Modeling and Predicting Cancer From ToxCast Phase I Data

Richard Judson, David Dix, Keith Houck, Thomas Knudsen, Imran Shah, David Reif, Matthew Martin, Holly Mortensen , Daniel Rotroff, Robert Kavlock

U.S EPA, ORD, Computational Toxicology Research Program

research & development

Science Question

The goal of the ToxCast program is to link *in vitro* assays with *in vivo* toxicity in humans. One strategy to approach this goal is to look for statistical associations between the ToxCast assays and *in vivo* phenotypes for animal models. The questions we address here are:

1. Can we find statistically significant associations between in vitro assays and cancer endpoints in rodents?
2. Can multi-assay models be constructed with high enough sensitivity and specificity to be useful for screening and prioritization?
3. Does our understanding of cancer biology (through knowledge of genes and pathways) support these models?

Research Goals

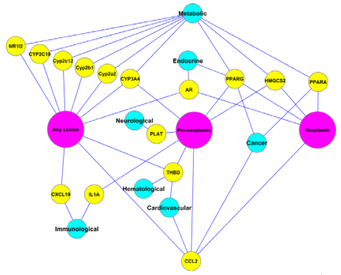
1. To develop a bioinformatics infrastructure that allows us to compile all ToxCast and ToxRefDB together in a way that allows high-throughput statistical analyses
2. To use statistical and machine learning algorithms to construct and test models linking *in vitro* assays with cancer endpoints in rats and mice
3. To qualify these models using information from the literature on general cancer biology and mode of action information on ToxCast chemicals. In particular we want to understand toxicity in terms of activation of specific genes and pathways.

Assays

A total of 687 in vitro assay endpoints were used from ToxCast. This includes multiple time points for several assays. All assays were run in concentration-response format and we extracted a characteristic concentration for each chemical-assay pair. These were either AC50 or LEC (lowest effective concentration, or the lowest concentration at which the response was significantly different than negative control). For these calculations, we did not use information on magnitude of effect (Max fold change or Emax). Assay AC50/LEC values were log transformed prior to performing statistical calculations. In addition to the *in vitro* assay data, we also used chemical class information, limited physico-chemical values calculated from chemical structure, and structure fragment composition fingerprints.

Endpoints

Cancer-related endpoints from rat and mouse chronic-cancer studies were extracted from ToxRefDB and dichotomized (cause / do not cause cancer)



Disease Progression

Univariate Correlations With Rat Liver Tumor Progression

Network of genes associated with the progression of rat liver tumor endpoints. Associations were calculated using Fisher's exact test, with assay AC50/LEC values ≤100 mM set to 1, and >100 mM set to 0. Only associations with a p-value <0.01 are included. Links between genes (yellow) and *in vivo* endpoints (pink) are shown where there is a statistical association based on the *in vitro* assay results. The "Any Lesion" category contains the "Pre-neoplastic" category, which in turn contains the "Neoplastic" lesions category. Disease or disorder classes (cyan) are linked to genes.

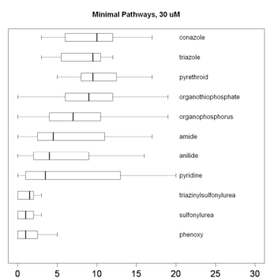
Endpoint	ATG BMPR2	ATG ESR1	CLZD HMGCS2	NVS HTR2C	ATG NFE2L2	ATG NR1I2	NVS NR1I2	NVS OPRL1	ATG PPARG	NVS PPARG	CLZD CYP1A1 (06)	CLZD CYP1A1 (24)	CLZD CYP1A1 (48)	CLZD CYP1A2 (24)	CLZD CYP1A2 (48)	NVS CYP3A4	CLZD CYP3A4 (48)	CLZD CYP3B6 (48)
Adrenal Gland Preneoplastic Lesion					0.32				0.38								0.36	
Bone Marrow Preneoplastic Lesion	3.72				9.73								3.40					
Kidney Preneoplastic Lesion				6.90														
Kidney Proliferative Lesions				6.64														
Liver Preneoplastic Lesion				2.65					2.84	3.79					1.55	3.13		
Liver Neoplastic Lesion			4.51						5.95									
Liver Proliferative Lesions	1.59		2.51						2.81	3.60						2.97		
Liver Tumors			4.53						6.00									
Mammary Gland Preneoplastic Lesion						4.49												
Pituitary Gland Preneoplastic Lesion									3.31		4.27	3.16		3.47	4.51			
Testes Preneoplastic Lesion															2.09			
Thyroid Preneoplastic Lesion		1.99				1.87		3.78				1.79				2.80		
Thyroid Neoplastic Lesion		2.43		2.73	4.49	2.14			2.31			2.26					2.14	2.25
Thyroid Proliferative Lesions		1.91						4.08								2.73		
Thyroid Tumors		2.29		2.41	3.04	2.03			2.10			2.12				2.08		

Associations between genes / assays and cancer endpoints: Only the 3 most reproducible dataset were used (Novascreen, CellzDirect and Attagene) and all single-assay, single-cancer endpoint associations were calculated. Results with p<0.05 were kept. By chance, 41 such associations were expected and 58 were found. The table shows relative risk values. Green indicates there is literature support for the association.

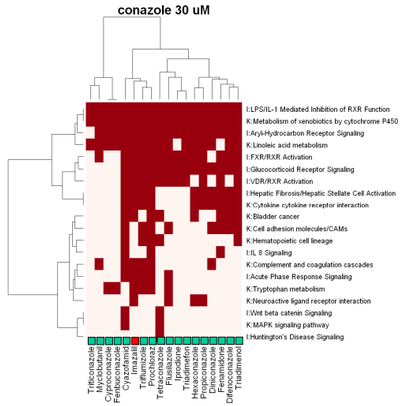
Gene/Pathway Perturbation Scores

Most assays were mapped to genes (313 human and 86 rat) and gene perturbation scores were calculated. These were the lowest AC50/LEL observed for a chemical for any assay mapped to that gene.

Genes were mapped to KEGG and Ingenuity pathways and pathway perturbation scores for a chemical were given the value of the lowest AC50/LEC value for any assay mapped to a gene mapped to that pathway. However, the chemical needed to be active in at least 5 assays mapped to that pathway for a pathway perturbation score to be calculated.



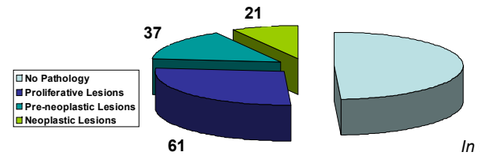
Number of pathways perturbed by chemicals varies widely both between and within classes



Conazole pathway coverage. Color band indicates (red) rat liver tumorigen

Multivariate Signature For Rat Liver Proliferative Lesions

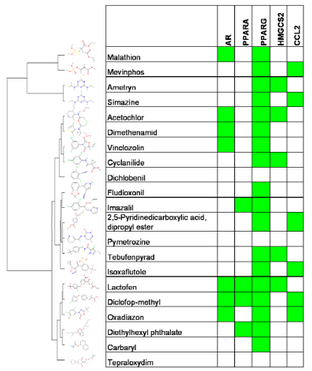
A Stepwise Logistic Regression model was created linking several assays with rat liver proliferative lesions. The signature genes are PPARG, HMGCS2 (surrogate for PPARA), CCL2 and AR, all of which are known to be associated with cancer, and all but CCL2 are linked to liver tumors in rodents and or man. The figure shows the hits of these assays against the subset of chemicals which are liver tumorigens in rats. 248/309 chemicals had rat data in ToxRefDB



In vivo data

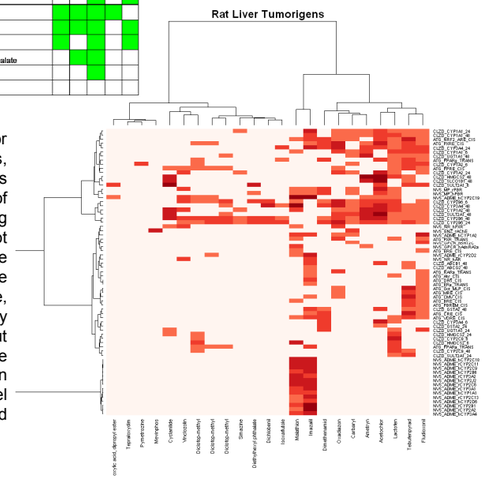
Signature	+	-
+	31	11
-	30	176

Sensitivity=51%
Specificity=94%



Prediction statistics for the Model
Specificity is high, so few false positives are seen, but sensitivity is low, indicating that many of these chemicals act through mechanisms not adequately probed by these assays, including the need for metabolic activation / deactivation

Model was built for proliferative lesions, but these figures examine the subset of chemicals causing liver tumors in rat. Not all tumorigens have hits in the predictive assays (above, chemicals ordered by chemical similarity) but they do have a wide range of activities in addition to the model assays (right, ordered by activity similarity)



Results/Conclusions

A simple predictive model using ToxCast assays and logistic regression produces a predictive signature for rat liver proliferative lesions. The genes showing up in the signature all have independent links to cancer. In particular, chemical-related PPAR activity is readily seen as a risk for rat liver proliferative lesions.

Impact and Outcomes

- 8 chemicals not in ToxRefDB were predicted to be positive for rat liver proliferative lesions
  - PFOA: Causes rat liver adenomas
  - PFOS: Causes rat liver adenomas
  - Diniconazole: rat liver hypertrophy
  - Chlorothalonil: rat liver enlargement, kidney tumors
  - TCMTB: testicular and thyroid adenomas
  - No data for Niclosamide, Methylene bis(thiocyanate), Phenoxyethanol

Future Directions

We are pursuing a variety of other analysis strategies. These include:

1. Focusing on particular chemical classes to better control for chemical structure
2. Including pharmacokinetic information as it becomes available
3. Focusing on specific chemicals where mode of action is well known and using these to understand the strengths and weaknesses of in vitro to in vivo mapping.

References

- M.T. Martin, R. Judson, D. Reif, D.J. Dix, R. Kavlock, "Profiling Chemicals Based on Chronic Toxicity Profiles from the U.S. EPA ToxRef Database", Environmental Health Perspectives, Vol. 117, 392-399 (2009)
- R.S. Judson, K.A. Houck, R.J. Kavlock, T.B. Knudsen, M.T. Martin, H.M. Mortensen, D.M. Reif, A.M. Richard, D.M. Rotroff, I. Shah, D. J. Dix "Predictive *In Vitro* Screening of Environmental Chemicals – The ToxCast Project", Environmental Health perspectives (Submitted 2009)