A1250 / P410

In Silico Assessment of Drug-Induced Liver Injury in Humans

V. Gombar[1], S. Thakkar[2], J.Rathman[1], A. Mostrag[1], W. Tong[2], and C. Yang[3]

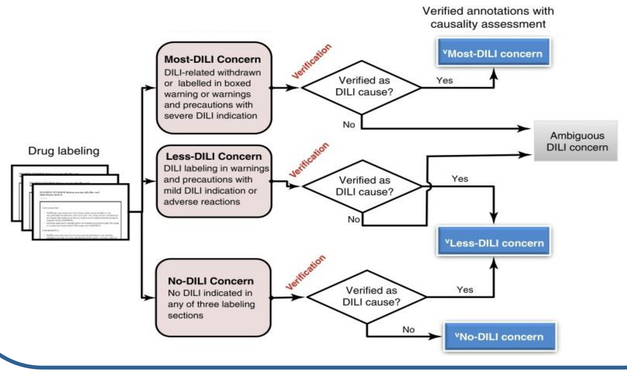
[1]Altamira LLC, Columbus, OH; [2]US FDA/NCTR, Jefferson, AR; [3]Molecular Networks GmbH, Nürnberg, Germany

Background

- **Drug-induced liver injury (DILI) is an adverse effect whose reliable assessment falls short by current *in vivo* toxicological studies[1-3] during drug discovery and development.**
- **1036 drugs administered to humans have been classified in terms of their DILI risk based on FDA drug labels, called DILIRank, after verification for causality[1,4]**
- **QSAR and Decision Forest (DF) models have been developed for predicting the potential of chemical compounds to cause DILI in humans by oral route of administration.[6]**
- **We present a novel *in silico* system developed for prediction of DILI risk, in which evidence from multiple QSAR (structural features and physical/chemical properties-based) and DF models are rigorously and quantitatively combined to arrive at a weight-of-evidence outcome with associated estimation of uncertainty.**

Data Curation

DILI classification based on FDA-approved drug labeling[1]



Training and test sets

Chemical structures QC

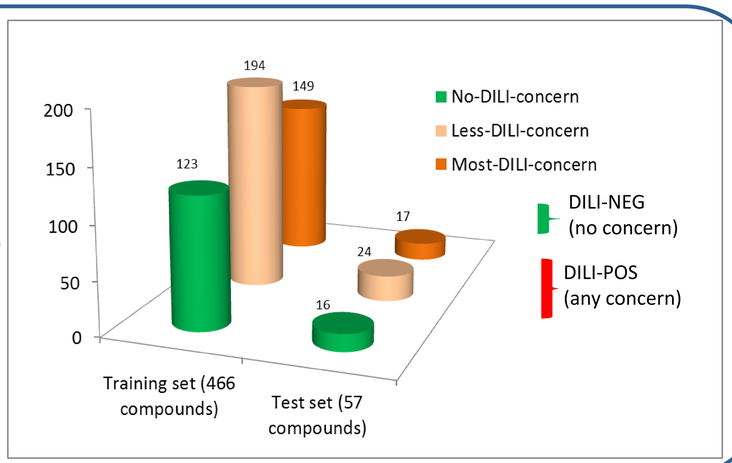
- Initial dataset: 1036 drugs with annotated DILI risk. This study focused on DILI risk to humans by oral route of administration.
- Verification of: atom/bond features, valence, radical states, charges, multiple fragments, stereochemistry and double bond geometry

Structures selection

- Removed IOMs (inorganic, organometallic, metal complexes and metals), natural products, mixtures, polymers, etc.

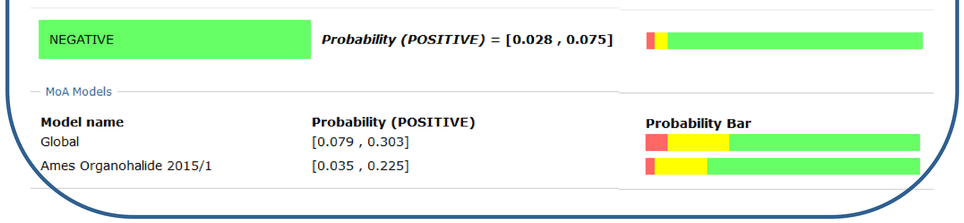
Computational structures

- CORINA CLEAN workflow in Corina Symphony (Molecular Networks)[7]
- Processing options: remove small fragments, neutralize, generate 3-D structure, flag duplicates for user decision



Model development

- **QSAR model building and cross validations performed exclusively on the training set**
- QSAR (logistic PLS, partial least squares regression) models were built and optimized by *k*-fold cross-validation with ToxPrint chemotypes[8] and physical/chemical properties as descriptors.
- The full global training set (466) was partitioned into three subsets to balance the POS/NEG ratios. Two local models were also created: Alcohols and FusedRings.
- Decision forest was developed using the training set of DILI-POS is 149 drugs (Most -DILI-Concern) and DILI-NEG is 123 drugs (No-DILI-Concern). 5- fold cross validations was performed for 1000 iterations. For generating the model no seed was generated. Six minimum and sixteen maximum trees were generated.
- The positive (PPV) and negative (NPV) predictive values provide the reliability metrics for each model.
- WoE prediction are computed by applying Dempster-Shafer decision theory.



Model validation and performance

Model	N	% no concern	Concordance	Sensitivity	Specificity	PPV	NPV
Global 1	267	52%	0.71	0.69	0.72	0.69	0.72
Global 2	267	52%	0.70	0.67	0.72	0.69	0.70
Global 3	267	52%	0.73	0.70	0.75	0.72	0.73
Alcohol	166	37%	0.75	0.79	0.67	0.80	0.66
Fusedrings	39	41%	0.67	0.65	0.69	0.75	0.58

Properties: HOMO-LUMO gap, logS, # hydrogen acceptors, HOMO, MW, TPSA

ToxPrints: As expected, due to the high structural diversity of this data set, a large number (~150) of ToxPrints factor into the global models

Model	N	% NEG	Concordance	Sensitivity	Specificity	PPV	NPV
Decision Forest	272	45%	0.76	0.76	0.68	0.74	0.71

QSAR		Predicted NEG*	Predicted Equivocal	Predicted POS**	Quality	
Verified Call	NEG*	16	0	0	16/16	Specificity
	POS**	5	1	35	35/40	Sensitivity
Applicability		98.2%		91.9%	Concordance	

Decision Forest		Predicted NEG*	Predicted Equivocal	Predicted POS**	Quality	
Verified Call	NEG*	8	6	2	8/10	Specificity
	POS**	5	19	17	17/22	Sensitivity
Applicability		56.1%		78.1%	Concordance	

Combined WoE		Predicted NEG*	Predicted Equivocal	Predicted POS**	Quality	
Verified Call	NEG*	15	0	1	15/16	Specificity
	POS**	6	0	35	35/41	Sensitivity
Applicability		100%		87.7%	Concordance	

***NEG:** no DILI concern

****POS:** any (less or most) DILI concern

Conclusions

- A high-quality database of 1036 drugs with verified DILI concern for oral route of administration in humans has afforded development of *in silico* predictive models.
- Application of DST allows combining models developed by different approaches to make WoE predictions with reduced uncertainty (fewer equivocal predictions and higher applicability).
- Overall performance enhancements are marginal, which is consistent with the fact that the models were developed from the same knowledgebase.

References

[1]Chen, M. et al. (2016) Drug Discov Today
[2]Olson, H. et al. (2000) Regul Toxicol Pharmacol 32, 56-67
[3]Willyard, C. (2016) Nature Medicine, 22(5), 450-451
[4]Garcí´a-Corte´s, M. et al. (2011) J Hepatol 55, 683-691
[5]Mold2, Center for Bioinformatics, US FDA NCTR, Jefferson AR, USA, http://www.fda.gov/ScienceResearch/BioinformaticsTools/Mold2/default.htm
[6]Chen, M. et al. (2013) Toxicol Sci 136(1), 242-249
[7]Corina Symphony, Molecular Networks GmbH, Erlangen, Germany
[8]Yang, C. et al. (2015) J Chem Inf Model 55(3), 510-528, www.toxprint.org