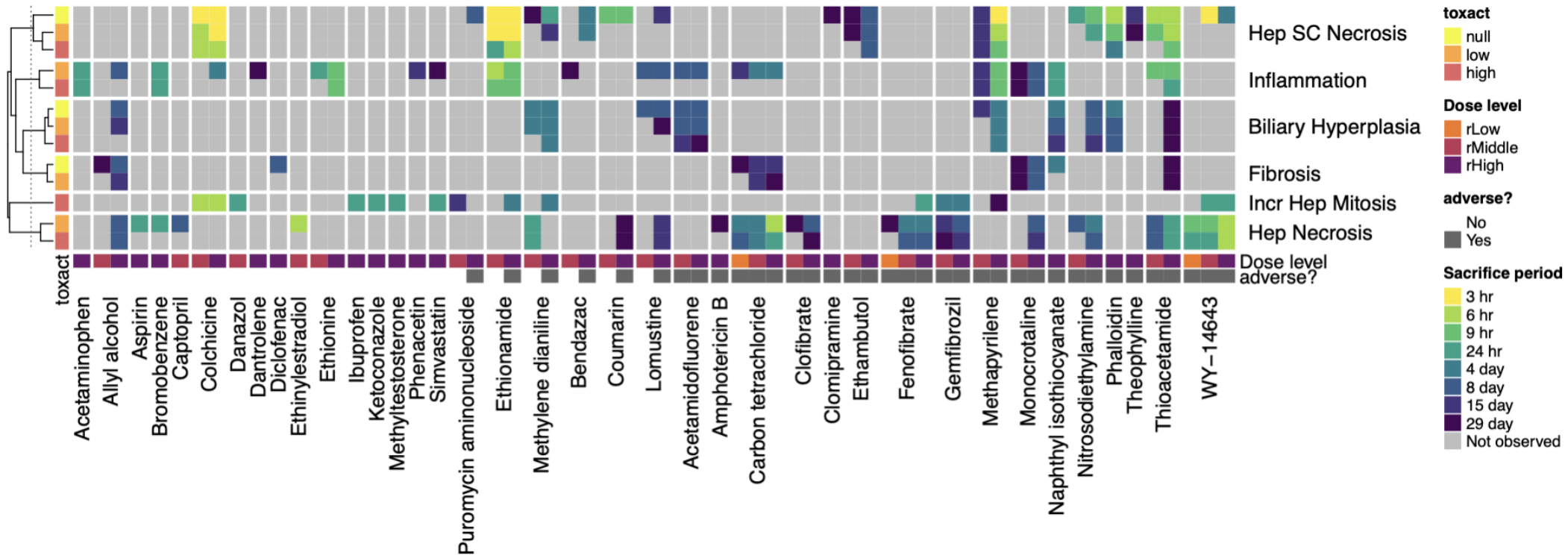
1. Histopathology
   - Toxscore > 0.1 → *Null*
   - Toxscore > 0.67 → *Low*
   - Toxscore > 1.34 → *High*

2. Pathway activation

   Significant difference to time and experiment-matched control group (pval < 0.05)

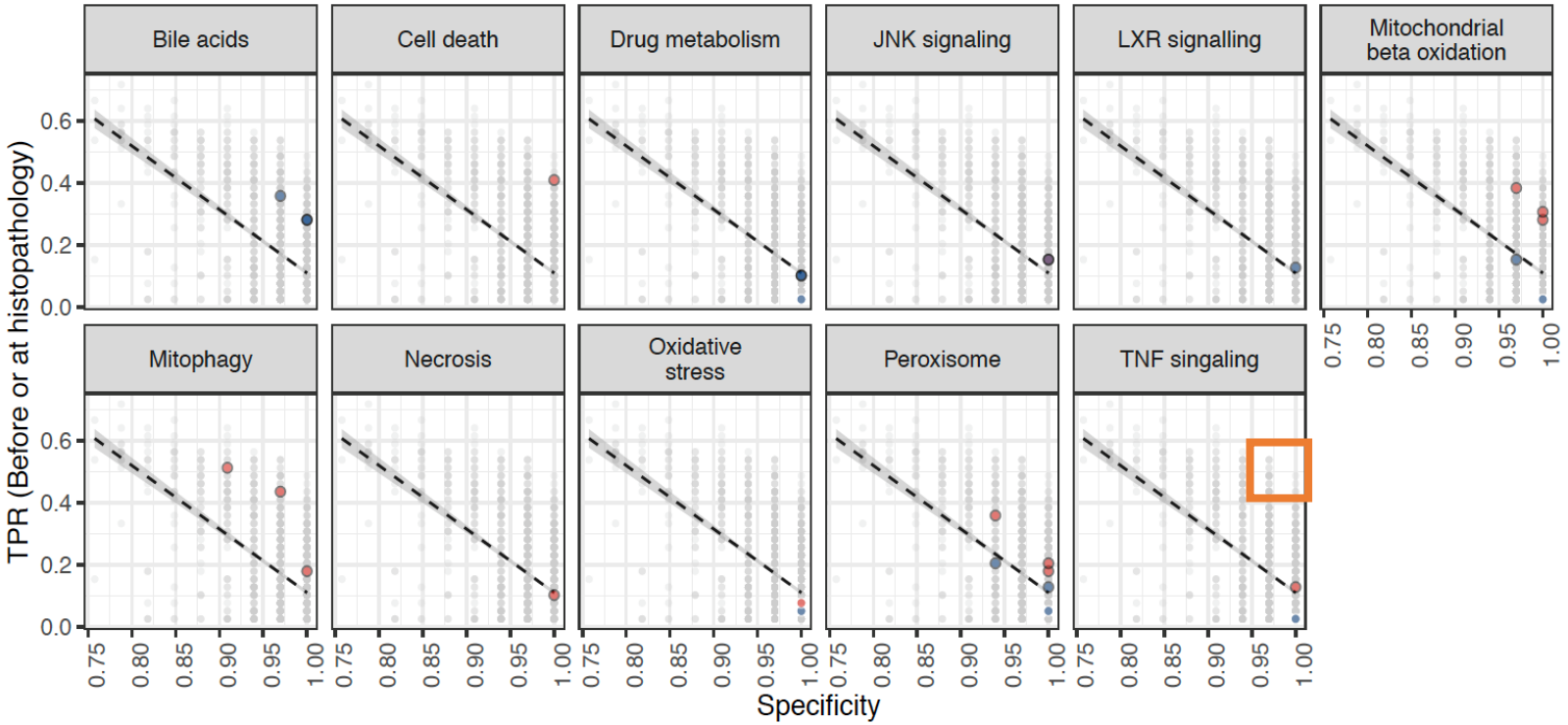
[1] S. Hänzelmann et al. , BMC Bioinformatics, 2013.
[2] A. Liberzon, Bioinformatics, 2011.
[3] J. J. Sutherland et al., Pharmacogenomics J., 2018.

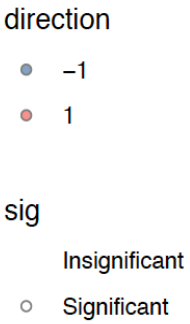
1st activation of any adverse histopathology label



All known key events are highly specific

Anchoring event	1st adverse histopathology
Temporal relation	Before or at
Background	Compounds-dose combinations without any histopathology





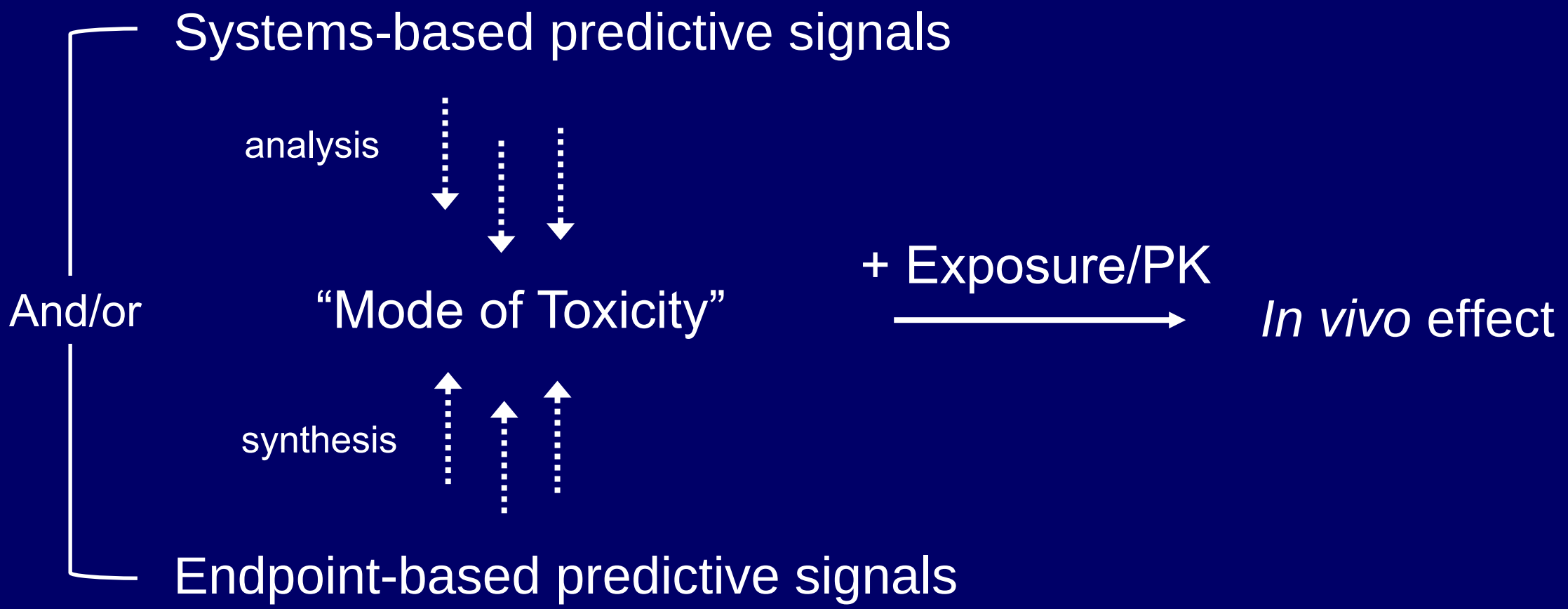
- Trade-off between specific and frequent events is not as pronounced as expected

My (personal) general learning w.r.t. high-dimensional biological readouts

Common practical difficulties with high-dimensional biology data (transcriptomics, also HCS *etc.*) are

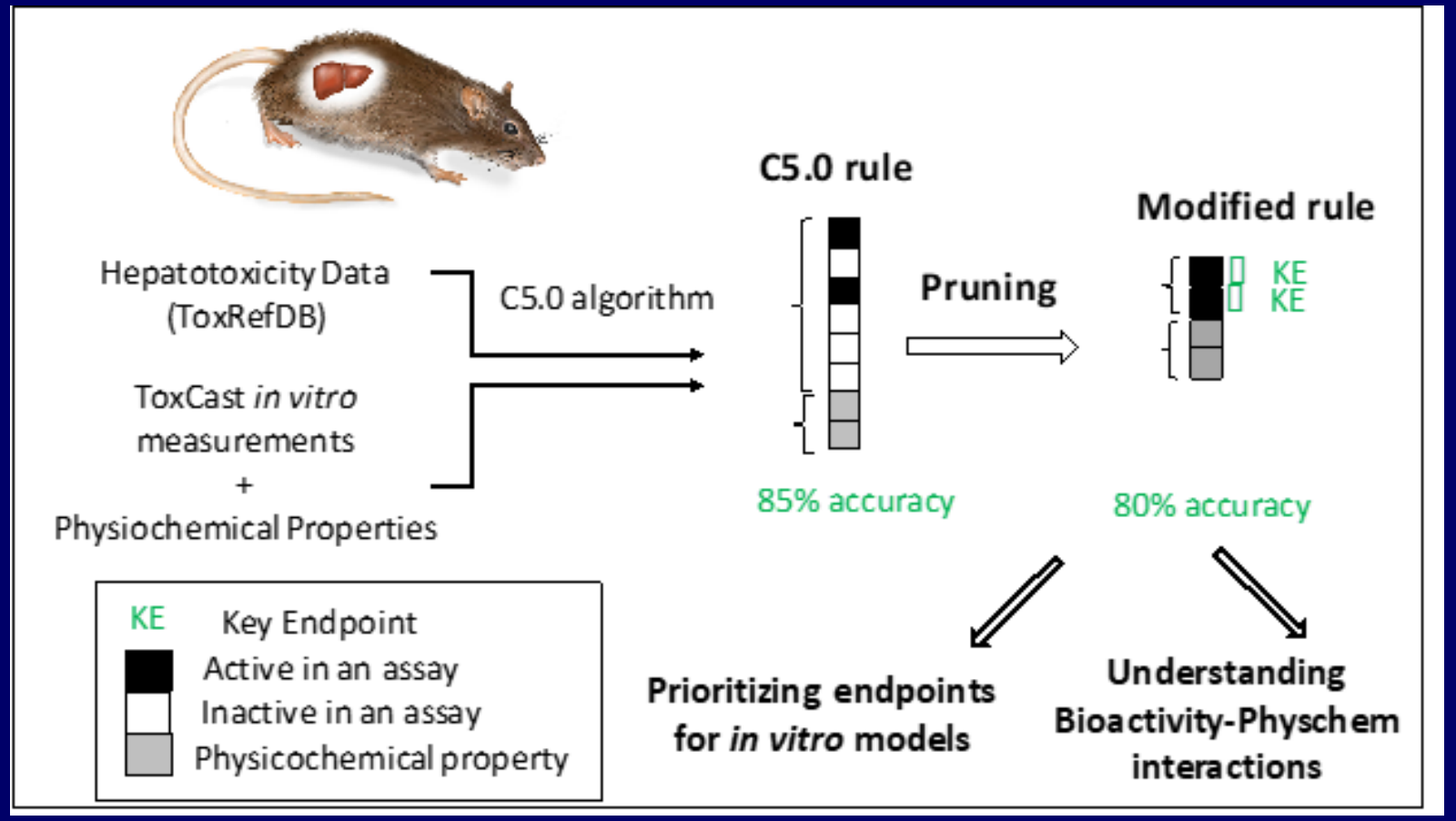
- *Many* choices to be made/issues with the data (biological system/dose/time point (!); reproducibility of controls, etc.)
- *Also* many choices to be made during analysis (choices determine what we see!)
- Data often contains sufficient signal for *signal detection* (but sometimes less so for 'modelling')
- Clear 'love/hate relationship' ☺ - 'works one third of the time, no (clear) signal one third of the time, too much signal one third of the time'… what to expect when?
- *What do we label/measure? Is it 'technology push', or 'science pull'?*
- **We need (a) relevance of the model system and (b) a hypothesis!**

‘Omics vs endpoint-based safety models: Conceptual differences and DIVI, DILI as case studies



Reverse-engineering organ toxicity from data

- Using modified rules to predict hepatotoxicity from ToxCast data; mechanistic, and PK-approximation
- Work of Samar Mahmoud

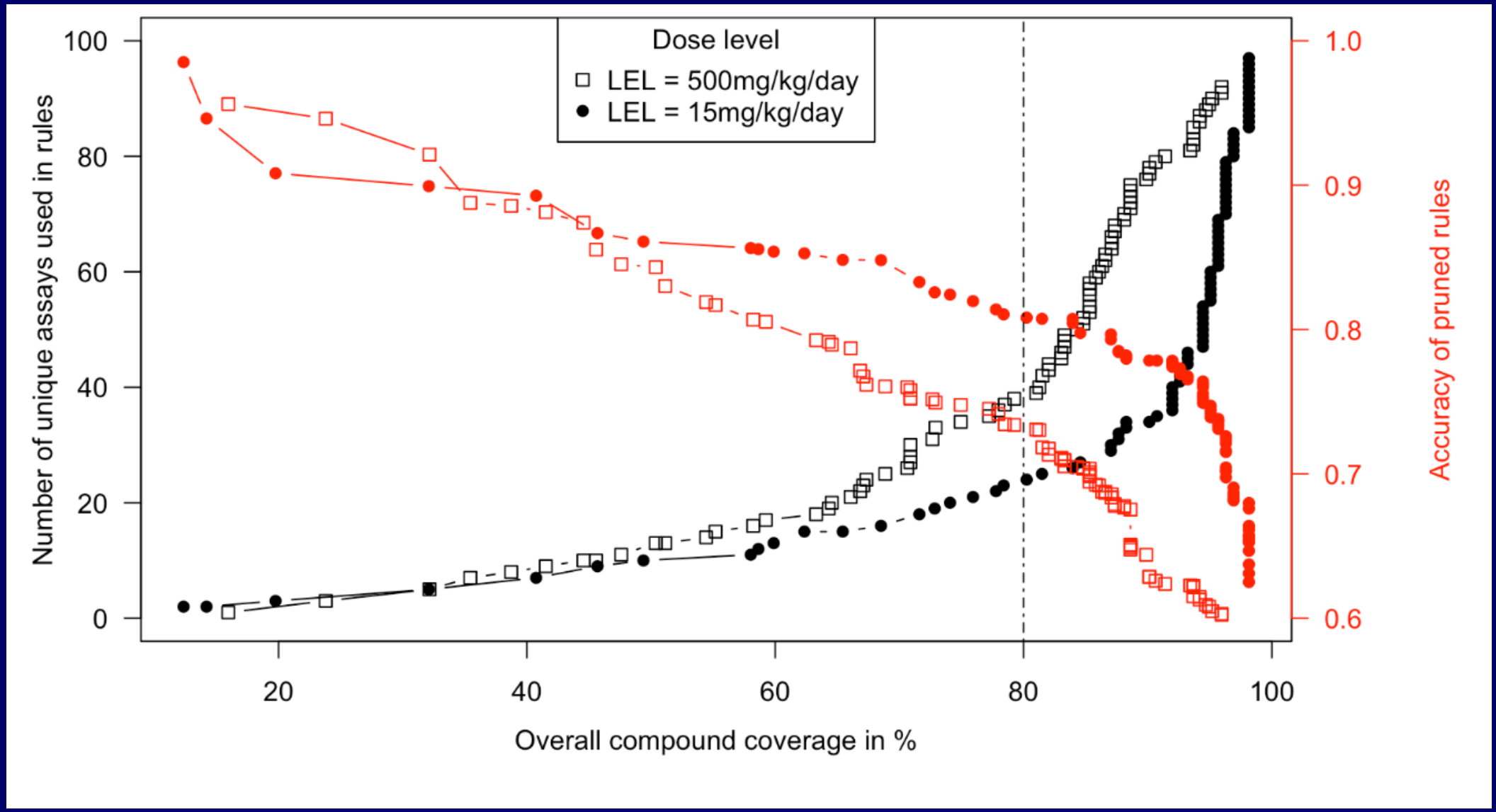


Data and Methods – ToxCast, ToxRefDB datasets, modified rules

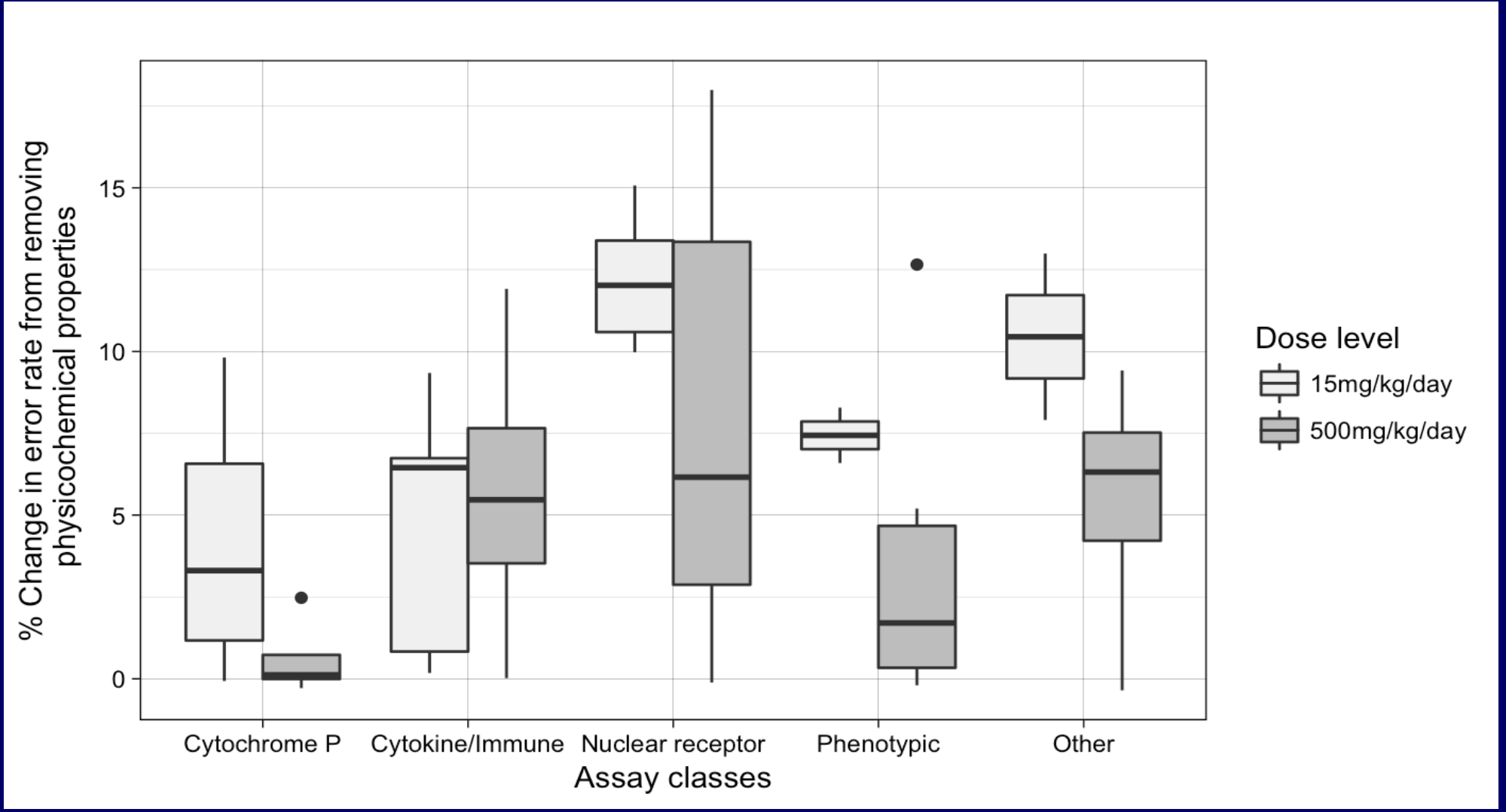
- Data: 673 compounds overlapping between 361 ToxCast assays (at least 5% valid AC50 values) and ToxRefDB hepatotoxicity readouts
- Hepatotoxicity at 15 and 500 mg/kg/day
- Added physicochemical properties as (crude) PK/PD approximation
- C5.0 classification rules, validated via 5-fold CV

- Manual rule modification: Retain rules that are meaningful (eg no negative activities for toxicity)
- Rule selection according to coverage and accuracy
- Note: Involves some manual steps, seems needed though (given limited data!)

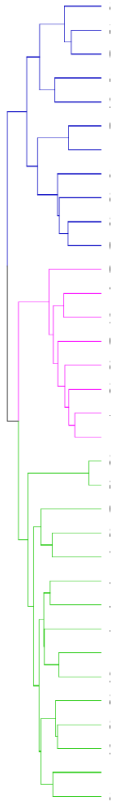
20 assays lead to 80% hepatotoxic compound coverage



Removing physiochemical properties led to deterioration of predicting low dose toxicity, less so high dose toxicity



Assays clustered based on shared rule membership gave CYP-related, immunological, and nuclear receptor-related groups



Bioactivity class	Index	Associated assay	Information gain (split)	Accuracy (rule)	Gene symbol	function
Activity against Cytochrome P	A.1	APR_HepG2_MitoMass_24h_up	0.019	0.921	NA	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
	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
Immunological activity	A.12	APR_HepG2_CellCycleArrest_72h_dn	**0.022**	0.749	NA	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
	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	NA	cell cycle
Nuclear receptor activity/ phenotypic readouts	A.22	APR_HepG2_MitoMembPot_72h_up	**0.020**	0.819	NA	cell morphology
	A.23	APR_HepG2_MitoMembPot_1h_dn	**0.029**	0.888	NA	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	NA	cell morphology
	A.26	APR_HepG2_OxidativeStress_1h_up	**0.025**	0.789	NA	cell cycle
	A.27	APR_HepG2_StressKinase_1h_up	**0.052**	0.956	NA	cell cycle
	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	NA	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	NA	cell cycle
	A.37	└ ATG_p53_CIS_up	0.014	0.946	TP53	dna binding

- Used for suggesting assays to evaluate hepatotoxicity
- Comparison to commercial hepatotoxicity assays gave mostly overlap, plus additional suggestions

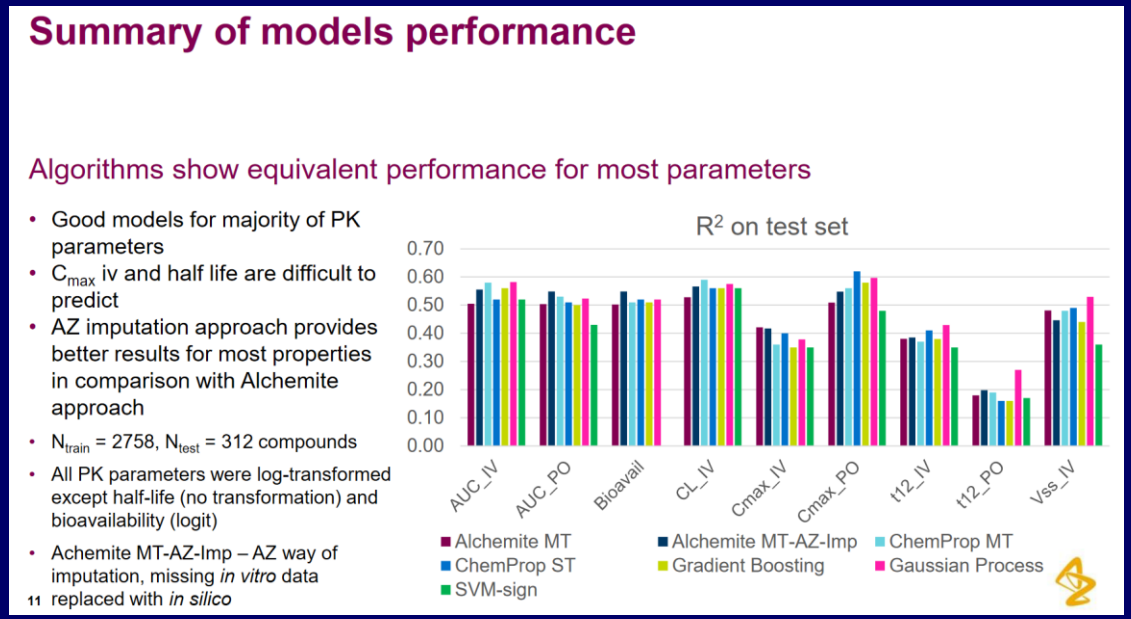
Machine learning models for PK

- *In vivo* PK data (rat, dog, mouse) available on large scale (1,000s-10,000s of compounds)
- ML models, based on ligand structure only
- Bayer, AstraZeneca, … models
- Don't require IVIVE; consider 'all' mechanisms
- Predictivity *en par* with/ better than e.g. well-stirred models

J Chem Inf Model. 2019 Nov 25;59(11):4893-4905. doi: 10.1021/acs.jcim.9b00460. Epub 2019 Nov 12.

Prediction of Oral Bioavailability in Rats: Transferring Insights from in Vitro Correlations to (Deep) Machine Learning Models Using in Silico Model Outputs and Chemical Structure Parameters

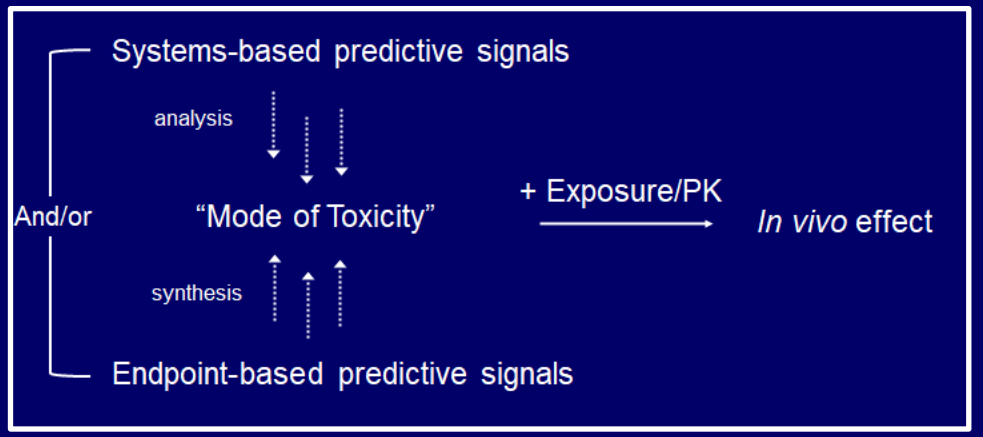
Sebastian Schneckener [1], Sergio Grimbs [1], Jessica Hey [1], Stephan Menz [2], Maren Osmers [2], Steffen Schaper [1], Alexander Hillisch [3], Andreas H Göller [3]



http://www.drugdiscovery.net/tox2020/data/obrezanova.pdf

So where do we stand with data in safety today?

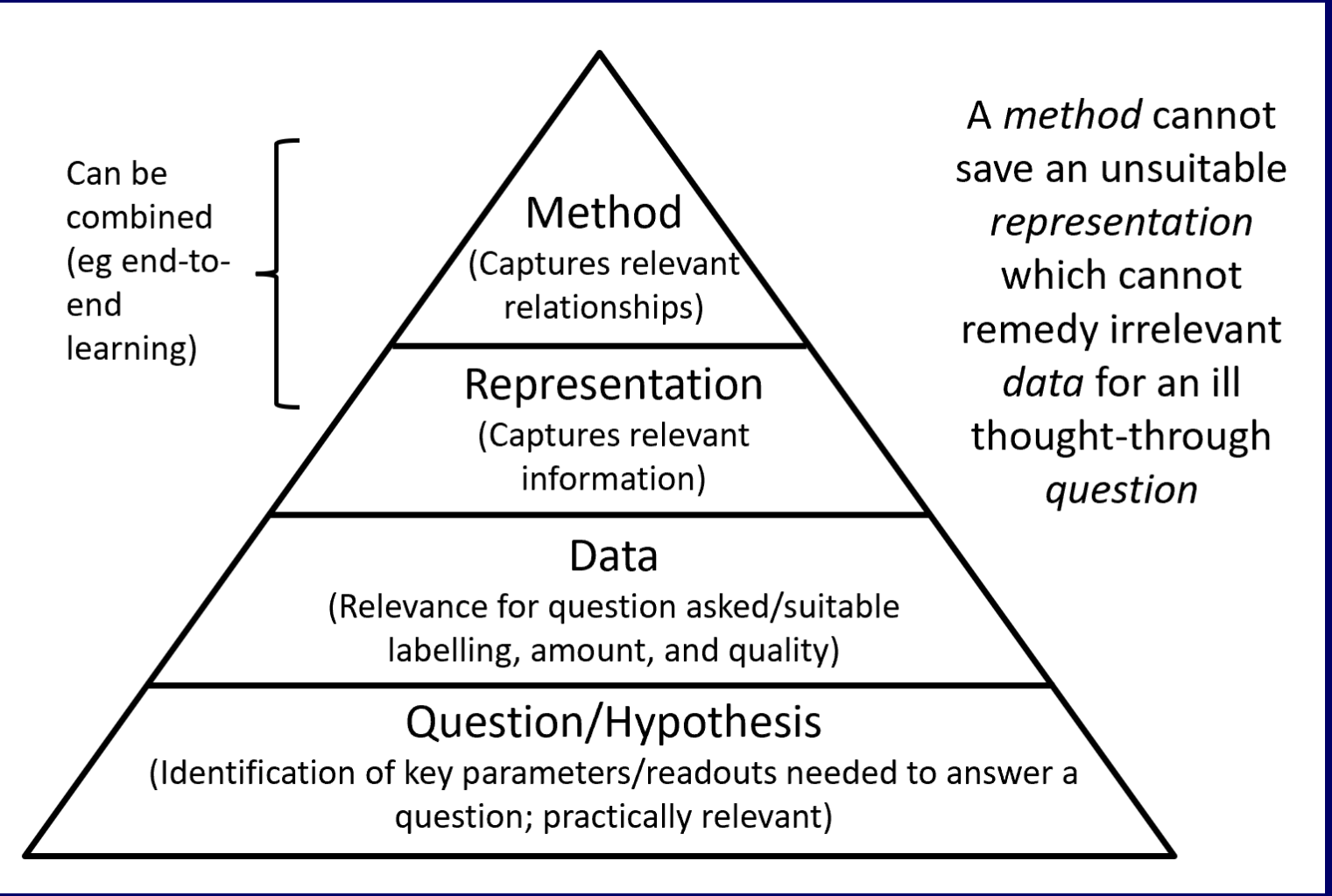
- Often proxy measures (to reduce cost)
- Historical data gets repurposed now ‘for AI’
- Not always relevant system/dose/time point



- “Models of models” – “the *in silico* model of the Glu/Gal mitotoxicity model” … is then meant to predict the *in vivo* situation
- We need to care more about modelling the actual endpoint of interest (say, organ risk), not the proxy (say, assay) endpoint!

- Often hypothesis-free (‘here we have our pile of data … anyone wants to have a go at it?’) instead of hypothesis-driven
- Often ‘technology push’, instead of ‘science pull’

The *question* needs to come first… and then the data, then the representation, and then the method
http://www.DrugDiscovery.NET/HowToLie



Lots of attention currently here…

But we need to care more about this

Summary

- Chemical and biological data is different from images, speech
- This makes applicability of ‘AI’ in drug discovery (and safety) not trivial
- Both ‘omics/high-dimensional biology, and target-based approaches have their value
- Impact of both experimental setup of data generation, and subjective choices during data analysis (!) not to be underestimated
- Currently a lot of computer science-driven approaches, some of which are more applicable in drug discovery than others (real translation is necessary, *but also better experimental design!*)
- Consortia on even larger scale are likely needed (for targeted data generation, not just sharing what is there already)

Resources

Artificial Intelligence in Drug Discovery – What is Realistic, What are Illusions?

Part 1: Ways to make an impact, and why we are not there yet

Part 2: a discussion of chemical and biological data

Andreas Bender and Isidro Cortes, *Drug Discovery Today* 2021 (in press)

http://www.DrugDiscovery.NET/AIReview

“How to Lie With Computational Predictive Models in Drug Discovery”

http://www.DrugDiscovery.NET/HowToLie

Thank you for listening!
Any questions?

Contact: ab454@cam.ac.uk
Personal email: mail@andreasbender.de
Web: http://www.DrugDiscovery.NET
Twitter: @AndreasBenderUK