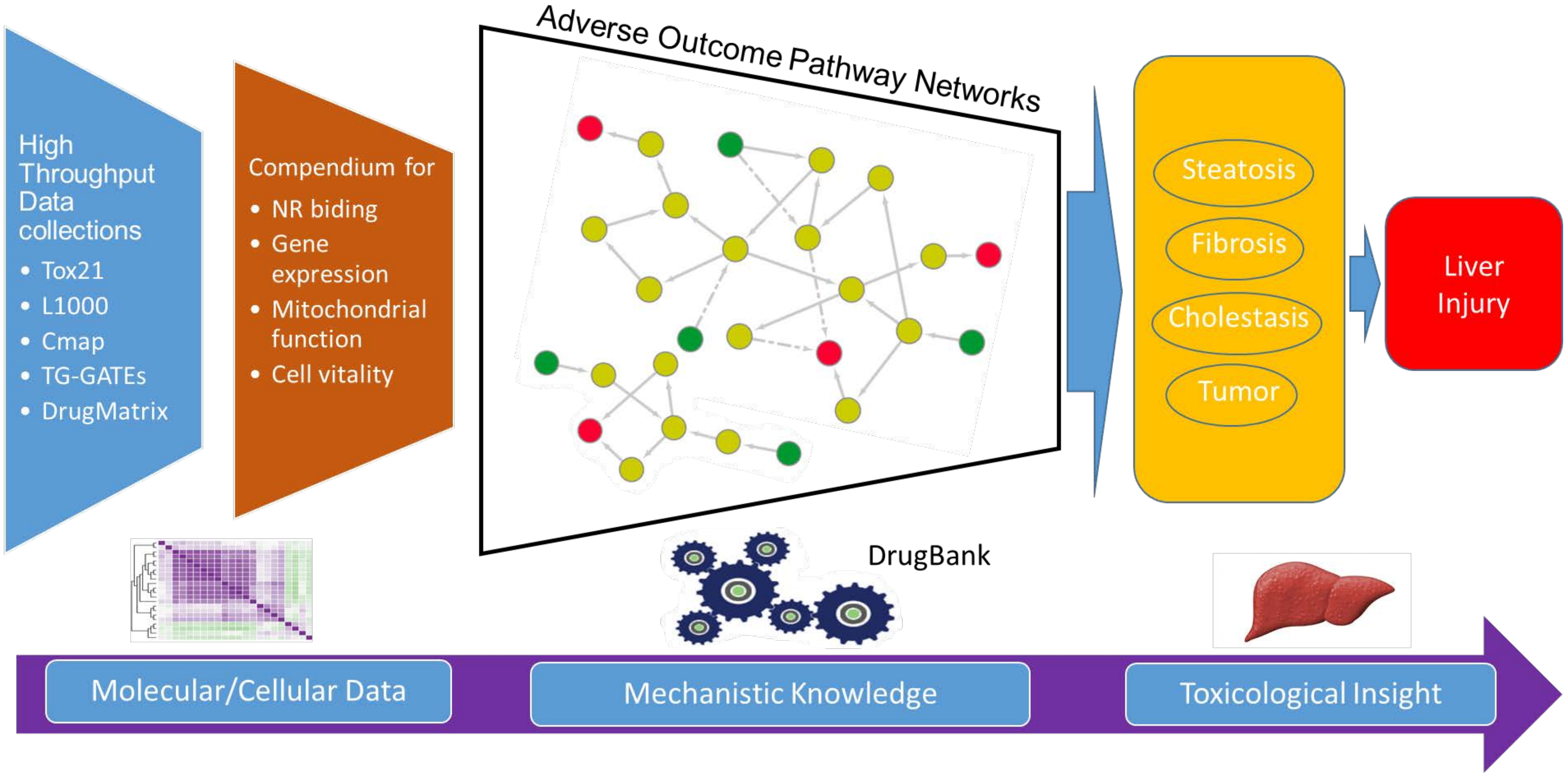
U.S. FOOD & DRUG ADMINISTRATION

Adverse Outcome Pathway Network Guided High-Dimensional Modeling for Drug Induced Liver Injury and Real-world Data

Dong Wang
Division of Bioinformatics and Biostatistics
National Center for Toxicological Research
US Food and Drug Administration

- *The opinion expressed in this presentation is of the author. It does not necessarily reflect the position of the US Food and Drug Administration.*

AOP guided modeling for drug induced liver injury (DILI)

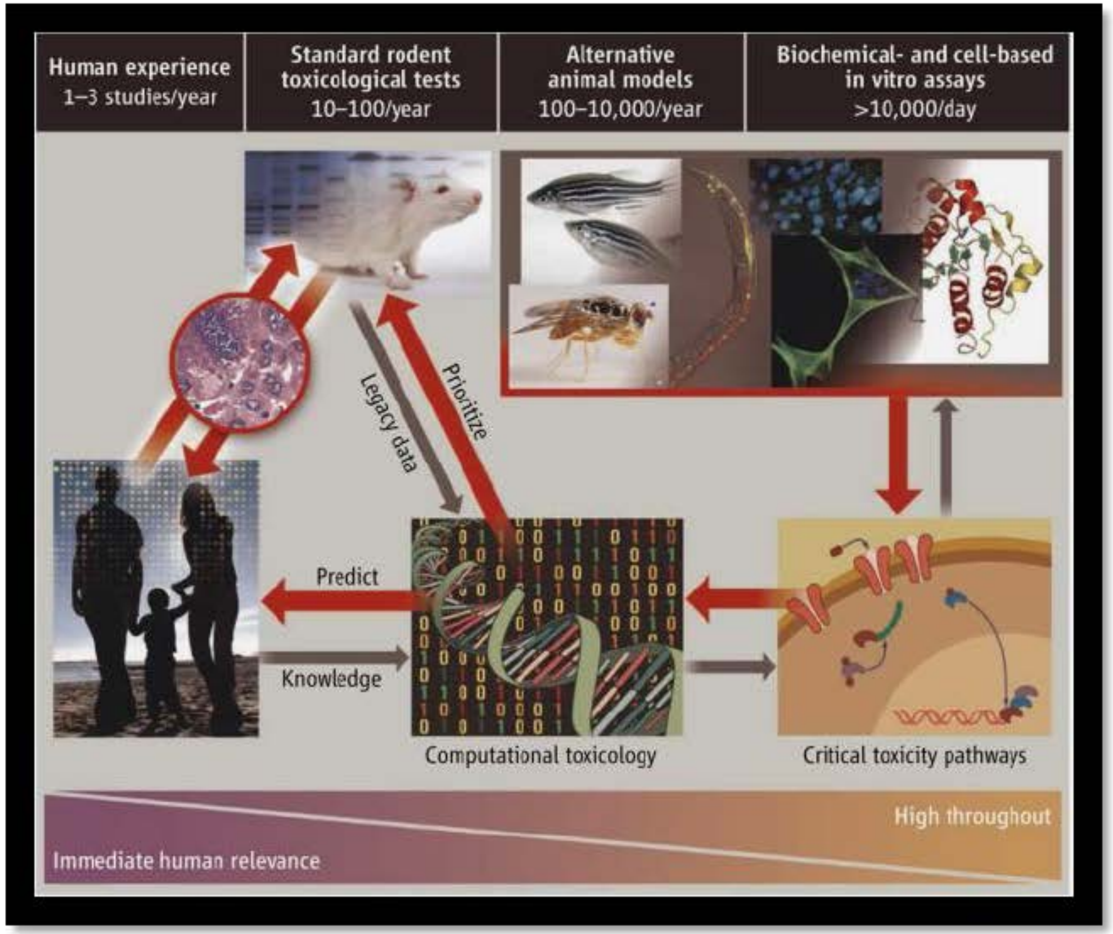


Integrating mechanistic knowledge with high throughput assays, a case study with DILI

- **Drug induced liver injury (DILI) is an important cause for drug withdraws**
- **Prediction of DILI rely heavily on the use of animals: often not accurate and also expensive**

- **Replacement with *in silico* models, and in vitro tests on high priority list of 21st century toxicity assessment**
  - ❖Faster/cheaper, broader, deeper, less animal use
- However, directly use high throughput data (Tox21, ToxCast, L1000, CMap, etc.) is difficult because the number of assays far exceeds the number of drugs, extremely high dimensional relative to sample size.
- Potential solution: use mechanistic knowledge to guide the modeling of high throughput data
- AOP networks provide a natural way to describe mechanistic knowledge of adverse events.

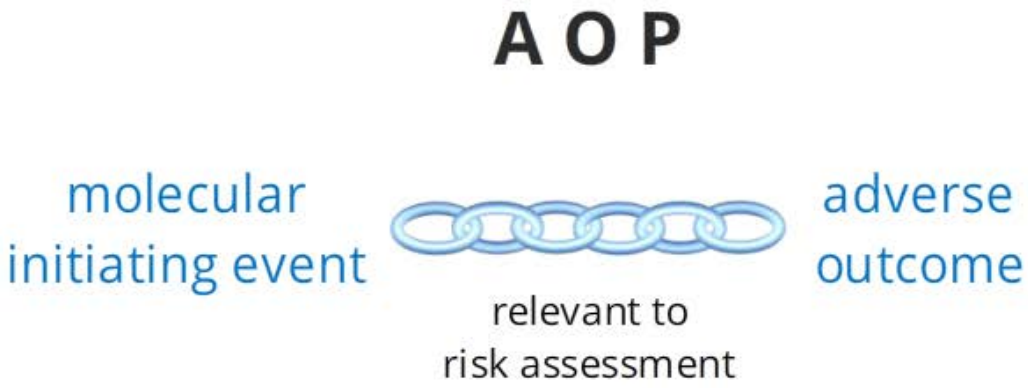
Tox21 Vision:
Transforming Toxicity Testing



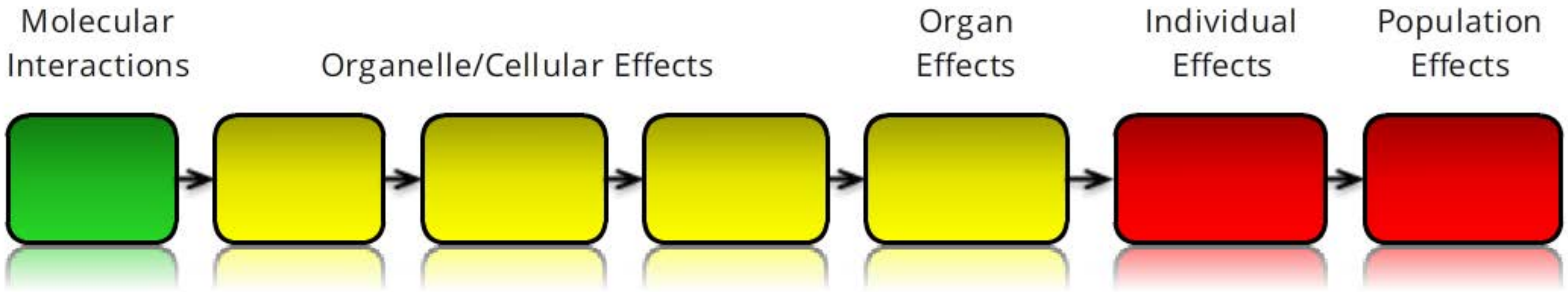
National Center for Advancing Translational Sciences (NCATS)
http://www.ncats.nih.gov/

SOURCE: Collins, Gray and Bucher (2008) Toxicology. Transforming environmental health protection. Science 319: 906

AOP Formally Defined in the Context of Risk Assessment



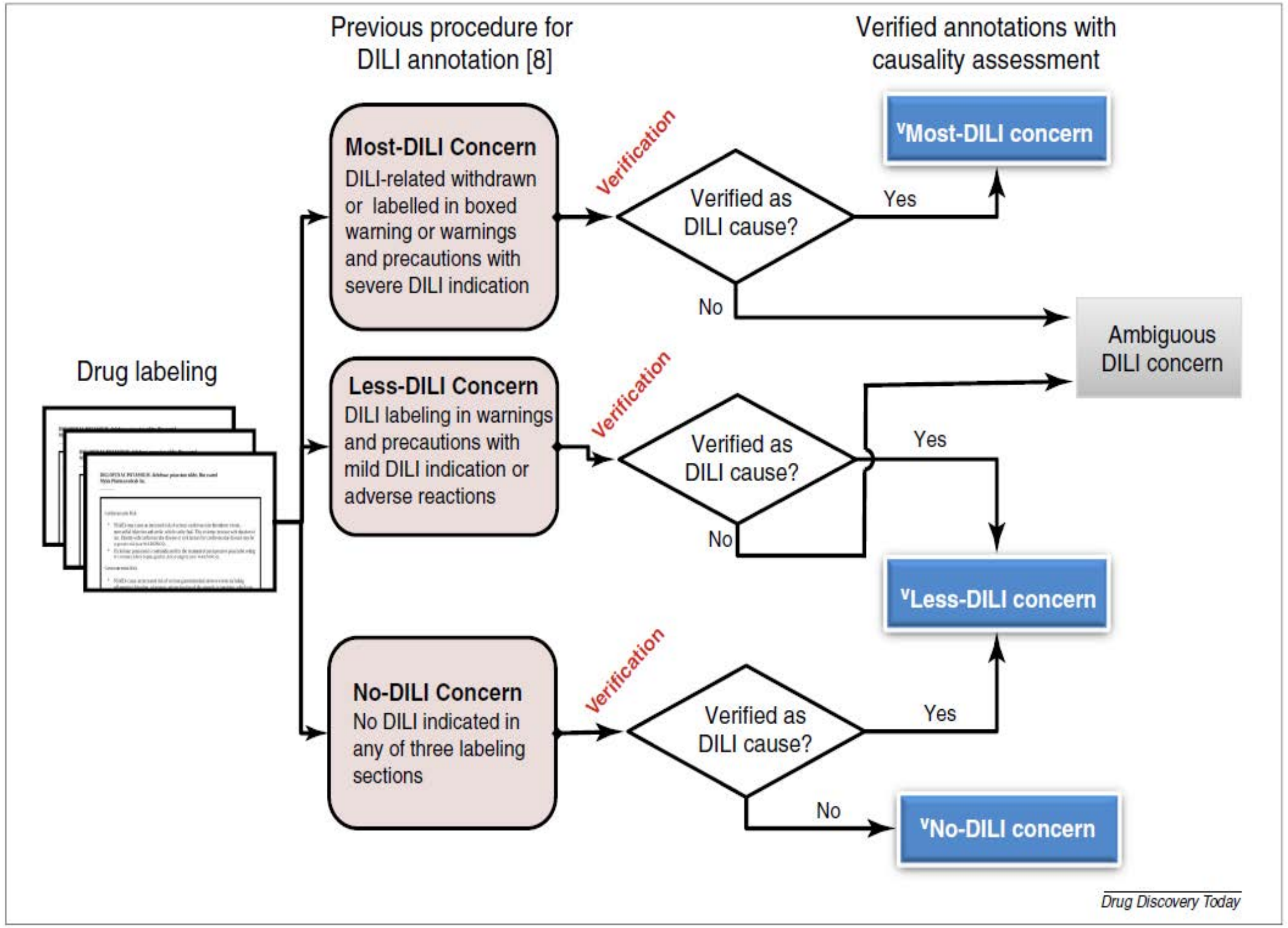
- Adverse Outcome Pathways (AOP) is an analytical construct that describes a sequential chain of causally linked events at different levels of biological organisation that lead to an adverse health or ecotoxicological effect (see figure below). AOPs are the central element of a toxicological knowledge framework being built to support chemical risk assessment based on mechanistic reasoning.



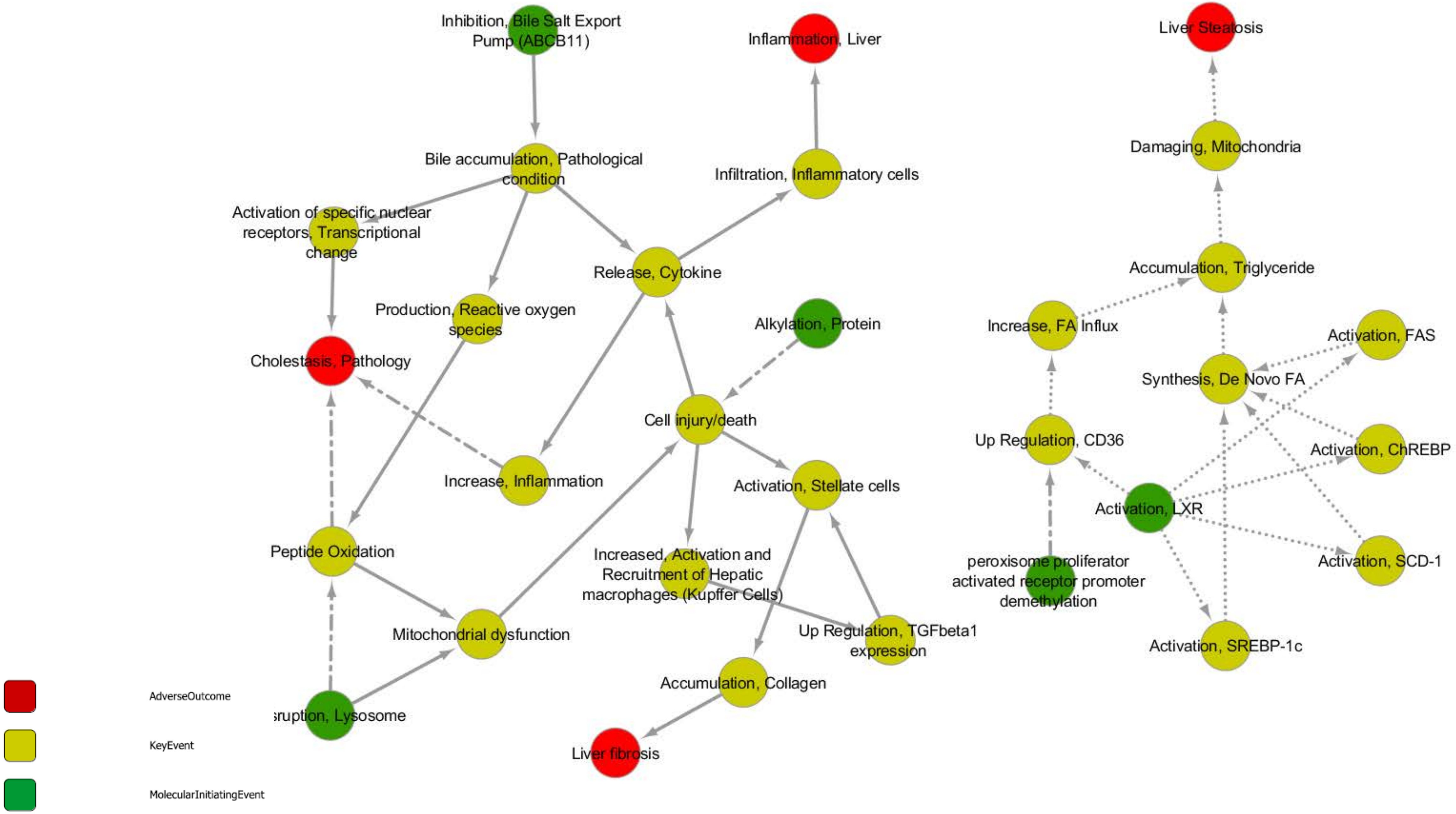
DILI AOP Network
Map Drugs to AOPs using information in DrugDB and AOPwiki

- AOP pathways form networks form pathways due to shared elements
- Currently 270 AOPs in AOPwiki
- Are they useful for modeling DILI?

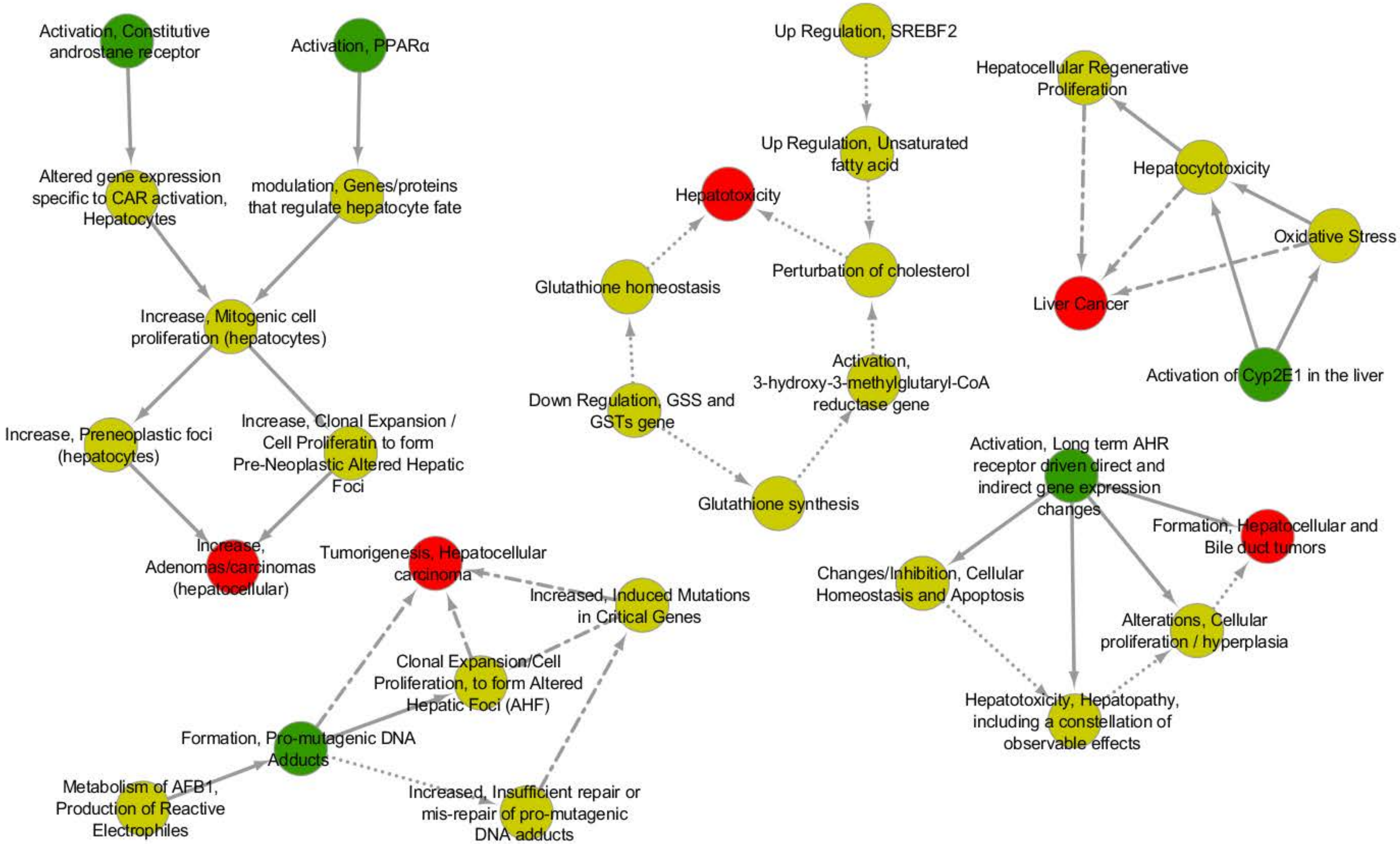
- DILI concern data on 364 drugs (Chen et al. 2016) from Liver Toxicity Knowledge Base
- Drug information (targets, enzymes) taken from drugbank (https://www.drugbank.ca/)
- Drug information used to map AOPs from AOPwiki (aopwiki.org)
- 27 unique AOPs found for the drugs investigated



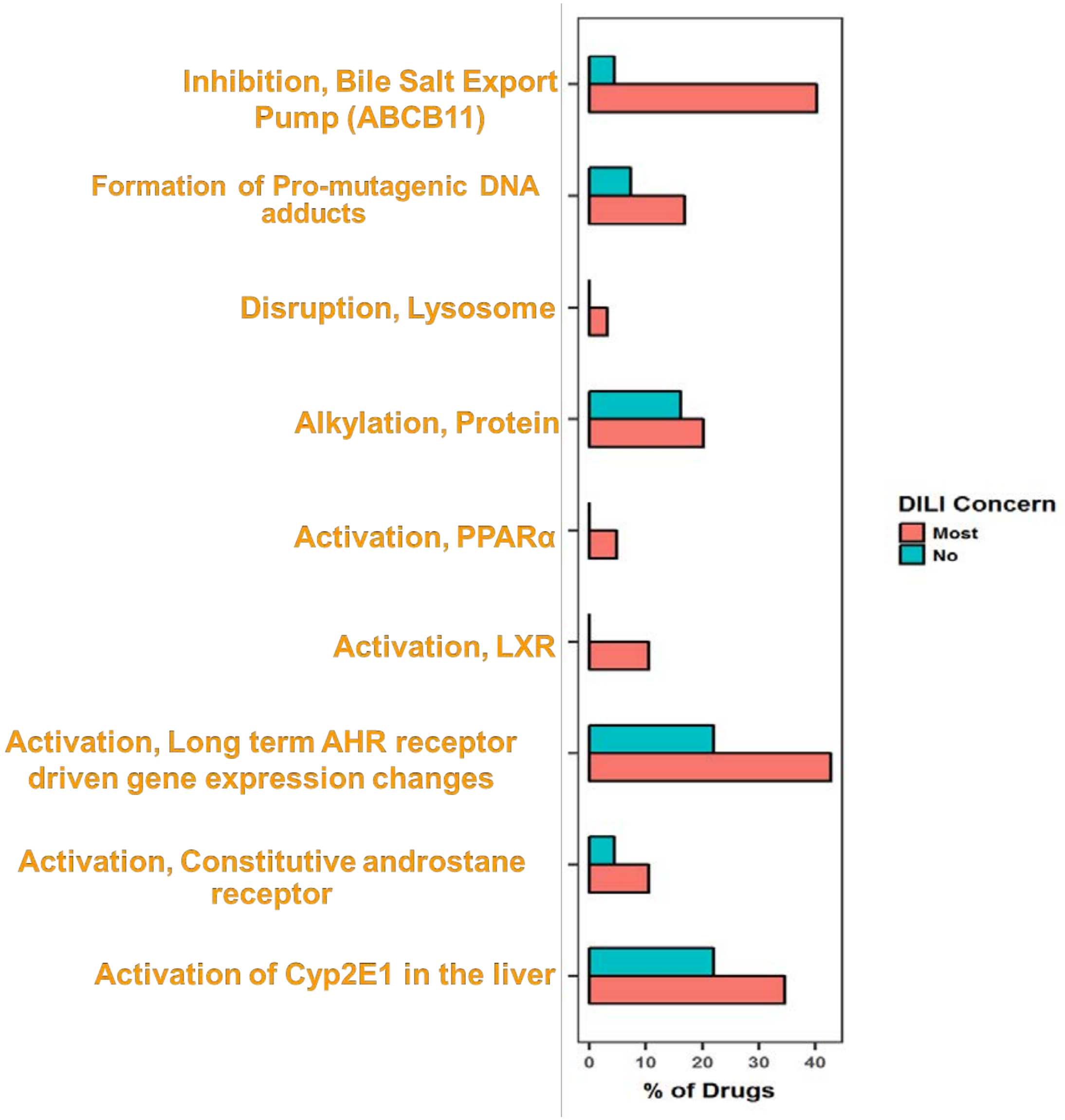
DILI Network



DILI networks II



Drug-AOP combinations for most-DILI-concern and no-DILI-concern drugs



- Adverse outcome pathways are hit with different frequencies by most-DILI-concern and no-DILI-concern drugs
- Some AOPs are more discriminative than the others
- This confirms the utility of AOP networks as knowledge source for adverse effects of drugs, even though the majority of AOPs are still under review by OECD

Building AOP guided predictive models for DILI

- Extract AOPs related to liver toxicity: steatosis, cholestasis, fibrosis, and liver tumor.
- Use AOPwiki and some publications
- From these AOPs, extract relevant nuclear receptors, cell function indicators, and genes as potential DILI predictors
- For drugs in Chen et al. (2016), extract the nuclear receptor binding properties and mitochondrial function measure from Tox21, and extract gene expression data from L1000.
- Use these high throughput predictors to build a model to predict DILI potential.

High throughput predictors extracted from AOP networks

Name	Type	AOP network	Source
CD36	Gene	steatosis	L1000
ChREBP	Gene	steatosis	L1000
SREBP-1c	Gene	steatosis	L1000
FAS	Gene	steatosis	L1000
SCD-1	Gene	steatosis	L1000
HSD17B10	gene	steatosis	L1000
PCK1	gene	steatosis	L1000
LDLR	gene	steatosis	L1000
CYP1A1	gene	steatosis	L1000
ACC-1	gene	steatosis	L1000
FOXA2	gene	steatosis	L1000
CPT1A	gene	steatosis	L1000
HMGCS2	gene	steatosis	L1000
HSD17B4	gene	steatosis	L1000
MTTP	gene	steatosis	L1000
ApoB100	gene	steatosis	L1000
NTCP	gene	cholestatiss	L1000
OATP1B1	gene	cholestatiss	L1000
CYP7A1	gene	cholestatiss	L1000
OSTA	gene	cholestatiss	L1000
OSTB	gene	cholestatiss	L1000
MRP2	gene	cholestatiss	L1000
UGT2B4	gene	cholestatiss	L1000
CYP3A4	gene	cholestatiss	L1000
SULT2A1	gene	cholestatiss	L1000

Name	Type	AOP network	Source
MRP3	gene	cholestatiss	L1000
CYP2E1	gene	tumor	L1000
apoptosis	general	fibrosis	
PPARgamma	NR	steatosis	TOX21
PPARalpha	NR	steatosis, tumor	
PPARbeta	NR	steatosis	
AHR	NR	steatosis, tumor	TOX21
NR1I3 (CAR)	NR	steatosis, cholestasis	
HNF4A	NR	steatosis	
PXR	NR	steatosis, cholestasis	
NRF2	NR	steatosis	TOX21
FXR	NR	steatosis, cholestasis	TOX21
SHP	NR	steatosis, cholestasis	
LXRalpha	NR/Gene	steatosis	
LXRbeta	NR/Gene	steatosis	
PPARgamma	NR/Gene	steatosis	L1000
AKT2	protein	steatosis	L1000
ATAD5	protein	fibrosis	Tox21, L1000
ATF6	protein	fibrosis	Tox21 , L1000
ESR1	NR	steatosis	Tox21
NR3C1	NR	steatosis	Tox21
MMP	mitochondria	fibrosis	Tox21
NFkB	NR	fibrosis	Tox21
RARA	NR	steatosis	Tox21
RXRA	NR	Steatosis	Tox21

Data Processing

- A total of 44 predictors were used to build a model to predict DILI potential.
  - Three drug properties used by Chen et al. (2016): daily dose, logP, RM formation
  - 10 assay endpoints from Tox21 assays on nuclear receptor binding, mitochondrial function, and apoptosis reporters
  - 31 indicators for differential gene expression from L1000
- Preprocessing
  - For Tox21 assays, a standardized score was obtained by dividing the maximum median response by the baseline median absolute deviation
  - For L1000 data:
    - Experiments using HepG2 were used if available. Otherwise, data from MCF7 experiments are used.
    - If multiple experiments are available for the same cell line, the experiment designated as exemplar by the L1000 team is used. Otherwise, the one with the highest transcription activity score is used.
    - The L1000 z-score (level 5 analysis) is used as measurement of disturbance of expression by the drug.
- Drugs: we were able to collect data on 287 drugs: MostDILIConcern 92, LessDILIConcern 141, No DILIConcern 54.

Predictive model for DILI potential using predictors guided by mechanistic knowledge

- Use data from most DILI concern and no DILI concern drugs to fit a sparse logistic regression model with elastic net penalty with 44 predictors.

- $$log\frac{P(y=1|X=x)}{P(y=0|X=x)} = \beta_0 + \beta^T x$$

$$(\hat{\beta}_0, \hat{\beta}) = \underset{(\beta_0,\beta)}{\text{argmin}} - \frac{1}{N}\sum_{i=1}^{N}\left[y_i(\beta_0 + x_i^T\beta) - \log\left(1 + e^{\beta_0 + x_i^T\beta}\right)\right] + \lambda\left[\frac{(1-\alpha)\|\beta\|_2^2}{2} + \alpha\|\beta\|_1\right].$$

- Model fitting:

  The parameter α is set to 0.9. The penalty parameter λ is determined by cross-validation. Model fitting is carried out in R.

Sparse logistic regression model

- Model fitting and cross validation.
- Model Fit

		Actual	
		No Concern	Most Concern
Predicted	No Concern	45	4
	Most Concern	9	88

Cross-validation

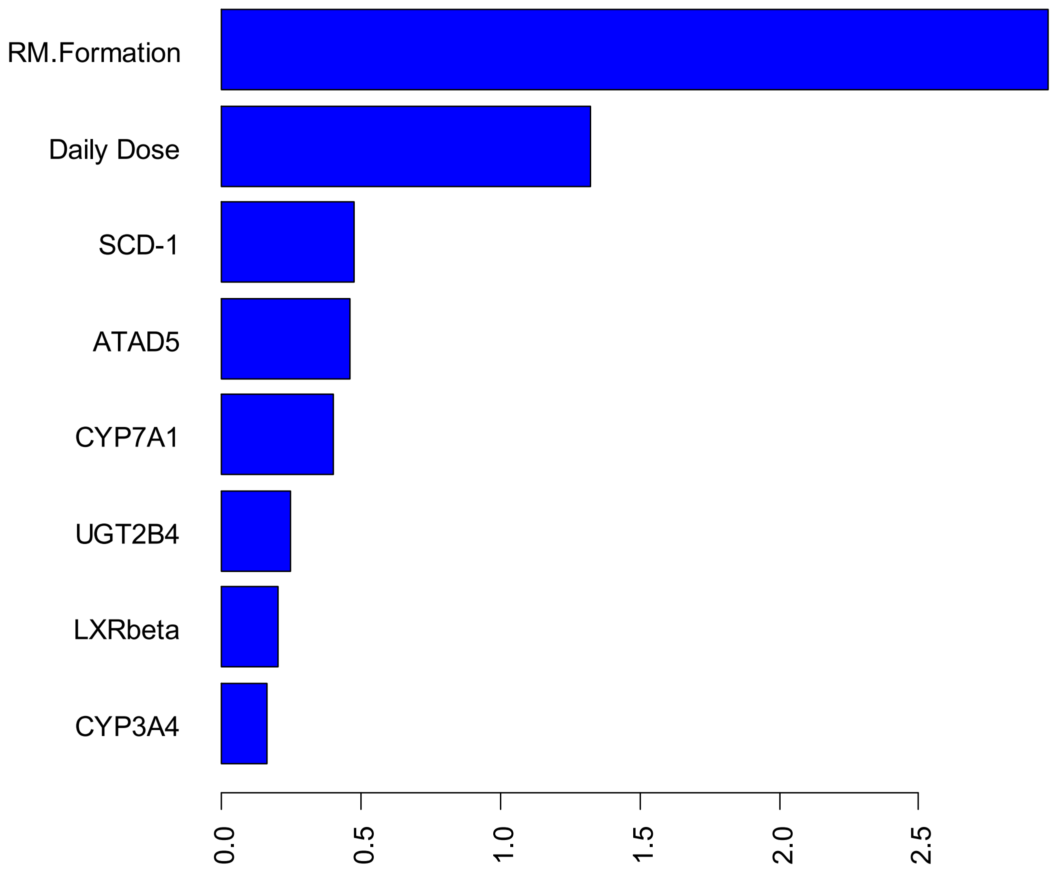
		Actual	
		No Concern	Most Concern
Predicted	No Concern	39	11
	Most Concern	15	81

- Model fit: sensitivity 0.96, specificity 0.83
- Cross-validation: sensitivity 0.88, specificity 0.72

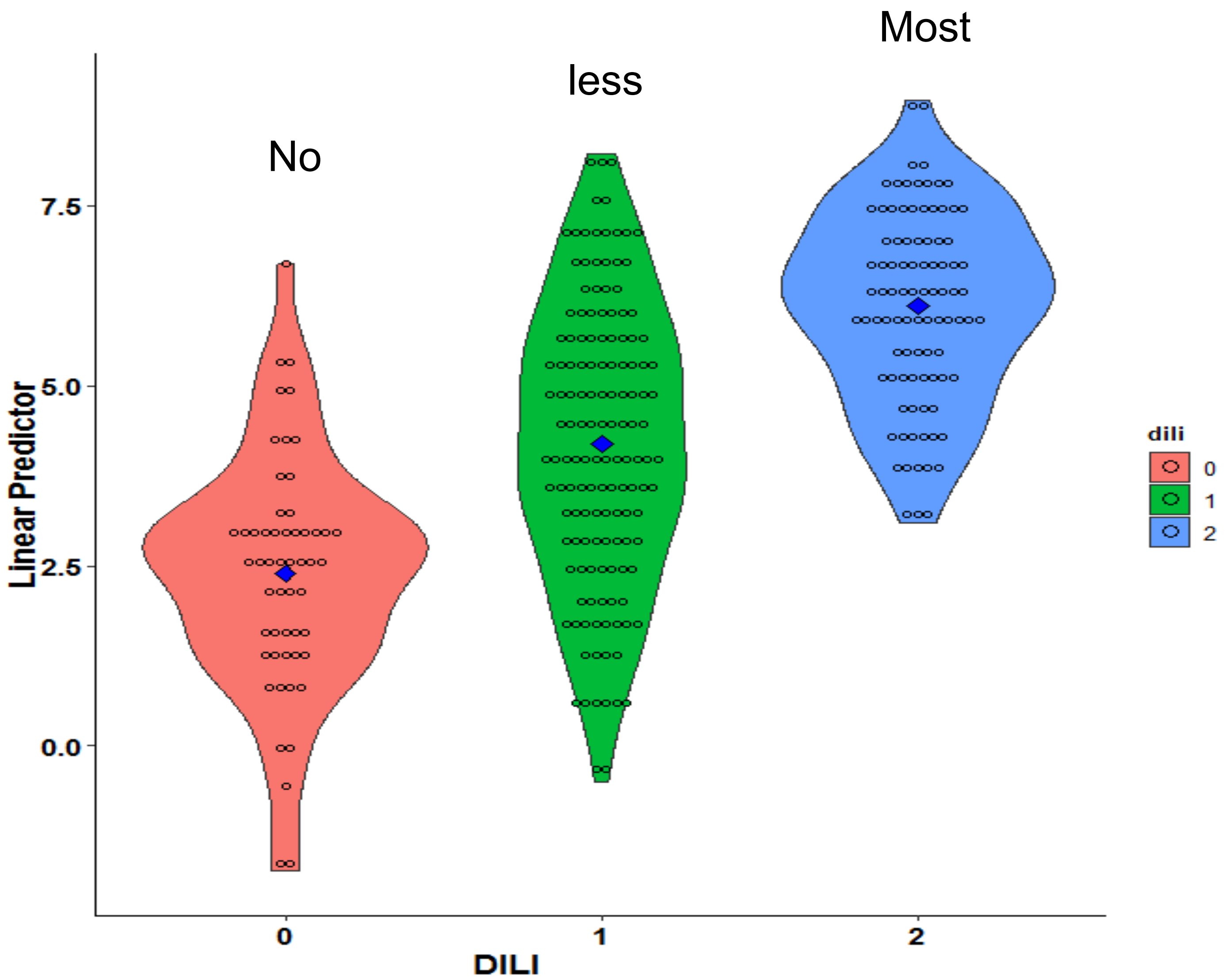
- Nearly random results without AOP filtering

Important predictors for DILI

- In this case, 19 predictors have nonzero coefficient
- As usual for logistic regression, the linear combination of these 19 assays form the linear predictor for the probability of DILI on the logistic scale.
- The two predictors with the largest coefficients are Reactive Metabolite formation and daily dose.
- These are followed by some important enzymes and receptors
- Future path: develop specified assays for RM formation and molecular reporters

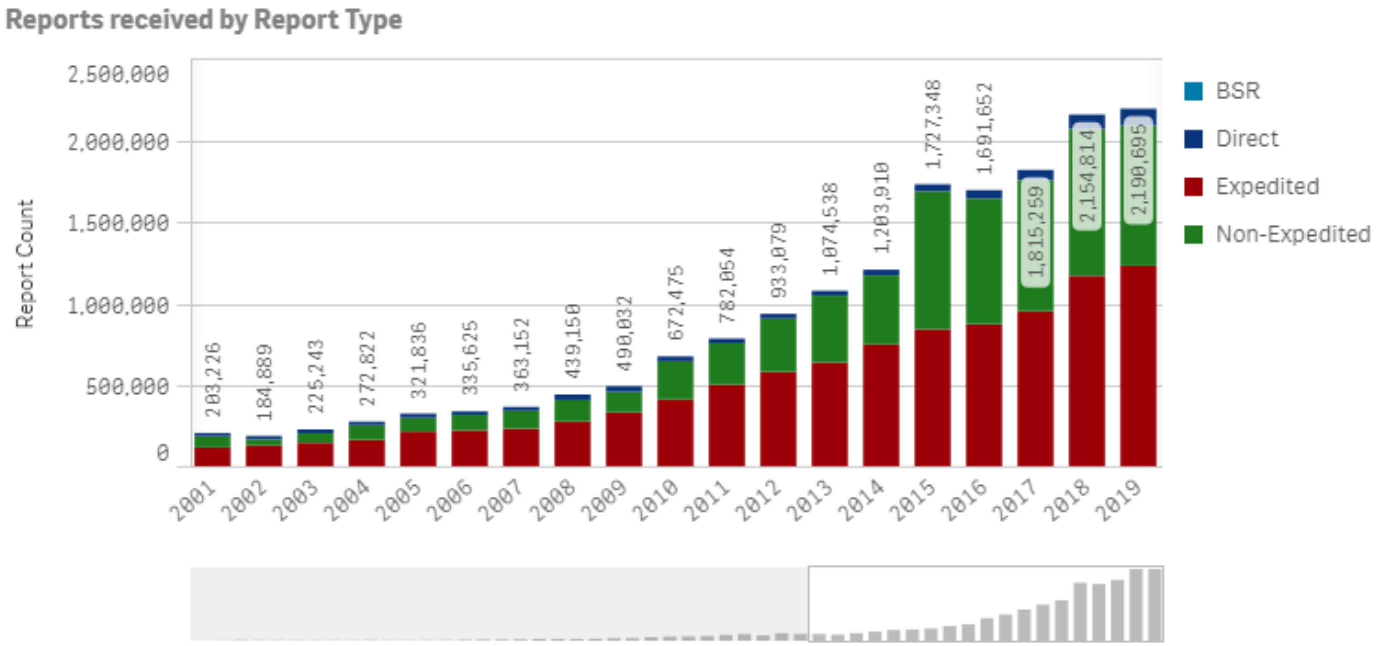


The linear predictor can be used as a score to rank drugs for DILI potential



Validation with Data from Postmarket Surveillance

- FDA Adverse Event Reporting System (FAERS) is a database that contains spontaneous adverse event reports that are submitted to FDA from the product manufacturer or directly from the consumer, healthcare professional, or other reporter. The database supports the FDA's post marketing safety surveillance program.
- Data are released quarterly.
- The database consists of more than 19 million reports. It is added at a rate of over two million adverse events per year.
- The report by itself does not establish causation.



Signal generation: one way around the no-denominator problem, use expected count

2x2 Table for the combination of adverse event AE i and Drug j

	AE i	no AE i	total
drug j	n_{ij}	$n_{.j}$-n_{ij}	$n_{.j}$
no drug j	$n_{i.}$-n_{ij}	$n_{..}$-$n_{.j} - n_{i.} + n_{ij}$	$n_{..}$-$n_{.j}$
total	$n_{i.}$	$n_{..}$-$n_{i.}$	$n_{..}$

- If drug and AE distributions are independent, then n_{ij} should be close to the baseline frequency
- $$E_{ij} = \frac{n_{i.} n_{.j}}{n_{..}}$$
- Declare a signal if n_{ij} is much higher than E_{ij} , usually based on a Poisson model
- It may be then investigated whether drug j tends to cause event i
- Important methods include empirical Bayes multiitem gamma Poisson shrinker (MGPS, Dumouchel 1999), likelihood ratio test (Huang, Zalkikar, Tiwari 2011, FDA), proportional reporting ratio (Evans et al. 2001), logistic regression LTR (Nam et al., 2017, FDA), and others.

Incorporating Covariate for Poisson Model with FAERS

- Available methods focus on signal detection
- We want to test for covariate (e.g., our linear score from in vitro assays)
- Test based on LTR:
- Model drug-AE combination counts to follow a $Poisson(\mu_{ij})$ distribution, where $\mu_{ij} = \lambda_{ij} E_{ij}$, $log(\lambda) = \beta_0 + \beta_1 s$.
- Here ***s*** is the linear predictor from our elastic net logistic regression.
- This is a Poisson regression model with drug specific offset.
- Modified LRT, needs Monte Carlo simulation for calibration

Corroboration of the prediction model with FAERS data

- The linear predictor *s* is significantly associated with the reporting rate of DILI in FAERS ($p < 10^{-6}$).
- The coefficient $\beta_1 = 0.26, e^{\beta_1} = 1.30$
- This means that if our linear predictor ***s*** increase by 1, the reporting counts will increase 30% from baseline level on average.
- This provides strong confirmation of our AOP based prediction model
- This model is also applicable to other covariates other than the linear score (e.g., host factors).

Integrating Adverse Outcome Pathways (AOPs) and High Throughput In Vitro Assays for Better Risk Evaluations, a Study with Drug-Induced Liver Injury (DILI)

Khadka KK, Chen M, Liu Z, Tong W, Wang D. ***ALTEX***.

https://doi.org/10.14573/altex.1908151

Society of Toxicology (2020) Computational Toxicology Specialty Section SS Pfizer Paper of the Year Award

Conclusion

- AOP is an efficient approach to organize mechanistic knowledge
  - Drug properties (e.g., DrugDB entries) can be mapped to AOPs to discover potential adverse effect
  - Only useful when the drug properties are properly annotated.
  - Multiple drugs to multiple AOPs
- Use AOP networks to guide the modeling of high throughput data
  - Meaningful predictors can be extracted from AOP networks
  - Information for these predictors can then be obtained from high throughput experiments
  - Effective dimension reduction, good model performance
- Potential use
  - The current example is far from ideal: some predictors indicated from AOP are not available; information from L1000 and Tox21 are under different doses, time, cell lines; not specifically designed for this purpose
  - Good model performance points to potential to design test panels
- Utilize postmarket surveillance data to assist model development

Acknowledgement

- Kapil Khadka
- Minjun Chen
- Zhichao Liu
- Weida Tong
- Qais Hatim
- Jane Bai
- Ayoko Suzuki

Thank you!

FDA U.S. FOOD & DRUG ADMINISTRATION