Transcriptomics!

Roland Grafström

Karolinska Institutet
Karolinska Innovations/Predictomics AB
Stockholm, Sweden

Challenges in Public Health Protection in the 21st century: New Methods, Omics and Novel Concepts in Toxicology

15-17 Nov 2021, Berlin, Germany

Predictive Toxicogenomics Space Modelling: Aims and Purpose

Modelling together **large collections of gene expression and high-throughput cellular screening profiles** (i.e., “Big Data “) should generate variants of **toxome** descriptions

Such a description should be able to serve as a “**Predictive Toxicogenomics Space (PTGS)**“ as it should capture **toxicity mechanisms and pathological effects**

Bioinformatics-based validation against existing and generated “big data” sets should prove the extent of usefulness of a potential “**high-throughput PTGS-based scoring concept**” for:

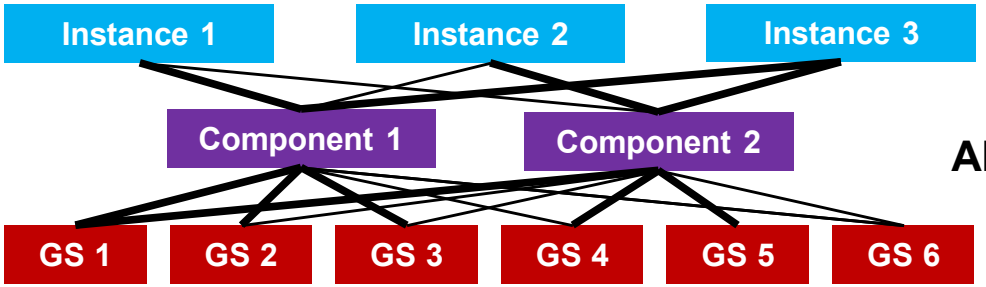
predicting Key Events for cellular and organ toxicity effects,

analyzing dose-dependent relationships for diverse agents,

all to be useful to **Adverse Outcome Pathway (AOP)** studies

”Toxicogenomics Space” is defined by ”omics” components predictive of cytotoxicity

Connectivity Map (3062 instances)



Latent Dirichlet Allocation component models (100)

Molecular Signatures Database (1321 gene sets)

Connectivity Map
(1217 compounds)
(100 components)
(MCF7, HL60, PC3)

→

Cross-over data set
222 compounds
492 instances

↓

←

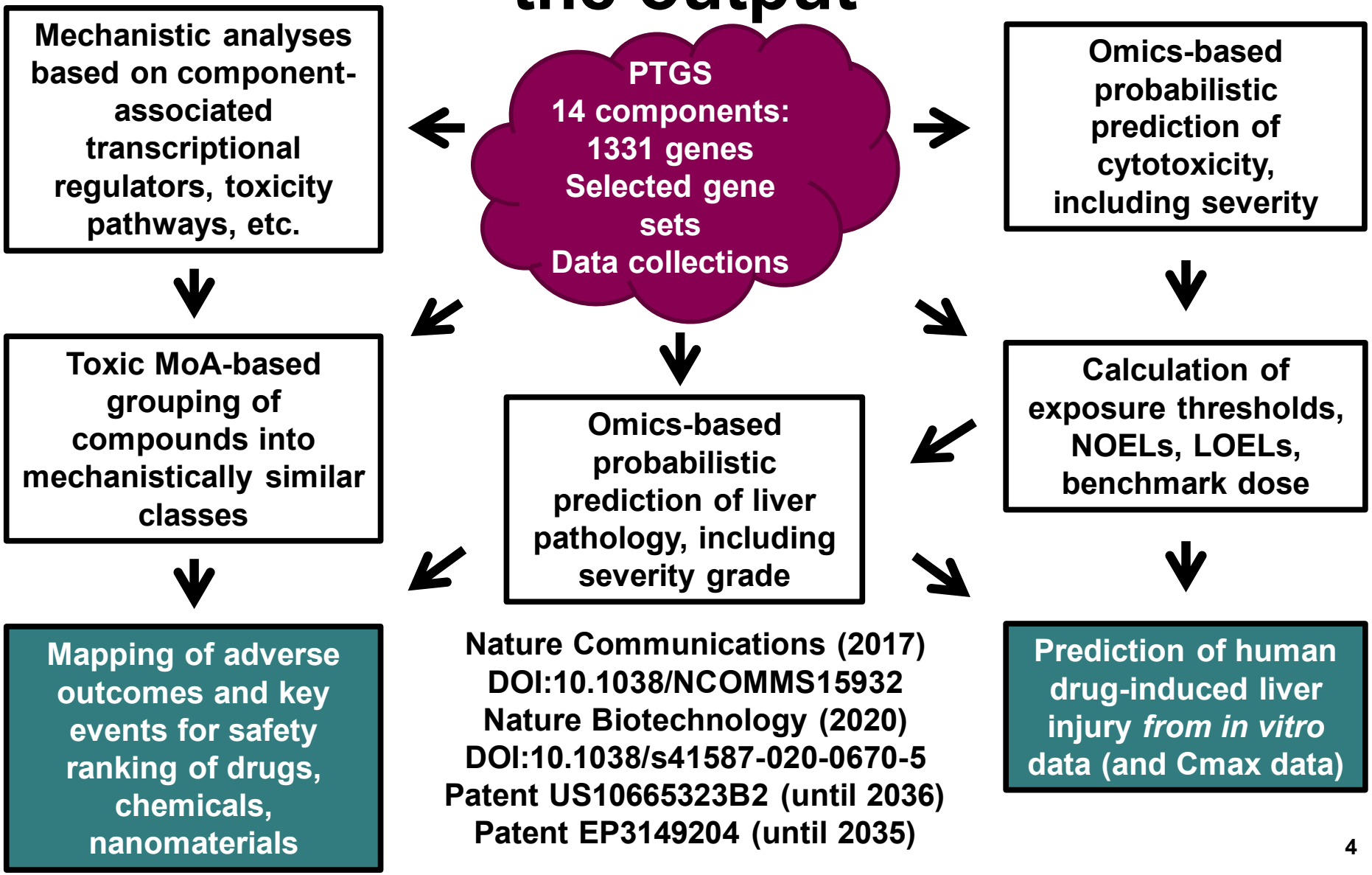
NCI-60 DTP
(100000 compounds)
(59 cell lines)
(GI50/TGI/LC50 data)

Predictive Toxicogenomic Space (PTGS)

14 of the 100 component models, 1331 genes

Component-based scoring	**Gene set-based scoring**
(tests if the 14 components are more active than the other 86 components)	(tests if the PTGS-associated genes are more active in the treated vs. non-treated)

The PTGS safety scoring concept – the output



Nature Communications (2017)
DOI:10.1038/NCOMMS15932
Nature Biotechnology (2020)
DOI:10.1038/s41587-020-0670-5
Patent US10665323B2 (until 2036)
Patent EP3149204 (until 2035)

DILI assessments with the Predictive Toxicogenomics Space (PTGS) concept (>500 million data points applied)

Data sets	Compounds	Tests	Samples	Data points
TG-GATEs rat repeated dose 28-day study, prediction concept/validation (MA)	143	1689	6765	128 651 751
TG-GATEs human hepatocytes, prediction concept/validation (MA)	157	941	2605	90 669 391
TG-GATEs rat hepatocytes, prediction concept/validation (MA)	145	1260	3370	76 692 128
DrugMatrix rat liver, in vivo, repeated dose, validation (MA)	201	654	2218	56 752 735
DrugMatrix rat hepatocytes, validation (MA)	126	268	939	25 671 374
Benchmark dose (BMD) rat liver, *in vivo*, validation (RNA-seq/MA comparison)	1	12	60	986 788
HepG2 cell model, TempO-Seq S1500+, validation (HTTr)	81	160	489	5 730 967
DILI prediction, rat liver, *in vivo*, blinded study, validation (MA)	1	4	24	891 746
Human and rat liver, *in vitro* systems comparison, blinded study, validation (MA)	3	87	439	14 427 646
DILI prediction, rat liver, *in vivo*, blinded study, validation (MA)	1	6	45	1 236 342
DILI prediction/BMD, human liver spheroids, validation (HTTr)	28	560	2774	67 779 442
DILI prediction, human liver spheroids, LINCS L1000+Inferred, validation (HTTr)	28	560	2774	44 018 024
DILI prediction, human liver spheroids, blinded study, validation (RNA-seq)	27	87	269	6 084 001
Total	***942***	***6288***	***22 771***	***519 592 335***

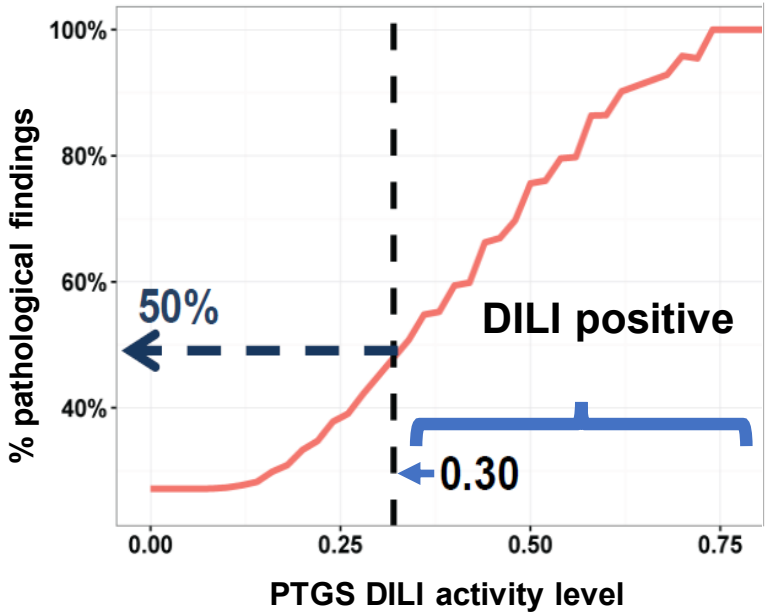
Unique compounds 453; 231 with DILI information (FDA DILIRank DB): 85 Most-DILI-concern, 87 Less, 27 Ambiquous and 32 No-DILI-concern; 119 compounds in total (74 Most, 36 Less, 14 No, 13 Ambiguous, 18 Unclassified).

In vitro model predictions included 119 compounds of which *92 were correctly predicted (77%).*

Study calculations include raw data and derived analyses data of gene expression at transcriptome level.
"Blinded study" indicates unrevealed compound identity and/or DILI classification at start of analysis.
MA = microarray technology, HTTr = High-throughput platforms, RNA-seq = RNA sequencing technology

Scoring concepts for DILI and cytotoxicity prediction: defining LOELs

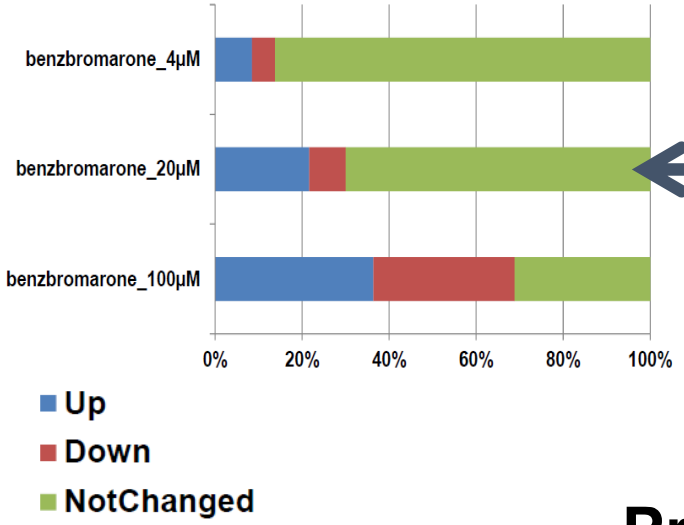
- **Gene set enrichment analysis, adjusted p-value (stat significance; FDR <0.05)**
  - ***R/Bioconductor limma*** **rotation-based testing (10000 rotations)**
  - **Tests whether PTGS changes relative control**
  - **Uses all gene expression information, not just DEGs**
- **Activity score relative TG-GATEs or the Connectivity Map (biological effect, > 50% effect probability)**
  - **Use the proportion of genes in set(s) altered by the exposure as a score**
  - **Compared directly to the TG-GATEs rat 28-day liver data (1667 treatments) for DILI and Connectivity Map /NCI-60 DTP for cytotoxicity (492 treatments)**
  - **Point where at least 50% of treatments have pathological/cytotoxic effects (DILI/GI50) is used as the threshold**



Nature Communications (2017) DOI:10.1038/NCOMMS15932

The PTGS safety scoring concept examplified with TG-GATEs data (Benzbromarone, 4-100 µM; 8 h, human hepatocytes, therapeutic $C_{max, total}$ 6.6 µM)

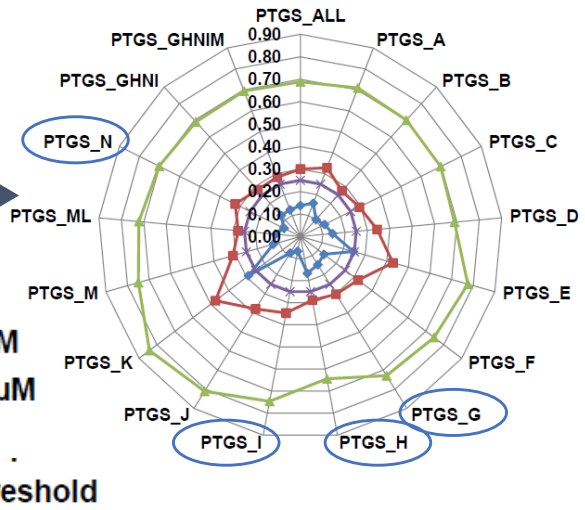
PTGS gene set/space activity



PTGS

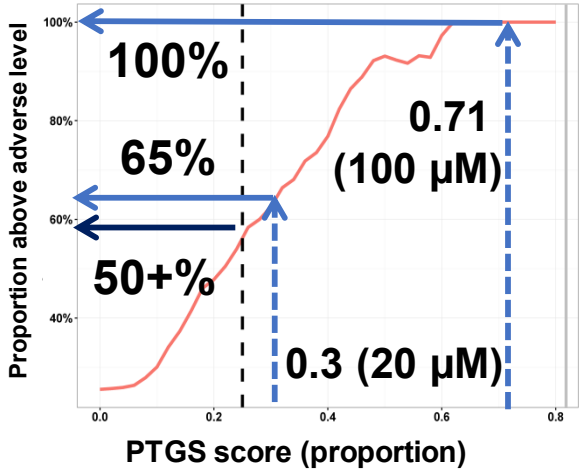
14 components
1331 genes
Selected gene sets, Data collections

PTGS component MoA

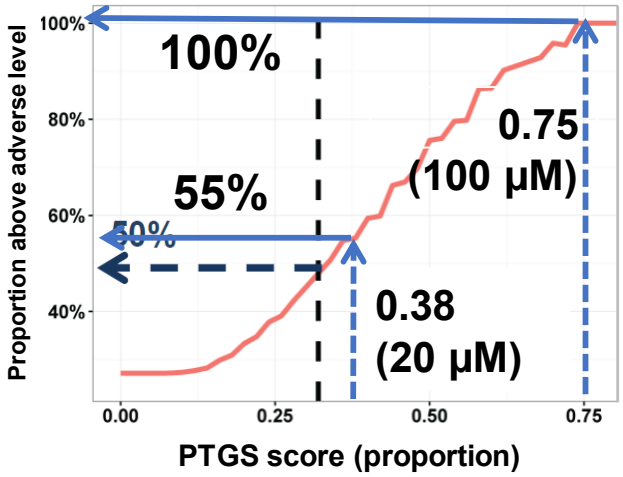


Probabilistic prediction of

Cytotoxicity LOEL (20 µM)

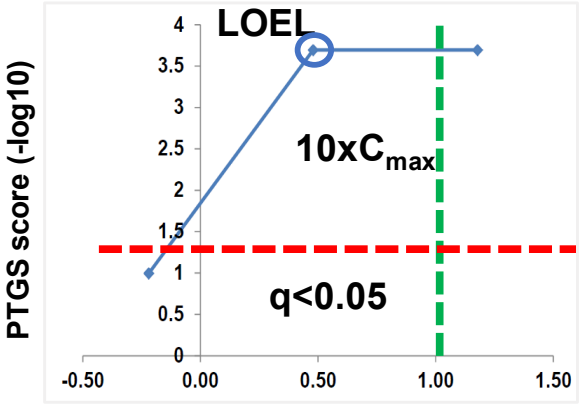


DILI LOEL (20 µM)



Human DILI (DILI +)

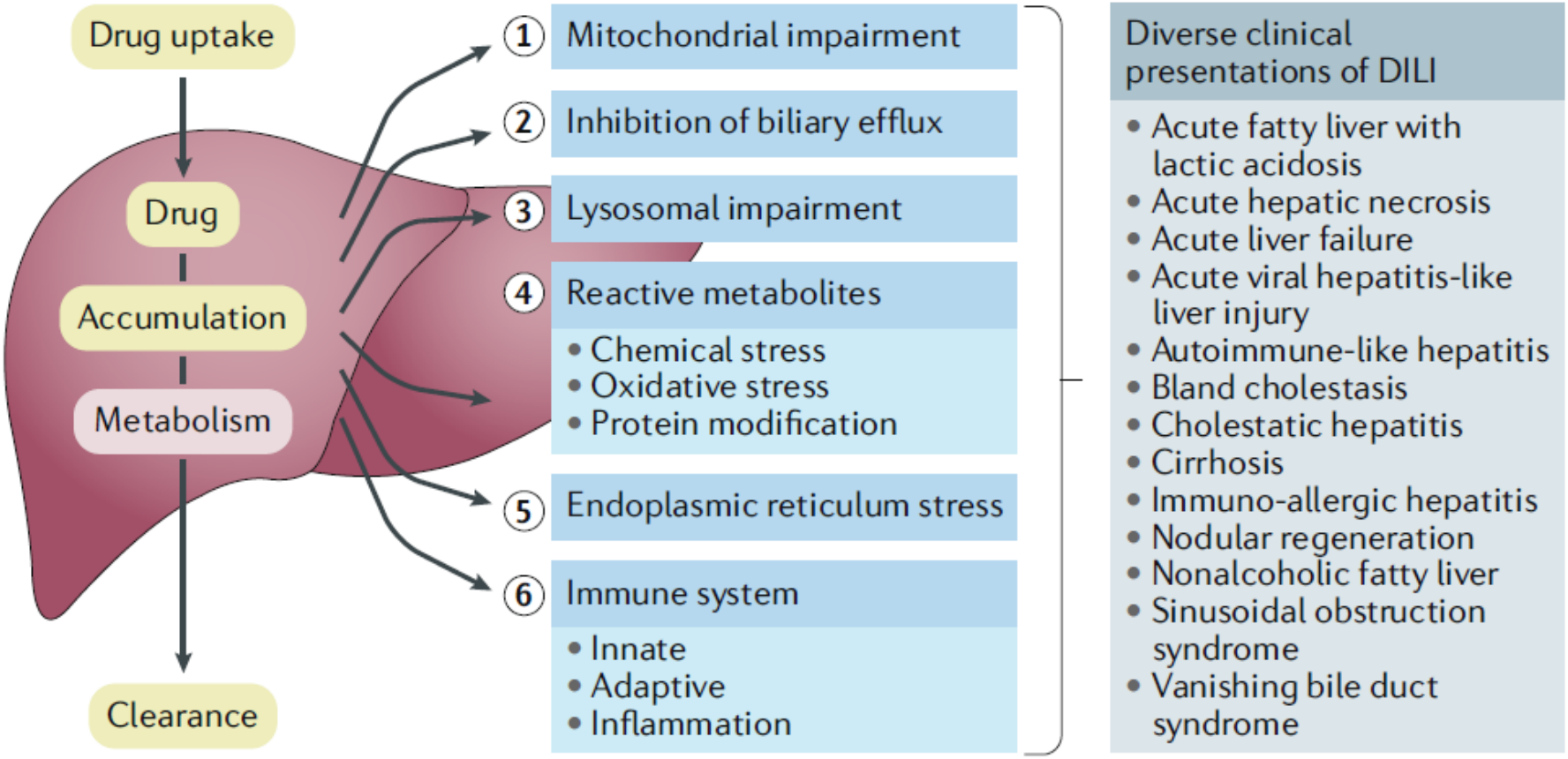
Safety margin (0.5; 3.3x)



PTGS concept enables component-based MoA analysis; G,H, I, N components are applied to liver toxicity prediction; (Component activations also serves for compound grouping)

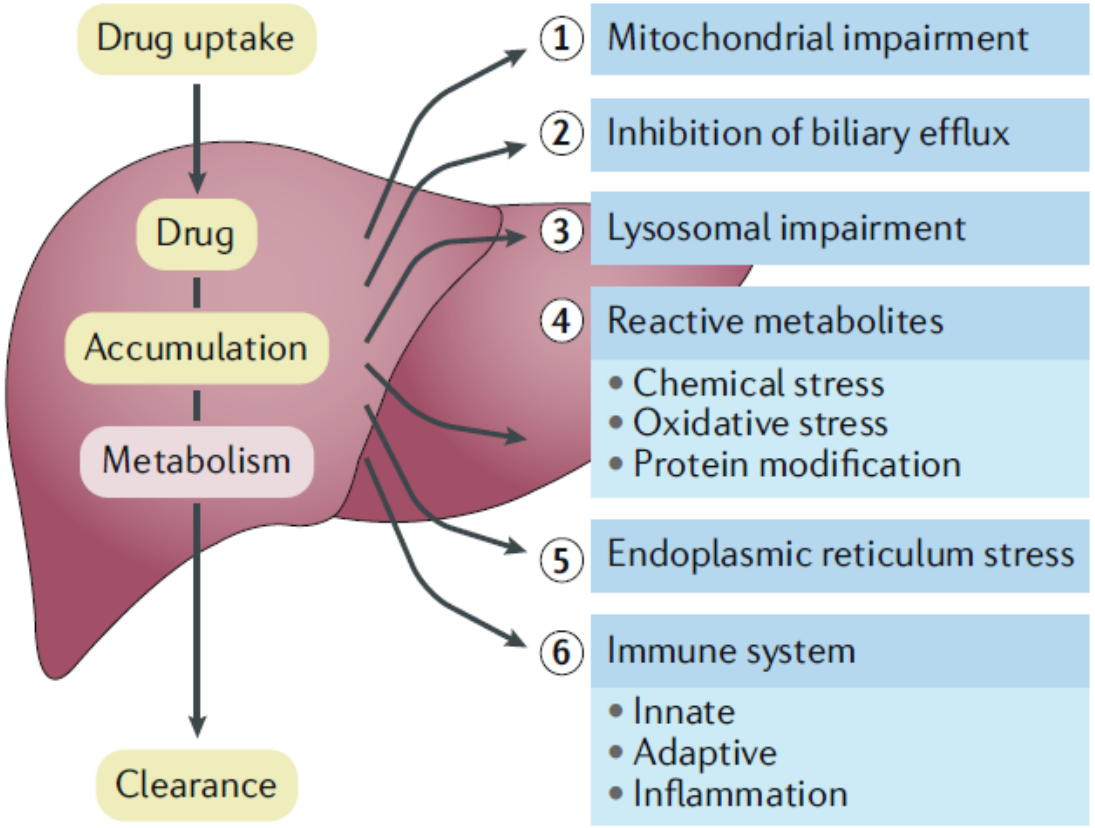
Component	Toxicity-associated biological and cellular mechanisms
A, B, C, G, H, I, N	PPARa/RXRa Activation, Peroxisome Proliferators via PPARa, LXR/RXR Activation, VDR/RXR Activation, RAR Activation, Aryl Hydrocarbon Receptor Signaling, NF-kB Signaling, Oxidative Stress, NRF2-mediated Oxidative Stress Response, TGF-b Signaling, Transmembrane Potential of Mitochondria, Anti-Apoptosis, Cell Cycle: G1/S Checkpoint Regulation, p53 Signaling
D	TGF-b Signaling, PPARa/RXRa Activation
E, K	Cell Cycle: G1/S Checkpoint Regulation and G2/M DNA Damage Checkpoint Regulation, Aryl Hydrocarbon Receptor Signaling, p53 Signaling, Notch signaling, E2F/MYC targets, peroxisome
L	Cellular aldehyde metabolic process (HMGCL, ABAT, ADH5, PGD)
F	Regulation of transcription, DNA-dependent, positive regulation of transcription from RNA polymerase II promoter, UV response
J	RNA polymerase II promoter regulation, IL2-STAT5 signaling
M	tRNA charging, unfolded protein response, MTORC1 signaling

PTGS tool captures chemical insults that lead to diverse clinical manifestations of DILI



Drug-induced liver injury (DILI) can be caused by various chemical insults (steps 1–5) and can present as an array of different pathologies, dependent on the specific function of the liver that is impaired. Furthermore, recruitment of the immune system (step 6) can result in a prolonged or altered pathological phenotype, adding further complexity to the clinical presentation of the condition.
(Fig from Weaver et al, Nature Reviews-Drug Discovery, 2020)

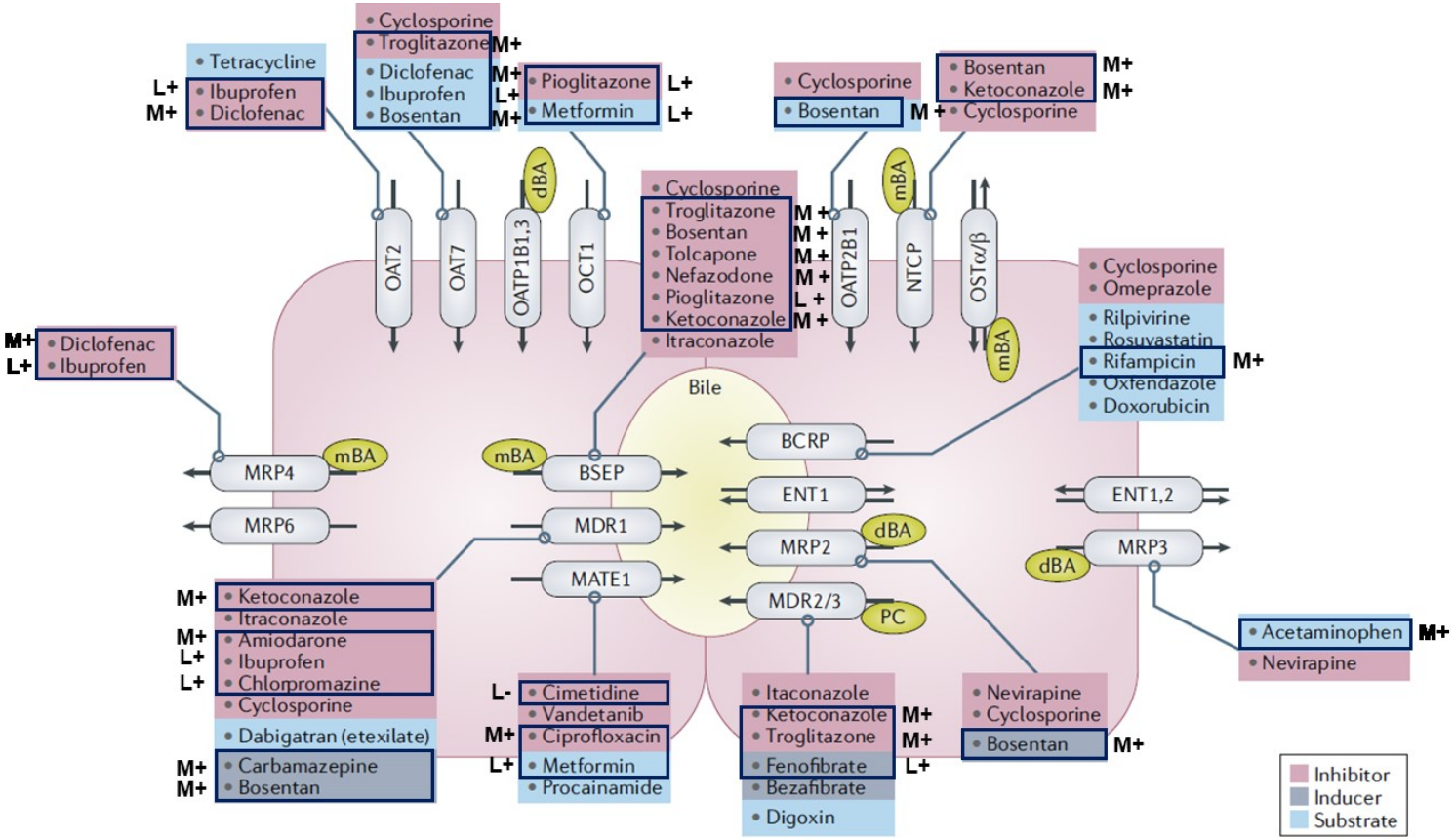
PTGS components capture toxic mechanisms associated to DILI



	Gene Ontology	toxLists (IPA)
1 Mitochondrial impairment	B, C, G, I, M	C, G, I
2 Inhibition of biliary efflux		A, B, C, D, E, F, G, H, I, K, N
3 Lysosomal impairment	H	
4 Reactive metabolites		
• Chemical stress	A, C, F, M	A, B, C, G, N
• Oxidative stress	B, D, G, I	A, B, C, D, E, G, H, N
• Protein modification	G	
5 Endoplasmic reticulum stress	B, D	
6 Immune system		
• Innate	B, N	
• Adaptive	B, C, J	
• Inflammation	B	
	B, G, I, N	

Figure adapted from Weaver et al, Nature Reviews-Drug Discovery, 2020

PTGS scoring in vitro captures DILI concerns related to hepatobilary transport (33 of 34 DILI-inducing drug molecular actions indicated as captured)



TGS analyses data ombined from public nd client liver model xperiments

= Most DILI concern ompounds
= Less DILI concern ompounds

= PTGS LOEL DILI ositive scores
= PTGS LOEL DILI egative scores

Cimetidine is negative in both TG-GATEs rat and human hepatocytes

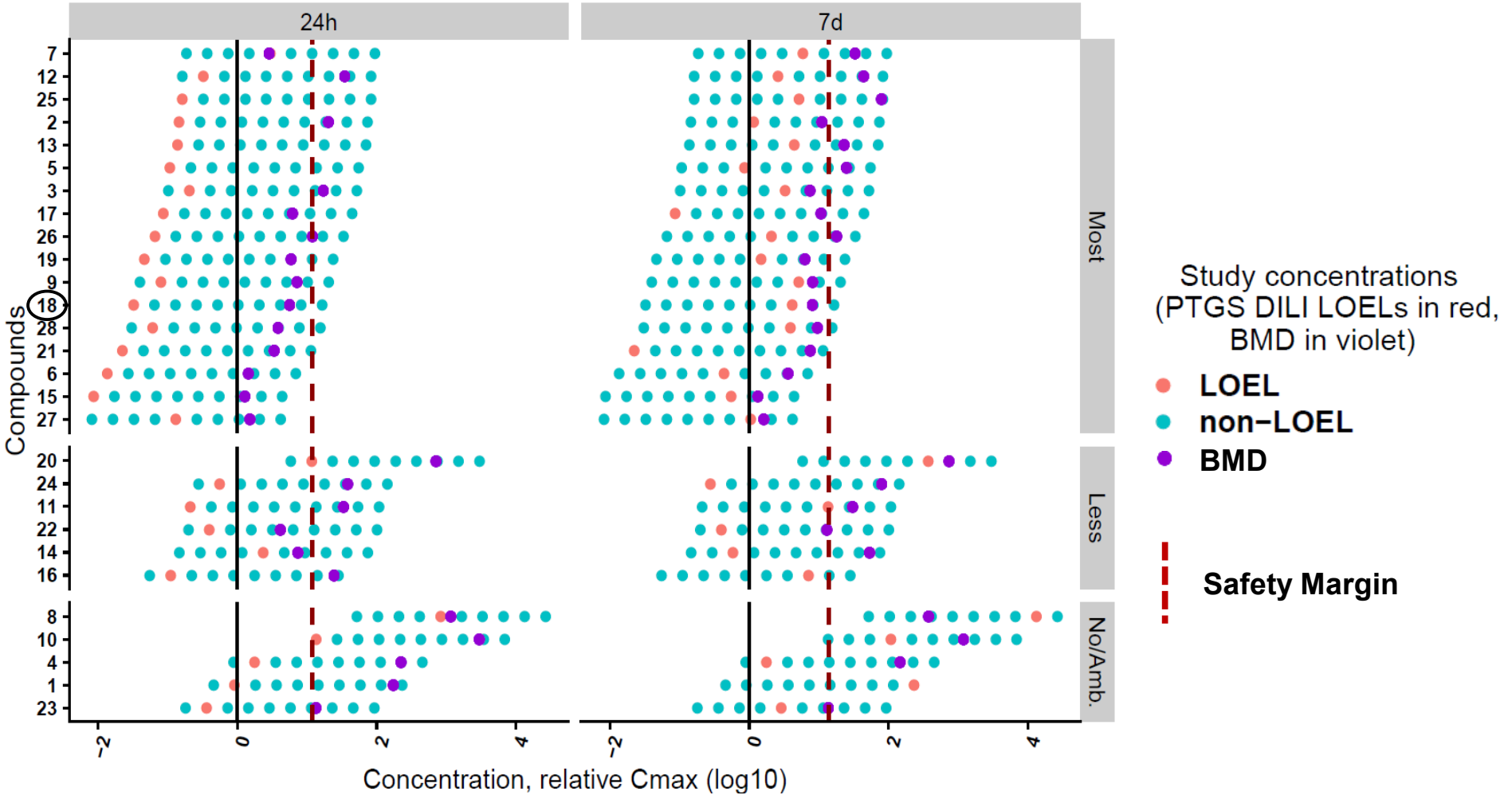
Figure adapted from Weaver et al, Nature Reviews-Drug Discovery, 2020

Types of PTGS Component MoA analyses

- **Analysis applying Latent Dirichlet Allocation (LDA) component models**: used for selecting component sets for tissues and cells, e.g., DILI
- **Self-contained Gene Set Enrichment Analysis (GSEA):** limma ROAST assesses activity of component genes in exposed vs. control; used for dose response analysis and deriving PTGS-LOEL data
- **Competitive GSEA**: version 1) limma ROMER analysis of component genes relative other PTGS genes, or version 2) relative other genes that are part of the 100 CMap LDA-modelled components; serves for MoA and AOP / KE analyses
- **BMD analysis with BMDExpress2:** analyses each measured gene expressing a dose response with 10+1 US EPA models; gives a summarized result with the optimal model(s) for each activated component
- **BMD analysis with BMDExpress2 using a novel single-sample GSEA method**: more sensitive than above method, less computation; gives a BMD with a single optimal model at the component level
- **Connectivity mapping (PharmaCoGX method, global weighted correlation)**: gives directional connectivity; used to connect components to an in-house generated LINCS perturbation class meta-signatures data set

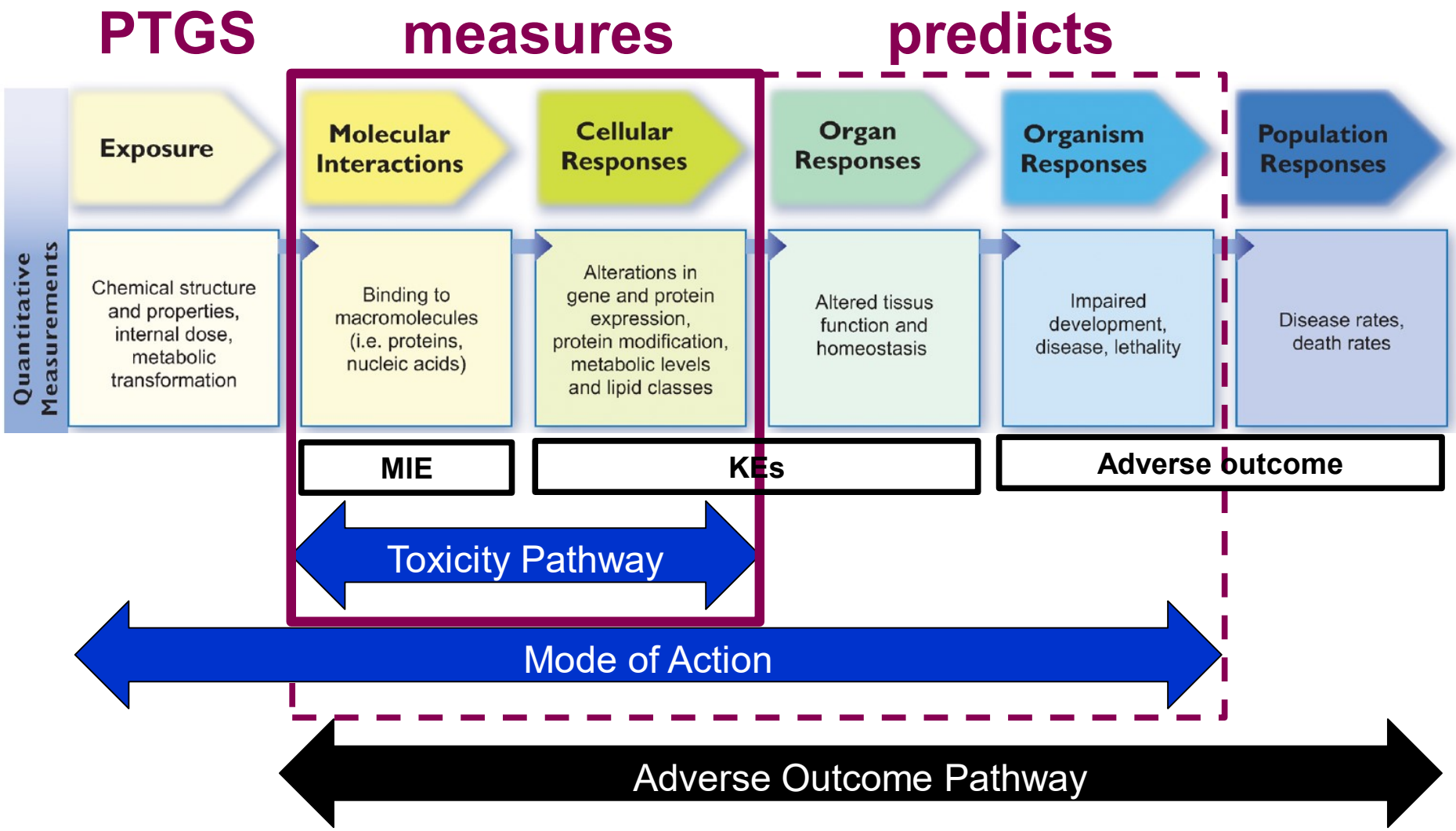
Example: PTGS and BMD analysis for a client

Excerpt from PTGS Toxicity/DILI Prediction Model (28 compounds)



PTGS test concept is highly sensitive. PTGS LOELs correctly predict DILI concern for 27 of 28 compounds (slide shown with permission from Predictomics AB)

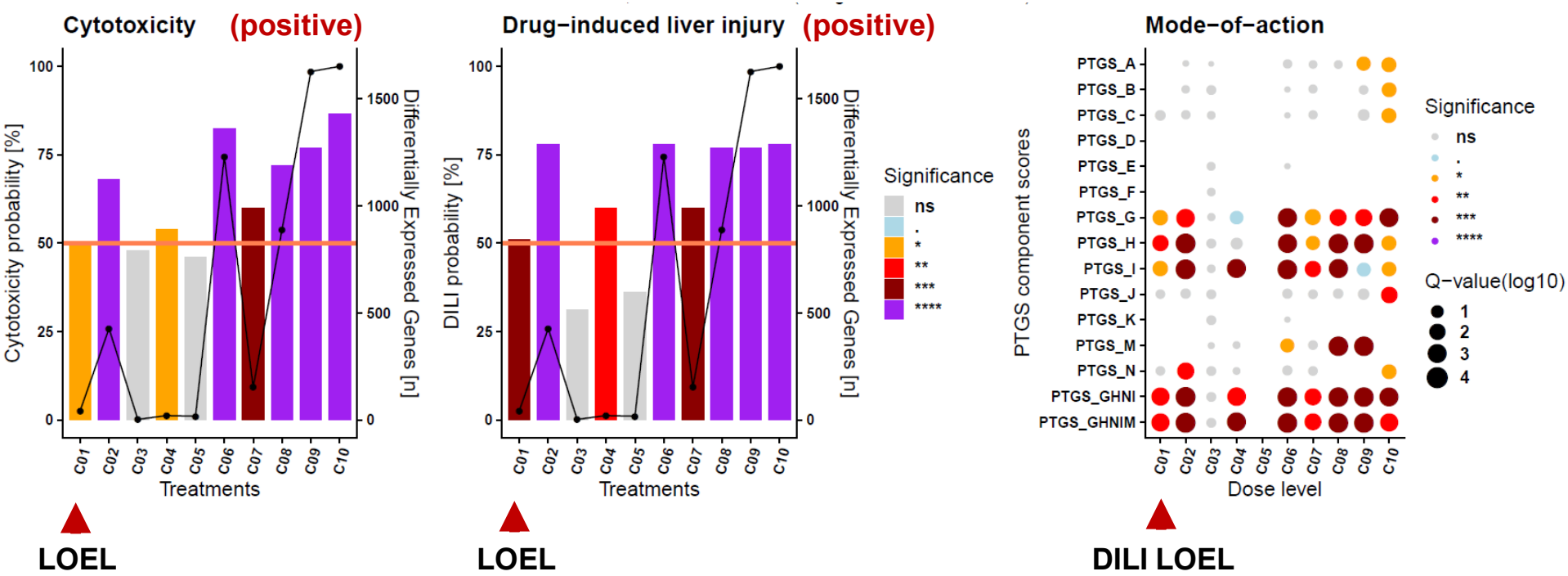
Programmatic PTGS-driven AOP analysis assesses 26 Liver AOPs coupled to 67/90 events (MIEs and KEs)



Ankley et al., 2010; Technical information on alternative methods (CADASTER workshop on the use of QSAR models in REACH, Slovenia, 1-2 September 2011) by Andrew Worth, European Commission, Joint Research Centre, Systems Toxicology Unit, Italy

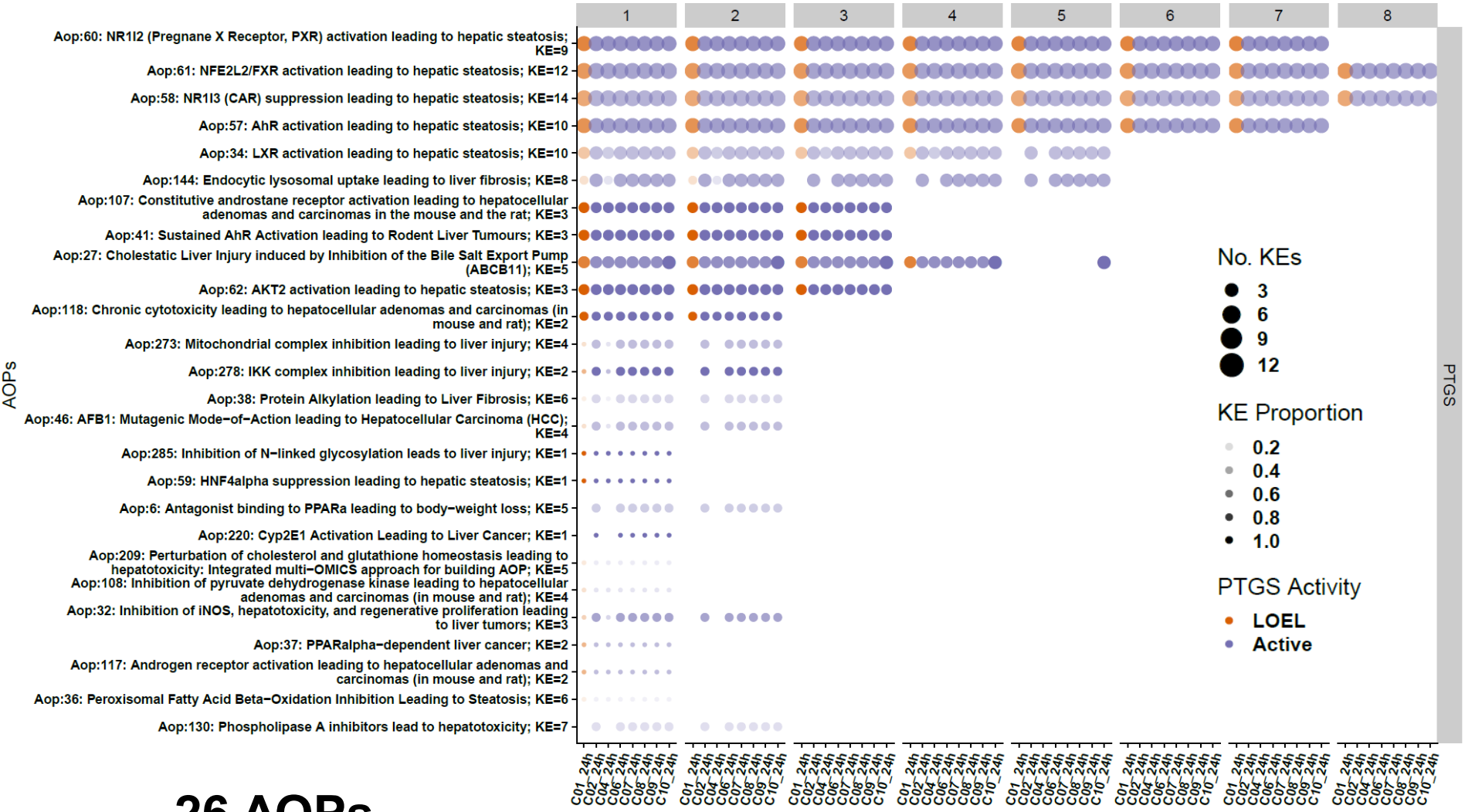
Sturla SJ, et al. Systems toxicology: from basic research to risk assessment. Chem Res Toxicol. 2014 Mar 17;27(3):314-29.

Compound 18, Most−DILI−Concern (24h) causes steatosis and liver fibrosis



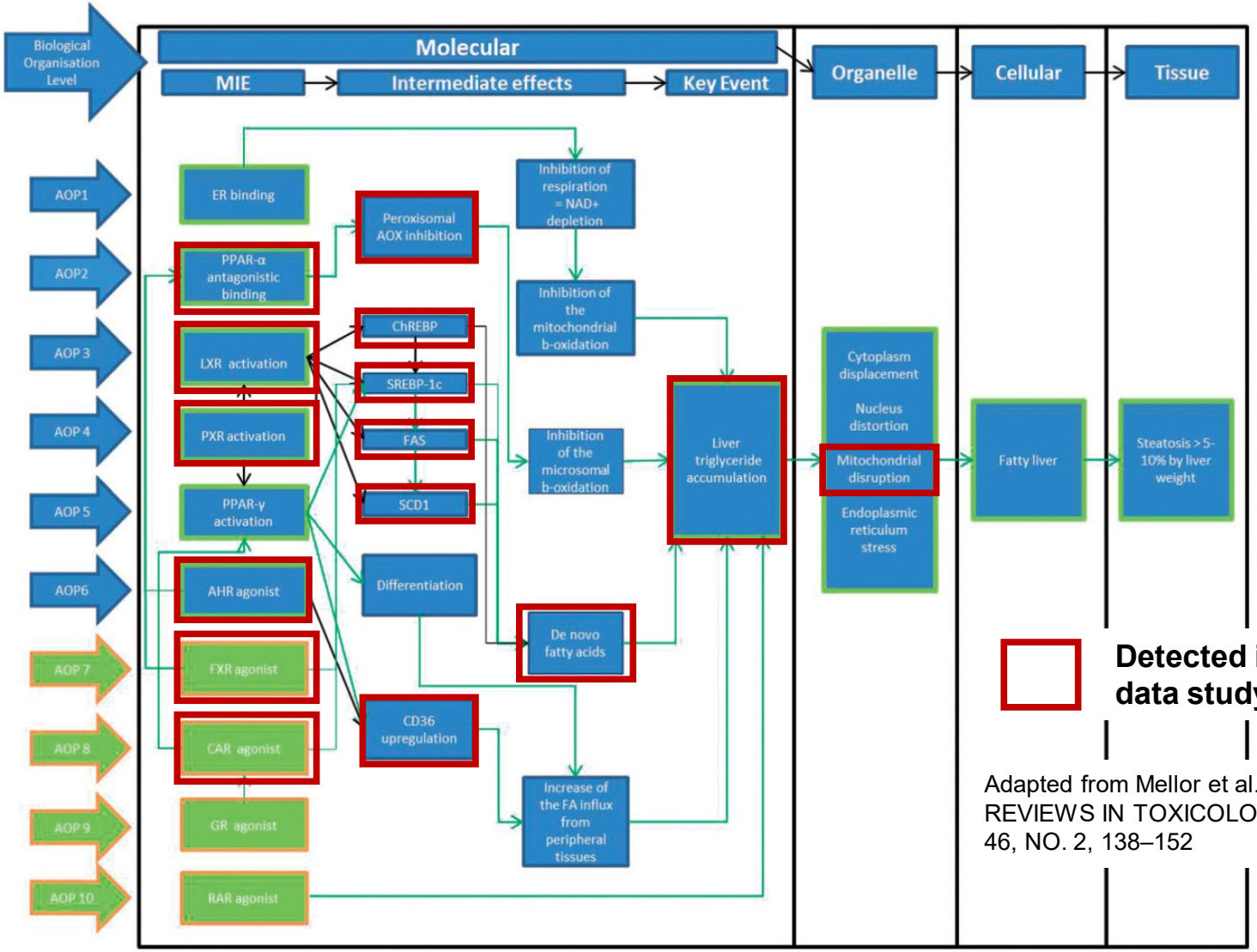
DILI LOEL: 0.98 µM, C01 (S.M. -1.49)
Cytotoxicity LOEL: 0.98 µM, C01 (S.M. -1.49)

Compound 18, liver AOPs: KE per AOP (PTGS DILI LOEL 24h: C01)



26 AOPs

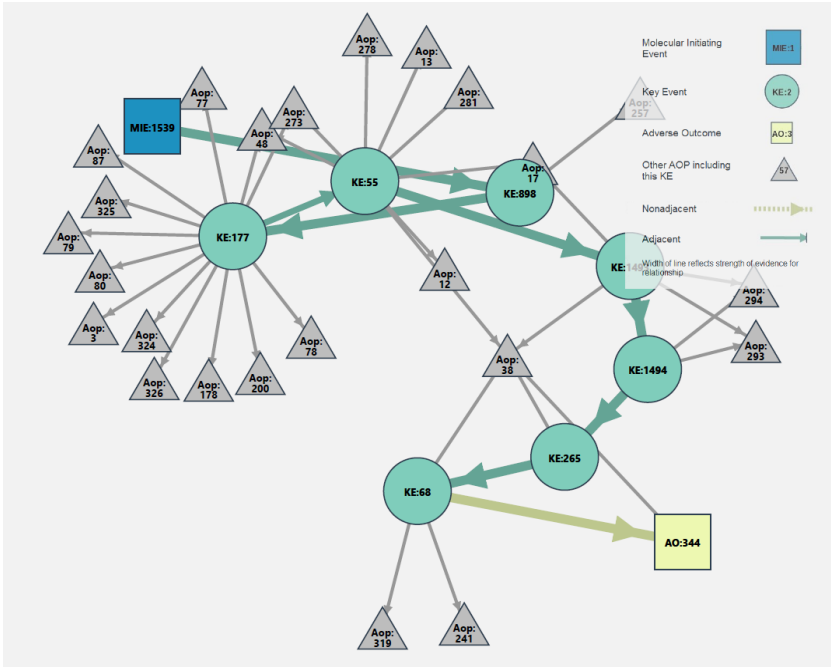
PTGS-derived AOPs for steatosis (Compound 18)



Adapted from Mellor et al. CRITICAL REVIEWS IN TOXICOLOGY, 2016 VOL. 46, NO. 2, 138–152

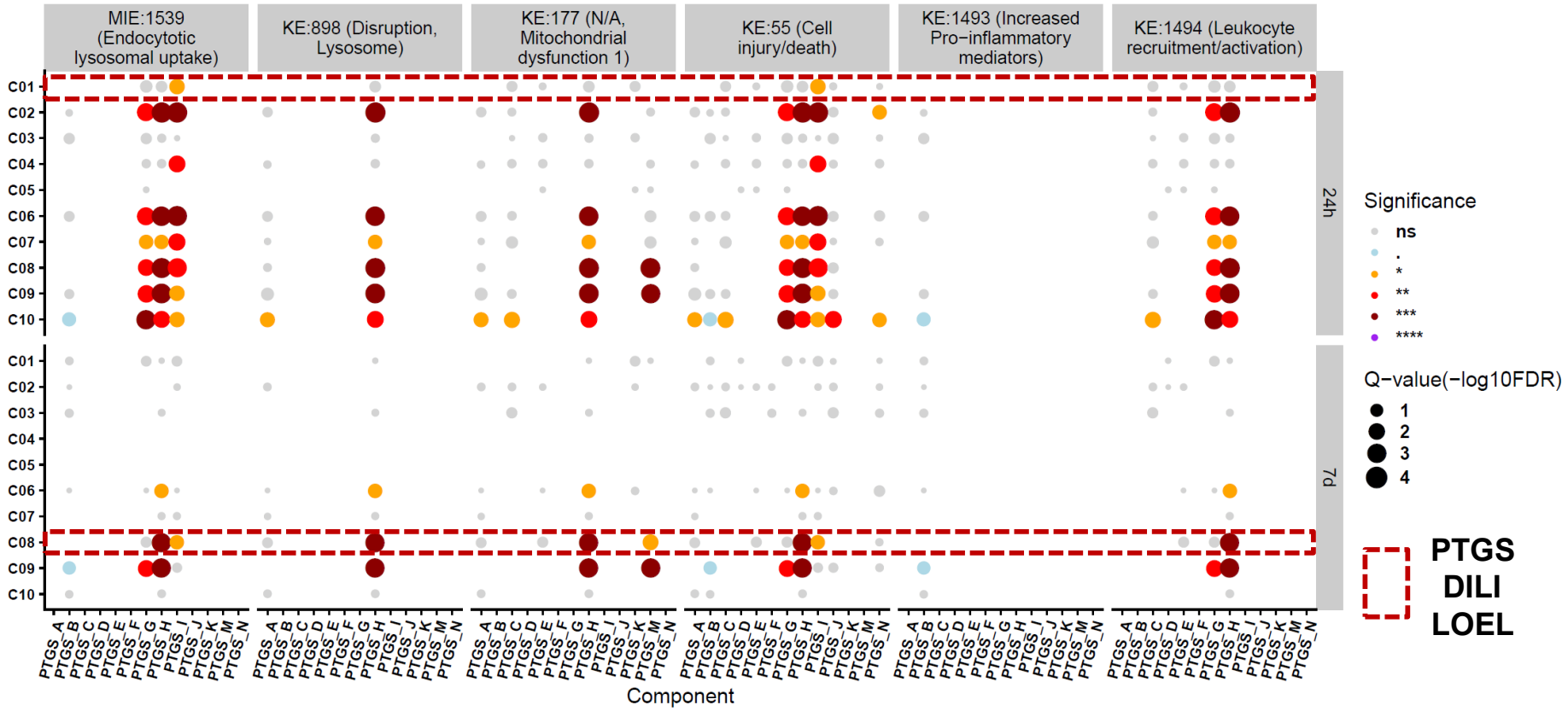
AOP 144: Endocytic lysosomal uptake leading to liver fibrosis (Compound 18)

Type	Event ID	Title	PTGS annotated
MIE	1539	Endocytotic lysosomal uptake	X
KE	898	Disruption, Lysosome	X
KE	177	N/A, Mitochondrial dysfunction 1	X
KE	55	Cell injury/death	X
KE	1493	Increased Pro-inflammatory mediators	X
KE	1494	Leukocyte recruitment/activation	X
KE	265	Activation, Stellate cells	
KE	68	Accumulation, Collagen	
AO	344	N/A, Liver fibrosis	X



Example compound 18. Aop:144 (Events=9), Endocytic lysosomal uptake leading to liver fibrosis (PTGS DILI LOEL 24h: C01, 7d: C08)

(Quantitative estimates via multiple testing corrected p-values)



PTGS Component MoA analysis using competitive GSEA (the 14 toxicity-associated vs the non-toxicity 86 components of the LDA-compressed cMap). Five of six PTGS-annotated events positive within the concentration series

The PTGS tool- inherent capabilities ensure broad industrial and regulatory applicability

- Serves as a giant AOP-applicable toxicity biomarker that captures/describes dose response and MoA
- Can be applied in high throughput manner to diverse types of model systems and types of transcriptomics data: microarray, RNA seq, feature sets (EPA/NTP S1500+; LINCS L1000)
- Initial selection of toxicity-related genes serves to cover multiple toxicity pathways and to avoid unspecific gene expression noise
- Analysis is standardized, driven by data completeness and quality considerations, arbitrary differential gene expression cut-offs are eliminated, and all gene expression levels are taken into account (FAIR principle considered in data handling)
- The algorithm and bioinformatics processing concept applying GSEA outperforms common tests for analyzing cytotoxicity; PTGS test is commonly 1-2 order of magnitude more sensitive
- Scoring concept enables IVIVE both with or without PBPK data/modeling results and clinical data (e.g., Cmax values)

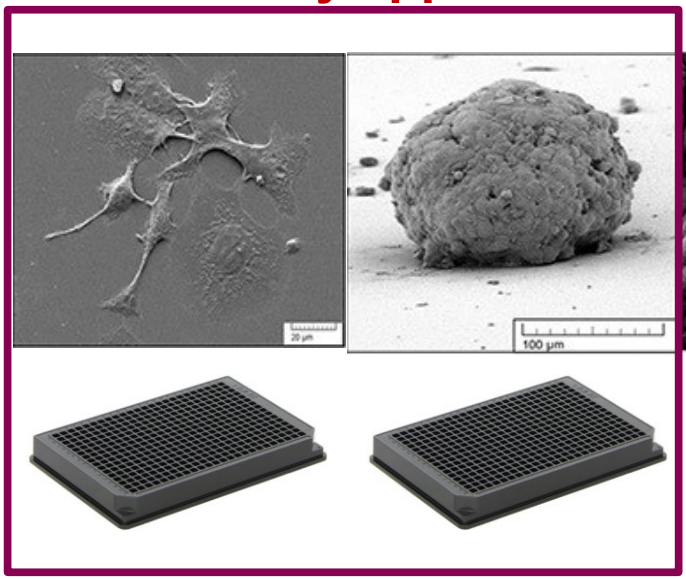
Human Cellular and Tissue Experimental Models

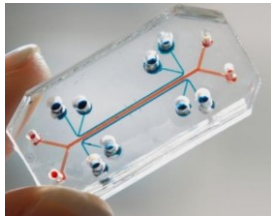
Level of human In Vitro Biomimetic/Structure Function

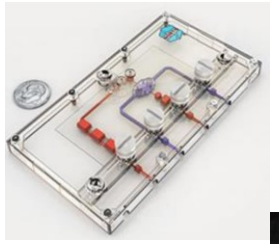
Recommended key concept: be as simple as you can but as complex as needed!
PTGS-driven DILI prediction is so far most accurate with 24h spheroid exposures!

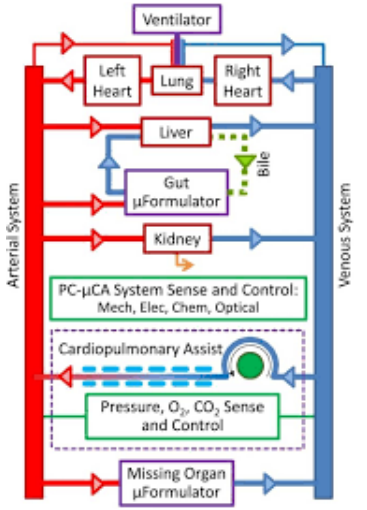
Static 2D culture	Static 3D spheroids	Static Organoid MPS	Organoids in Fluidic MPS	Biomimetic, Fluidic MPS	Integrated, Fluidic Organ MPS

Successfully applied to the HT analyses!



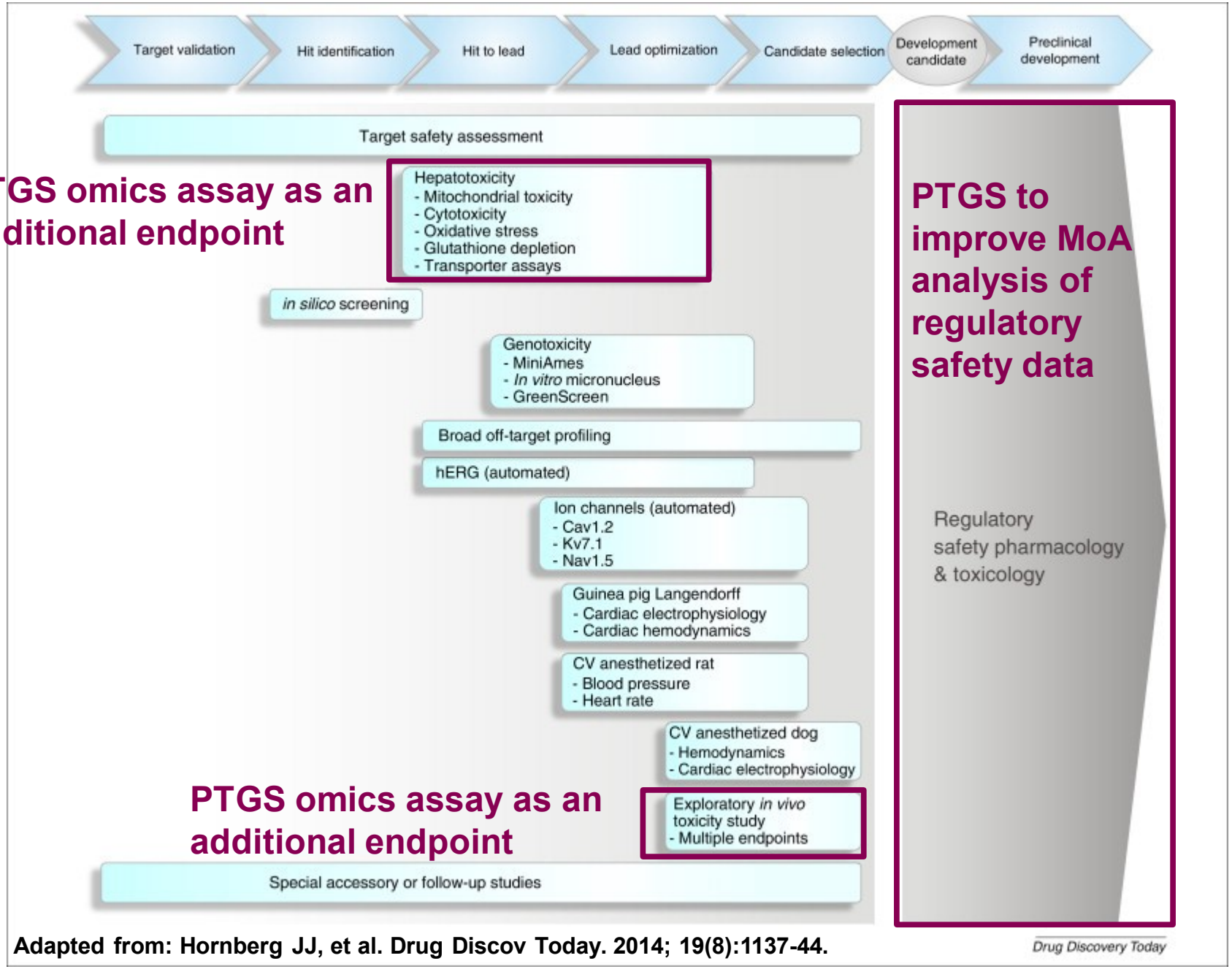






MPS = MicroPhysiological Systems

Applying PTGS scoring in drug discovery



PTGS omics assay as an additional endpoint

PTGS omics assay as an additional endpoint

PTGS to improve MoA analysis of regulatory safety data

Adapted from: Hornberg JJ, et al. Drug Discov Today. 2014; 19(8):1137-44.

PTGS concept/applications-state Nov 2021 (Predictomics & Karolinska Innovations)

Defined toxic MoA/Adverse Outcomes/AOPs (genes, gene sets, pathways, networks, components, pertubation classes) for 2533 agents (877 693 304 data points; 49% results data)

AI, ”Big Data”, sequential machine learning-driven drug side effects prediction (unique algorithms for cells/organs)

Broad dose-response coverage, including below overt toxicity and pathway pertubations

Drug development and repurposing based on defined gene targets, MoA and connectivity mapping (use of implicated/established ”opposing” drugs, gene constructs, etc. reflecting agonist or antagonist influences)

Broad potential pharmacovigilance applicability where DILI prediction is the primary proof-of-concept (accuracy is currently higher for sensitivity than for specificity)