Predictive Signatures from ToxCast Data for Chronic, Developmental and Reproductive Toxicity Endpoints

Richard Judson, Matthew Martin, Thomas Knudsen, Nicole Kleinstreuer, David Dix, Keith Houck, Robert Kavlock
U.S EPA, ORD, National Center for Computational Toxicology

Goals

The goal of the ToxCast and Tox21 programs are:
1.Screening of hundreds to thousands of environmental chemicals using *in vitro* HTS methods
2.Development of predictive models or signatures by training against chemicals with known toxicity
3.Signatures should have relatively high sensitivity (few false negatives)
4.Prioritization of chemicals for further targeted testing, based on positive response for signatures indicating potential for particular modes of toxicity

Abstract

The EPA ToxCast program is using *in vitro* assay data and chemical descriptors to build predictive models for *in vivo* toxicity endpoints. *In vitro* assays measure activity of chemicals against molecular targets such as enzymes and receptors (measured in cell-free and cell-based systems in binding, agonist and antagonist modes), in addition to cellular phenotypes and pharmacokinetic-related parameters.

Over 600 separate assays were run in concentration response format to derive AC50 or LEC values for each chemical-assay combination. This collection of AC50 values were used to build statistically significant predictors of *in vivo* toxicity endpoints derived from chronic/carcinogenicity studies in rats and mice, prenatal developmental toxicity studies in rats and rabbits, and two-generation reproduction studies in rats.

A collection of 309 unique chemicals were tested and used for predictive signature development. Machine learning techniques were used to build multivariate predictive signatures. K-fold cross validation was performed to quantify prediction accuracy.

Three sets of mechanistic signatures are presented.
1.Chronic rat liver tumors based on assays including targets in the peroxisome proliferator-activating receptor (PPAR) pathway.
2.Prenatal cleft palate in rat fetuses based on assays targeting Wnt signaling, retinoids, AhR, TGF-beta, glucocorticoids and chemokines, identifying pathway level disruption of the Forkhead-family of transcription factors and chromatin remodeling.
3.Reproductive fitness toxicity seen in rat reproduction studies and based on *in vitro* assays, pathways and chemical descriptors.

Data Availability

http://www.epa.gov/ncct/toxcast
http://actor.epa.gov
http://toxrefdb.epa.gov (coming soon)

Methods and Data

Univariate Association Analysis

Use AC50 for assays as input (~500 assays)
- Log transform [-log10(AC50(μM)/100000)]
- Inactive values set to zero
- For dichotomous calculations, any activity=1

Perform two univariate tests for each assay-endpoint pair
- T-test, using continuous assay data, dichotomous endpoint
- Chi-squared using dichotomous assay and endpoint data

Multivariate Signature Development

Simple linear model (*linmod*)
- Calculate t-test p-value for all assays
- Sort assays by increasing p-value
- Calculate ROC curve as function of:
  - number of assays included and cutoff value (hit vs. no-hit)
  - Only 2 free parameters
- Select model with best sensitivity and specificity
- Use k-fold cross validation

Model=intercept + sum(-log10(p-value) for included 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).

Pathway score were calculated by taking the average log(AC50) for all assays that mapped to genes included in a pathway. Pathway sources included KEGG, Ingenuity and PathwayCommons. Only human pathways were included. ToxCast assays mapped to a total of 303 human genes.

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.

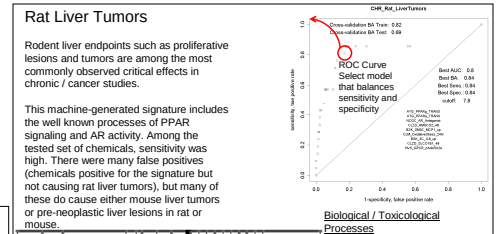
Endpoints from rat, mouse and rabbit chronic-cancer, multigenerational and developmental studies were extracted from ToxRefDB and dichotomized (cause / do not cause endpoint)

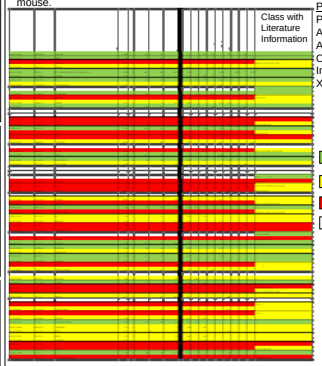
Cancer / Chronic Toxicity

Rat Liver Tumors

Rodent liver endpoints such as proliferative lesions and tumors are among the most commonly observed critical effects in chronic / cancer studies.

This machine-generated signature includes the well known processes of PPAR signaling and AR activity. Among the tested set of chemicals, sensitivity was high. There were many false positives (chemicals positive for the signature but not causing rat liver tumors), but many of these do cause either mouse liver tumors or pre-neoplastic liver lesions in rat or mouse.





Chemical Predicted Positive for Rat Liver Tumors Based on Signature

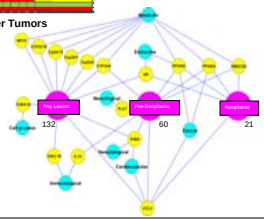
Biological / Toxicological Processes
PPAR Signaling
Androgen Receptor Activity
Angiogenesis (MCP1)
Oxidative Stress
Inflammatory response (IL8)
Xenobiotic removal (SLCO1B1)

Chemical by Chemical Results
- Rat Liver Tumorigen
- Rat Proliferative lesions or mouse liver tumorigen
- No Rat or mouse major liver effects
- No animal data

5 false positive are low-dose cholinesterase inhibitors

18/21 rat liver tumorigens identified

3 false negatives are active in no model assays



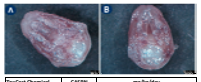
Summary of Univariate Associations

Look across disease progression
Includes gene-disease linkage

Developmental Toxicity

Cleft Palate (CLP)

CLP was selected as a prototype to build and evaluate a predictive model for dev-tox. CLP-actives are listed by categorical Lowest Effect Level (cLEL) else developmental LEL (dLEL) for rat and rabbit studies.

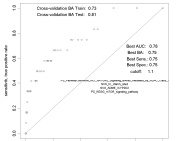


ToxCast Chemical	CASRN	mg/kg/day	
rat CLP		rat cLEL	rabbit dLEL
Carnoconazole	84361-06-1	24	>50
Triadimenole	55327-43-1	90	>400
Tri-allate	2303-17-5	90	not tested
Spiroxamine	118134-30-8	100	>80
Triadimefon	43121-43-3	100	>120
Triclopyr	55335-06-3	200	not tested
Fluazinam	79622-59-6	250	>12
Fluridazole	85509-19-9	250	>12
Myclobutanil	127740-64-3	512	>80
Dibutyl phthalate	84-74-2	750	not tested
Dibutyl phthalate	84-74-2	750	not tested
rabbit CLP		rat cLEL	rabbit cLEL
Cymoxanil	57966-95-7	>150	32
Diclobutrazol	1194-65-6	not tested	135

Univariate analysis in ToxMiner v16 showed 137 significant associations with ToxCast™ *in vitro* assays. Most (131) were rat-specific but 6 were rabbit-specific. The significant targets were used as inputs to build multivariate Toxicity Signatures, grouped by signaling system. No single association mapped to both species although close matches emerged for Forkhead Box transcription factor networks [FOXM1, FOXA] and signaling events mediated by Histone Deacetylase [HDAC-II, HDAC-I].

Assay Component - Rabbit	Position
PS_Pathway Commons_NCI_NATURE_FOXM1_transcription_factor_network	+0.000
PS_Pathway Commons_NCI_NATURE_Signaling_events_mediated_by_HDAC_Class_II	+0.000
PS_Gene_ESR1	+0.000
PS_Ingenuity_Estrogen_Receptor_Signaling	+0.000
KEGG_PPAR_CS	0.000
PUR_CYP19A1_all	0.000

Assay Component - Rat	Position
PS_Gene_TGFB1	0.000
PS_KEGG_TGF_beta_signaling_pathway	0.000
PS_Pathway Commons_CBX1_MAP_TGFBR	0.000
PS_Pathway Commons_REACTOME_Signaling_by_TGF_beta	0.000
PS_KEGG_Renal_cell_carcinoma	0.000
PS_Ingenuity_LPS/IL-1_Mediated_Inhibition_of_RXR_Function	0.000
PS_Ingenuity_PPARalpha/RXRa_Activation	0.000
PS_Ingenuity_RAR_Activation	0.000
PS_Pathway Commons_NCI_NATURE_Retinoic_acid_receptor_mediated_signaling	0.000
PS_Pathway Commons_NCI_NATURE_FOXA_transcription_factor_networks	0.000
PS_Pathway Commons_NCI_NATURE_Signaling_events_mediated_by_HDAC_Class_I	0.000
PS_Ingenuity_Aryl_Hydrocarbon_Receptor_Signaling	0.000
PS_Ingenuity_Wnt/beta_catenin_Signaling	0.000
PS_Ingenuity_Tight_Junction_Signaling	0.000
PS_KEGG_Neuroactive_ligand_receptor_interaction_Rattus_norvegicus	0.000
PS_Pathway Commons_NCI_NATURE_PDGFR_alpha_signaling_pathway	0.000
KEGG_ASTHMA_RCYP2C23	0.000
KEGG_ASTHMA_RCYP2D2	0.000
PS_Gene_CYP2C18	0.000
PS_Gene_Cyp2d6_Rattus_norvegicus	0.000
PS_Ingenuity_Xenobiotic_Metabolism_Signaling	0.000
PS_Pathway Commons_REACTOME_Inlo_Interactions_with_Influenza_Factors	0.000
PS_Pathway Commons_REACTOME_Influenza_Infection	0.000
PS_Pathway Commons_REACTOME_Influenza_Virus_Induced_Apoptosis	0.000



CLP pathways common to rat and rabbit

Diagnostic Pathway	TP	TN
rat CLP		
PS_Pathway Commons_NCI_NATURE_Signaling_events_mediated_by_HDAC_Class_I	15	0.00
PS_Pathway Commons_NCI_NATURE_FOXA_transcription_factor_networks	18	0.00
rabbit CLP		
PS_Pathway Commons_NCI_NATURE_Signaling_events_mediated_by_HDAC_Class_II	2	0.00
PS_Pathway Commons_NCI_NATURE_FOXM1_transcription_factor_network	2	0.00

Consensus model

A simple linear model predicted CLP for all CLP-actives except Tri-allate, Triclopyr, Cymoxanil and Diclobenil. On the other hand, the FOX and HDAC pathway models were of the highest diagnostic value in terms of sensitivity-specificity, including: estrogen receptor signaling, signaling events mediated by HDAC Class II, and the FOXM1 transcription factor network (BA = 0.83). Mutation or deregulation of FOX genes are associated with human congenital disorders, and CLP is one of the complications.

Conclusion

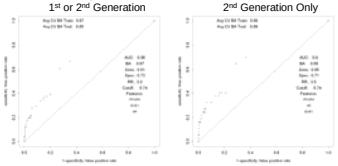
This toxicity signature posits a generalized role for the Forkhead-family of transcription factors and chromatin remodeling as targets for environmental chemicals in the mechanism underlying cleft palate.

Reproductive Toxicity

Reproductive Performance

Chemicals tested in multigenerational reproductive toxicity studies (primarily carried out for 2 generations) have been evaluated for their effects on reproductive performance. Reproductive performance is an endpoint made up of many specific effects, including effects on mating, fertility, gestation, implantation, and pregnancy. Each assay (where applicable) has been mapped to a gene and the average activity across all assays for a given gene have been used to develop this model, predicting reproductive performance effects.

Reproductive Performance Gene Model
1st or 2nd Generation | 2nd Generation Only



Assay Score
Positive = -log(AC50)+6
Negative = 0

Gene Scores
(AR,ESR1,PPARA) = Mean(Assay Score)

AR = Androgren Receptor
ESR1 = ERa = Estrogen Receptor Alpha
PPARA = Peroxisome proliferator-activated receptor alpha

GENE	ASSAY COMPONENT
AR	ATG_AR_TRANS
	NCGC_AR_Agonist
	NCGC_AR_Antagonist
	NVS_NR_hAR
	NVS_NR_rAR
ESR1	ATG_ERE_TRANS
	ATG_ERE_CIS
	NCGC_Ethagina_Agonist
	NCGC_Ethagina_Antagonist
	NVS_NR_hER
	NVS_NR_rER
PPARA	ATG_PPARa_TRANS
	NCGC_PPARa_Agonist
	NVS_NR_hPPARa

Conclusions
1. Improved model statistics when using 2nd Generation Only Endpoint
2. Simple linear model does not over-fit
3. Comparable sensitivity and specificity even though dataset is unbalanced
4. Biologically relevant targets identified
5. Predicts, accurately, many chemicals not used in training or testing of model (e.g., HPTE, Methoxychlor)

Reproductive Performance Gene Model (2nd Generation Only)		
True Positives (Total: 24)	**False Positives (Total: 59)**	%Z = Pos
Bisphenol A; Lactofen; Fenarimol; Fenthion; Imazalil; Vinclozolin; Fluxilazole; Triadimefon; Fipronil; Trifluminzole; Fluazinam; Clomazone; Diazinon; Emamectin benzoate; Triticonazole; Fenbuconazole; Maneb; Folpet; Boscalid; Resmethrin; Sulfentrazone; Propetamphos; Methidathion; Benfluralin	Diclofop-methyl; Fenhexamid; Bensulide; Flumetralin; Hexaconazole; Fludioxonil; Pyridaben; Malathion; Phosalone; Carbaryl; Linuron; Tetraconazole; Endosulfan; Oxyfluorfen; Acetochlor; Triadimenol; Dithiopyr; Propiconazole; Fenpropathrin; Dibutyl phthalate; Etoxazole; **38 more...**	
False Negatives (Total: 13)	**True Negatives (Total: 146)**	%Z = Neg
Trichlorfon; Dichlobenil; Mevinphos; Dichlorprop; Butafenacil; Carboxin; Daminozide; Dichlorvos; Norflurazon; Fluazifop-butyl; Propoxycarbazone-sodium; Prosulfuron; Tri-allate	Ethalfluralin; Cycloate; Mytobutanil; Ethofumesate; Cyclanilide; Cyromazine; MGK; Azinphos-methyl; Bendiocarb; Oxasulfuron; Cymoxanil; Bifenazate; Dichloran; Pyriproxyfen; Diphenylamine; Forchlorfenuron; Acibenzolar-S-Methyl; Pymetrozine; Propazine; Simazine; Tebufenpyrad; Icaridin; TCMTB; 2,4-D; **122 more...**	
Sensitivity = 65%	**Specificity = 71%**	
Predicted Positive - No ToxRefDB Study (Total: 26)	***Predicted Negative- No ToxRefDB Study (Total: 41)***	
HPTE; Methoxychlor; Bromoxynil; Clorophene; Fenitrothion; Perfluorooctanoic acid; Perfluorooctane sulfonic acid; Parathion; Napropamide; Prodiamine; MEHP*; Fluazifop-P-butyl; Tribufos; Cinmethylin; Isazofos; Chlorpyrifos oxon*; Milbemectin; Dimethomorph; Diniconazole; Prochloraz; Alachlor; Esfenvalerate; Fenoxycarb; Butralin; Hexythiazox; Disulfoton*	*Tefluthrin; Tebupirimfos; Amlazine; Spiroxamine; Cloprop; 6-Desisopropylatrazine; Oxytetracycline dihydrate; Clopyralid-olamine; Abamectin; Fluroxypyr-meptyl; Acifluorfen; Clofentezine; Bensulfuron-methyl; Boric acid; **25 more...***	

*Designates metabolites within chemical set

References

- Judson, et al., "A Comparison of Machine Learning Algorithms for Chemical Toxicity Classification Using a Simulated Multi-Scale Data Model" BMC Bioinformatics Vol. 9, 241 (2008)
- Martin, et al. "Profiling Chemicals Based on Chronic Toxicity Profiles from the U.S. EPA ToxRef Database", Environmental Health Perspectives, Vol. 117, 392-399 (2009)
- Knudsen, et al. "Profiling the Activity of Environmental Chemicals in Prenatal Developmental Toxicity Studies using the U.S. EPA's ToxRefDB", Reproductive Toxicology, Vol. 31, 241 (2009)
- Martin, et al. "Profiling the Reproductive Toxicity of Chemicals from Multigeneration Studies in the Toxicity Reference Database (ToxRefDB)", Toxicological Sciences, Vol. 110, 181 (2009).
- Judson, et al. "Predictive *In Vitro* Screening of Environmental Chemicals – The ToxCast Project", Environmental Health perspectives (EPub ahead of Print Dec 2009)

Building a scientific foundation for sound environmental decisions

This work may not necessarily reflect official Agency policy.