The ToxCast™ Pathway Database

Identifying Toxicity Signatures and Potential Modes of Action from Chemical Screening Data

Holly M. Mortensen, David Dix, Keith Houck, Robert Kavlock, Imran Shah, Richard Judson

National Center for Computational Toxicology, Office of Research and Development, United States Environmental Protection Agency, Research Triangle Park, NC.

Introduction

The US EPA ToxCast™ program is using *in vitro* HTS (High-Throughput Screening) methods to profile and model bioactivity of environmental chemicals. The main goals of the ToxCast program are to generate predictive signatures of toxicity, and ultimately provide rapid and cost-effective alternatives to animal testing. Application of HTS to environmental toxicants is a novel approach to predictive toxicology and health risk assessment, and differs in some important aspects from what is required for drug efficacy screening. The biochemical interaction of environmental chemicals are sometimes weaker than that seen with drugs and their intended targets. Additionally, the chemical space covered by environmental chemicals is much broader compared to that of pharmaceuticals (Knudsen, *et al.*, 2009).

The ToxMiner database has been created and added to the EPA's ACToR database (Judson *et al.*, 2008) (Figures 1 and 2). One purpose of the ToxMiner database is to link biological, metabolic and cellular pathway data to gene and *in vitro* assay data for the initial subset of chemicals screened in the ToxCast Phase I HTS assays. Also included in ToxMiner are human disease and species homology information, which correlate with ToxCast assays that affect specific genetic loci. This information is designed to make it possible to infer the types of human disease associated with exposure to these chemicals. This is of interest to the EPA because of these chemicals' potential to adversely affect human health.

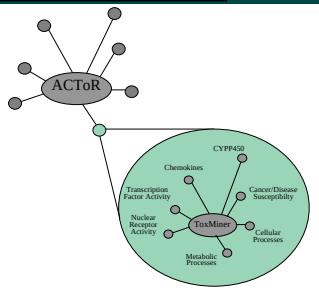


Figure 1: Schematic of the components of the ToxMiner database and its relation to the ACToR database

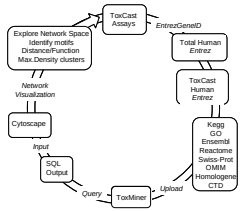


Figure 2: Schematic of the workflow of ToxCast data to publicly available information, to input for the ToxMiner database, followed by subsequent analyses

Primary Aims:

1. to assess application of HTS methods for use in the accurate characterization of toxicity profiles of the ToxCast Phase I chemicals
2. to provide a comprehensive resource of publicly available pathway information, with specific focus on chemicals of interest to the EPA
3. to derive pathway (biological, metabolic, and cellular) information that links each HTS assay and target chemical to its corresponding set of genes, proteins, transcription factors, etc.
4. use network analyses to visualize potentially complex pathway relationships, and delineate the types of pathways perturbed by the ToxCast chemicals, as well as to define putative modes of action, and human disease states

Table 1

ToxCast Phase I Chemical Summary (n=320)
309 unique structures
3 triplicates, 5 duplicates for QC
8 metabolites
291 pesticide actives
273 registered pesticide actives
22 pesticide inerts
33 antimicrobials

Table 2

MOA Classes (>3 chemical)	
acetylcholine esterase inhibitors	amide herbicides
conazole fungicides	aromatic fungicides
sodium channel modulators	chloroacetanilide herbicides
pyrethroid ester insecticides	chlorotriazine herbicides
organothiophosphate acaricides	growth inhibitors
dinitroaniline herbicides	organophosphate acaricides
pyridine herbicides	oxime carbamate insecticides
thiocarbamate herbicides	phenylurea herbicides
imidazolinone herbicides	pyrethroid ester acaricides
organophosphate insecticides	strobilurin fungicides
phenyl organothiophosphate insecticides	unclassified acaricides
aliphatic organothiophosphate insecticides	unclassified herbicides

More Information:
ToxCast: http://www.epa.gov/ncct/toxcast
ToxCast Data Analysis Summit: May 14, 15, RTP NC
Public Data Release June 2009 in ACToR: http://actor.epa.gov
Contact: mortensen.holly@epa.gov

Methods

Chemical Selection

Phase I of the EPA ToxCast program employs a chemical library containing 320 compounds (Table 1). The chemicals selected for Phase I are composed largely by a diverse set of pesticide active ingredients (Table 2), which had sufficient supporting *in vivo* data included as part of their registration process with the EPA. Other miscellaneous chemicals of environmental concern were also included. Chemical samples included were obtained from BioFocus DPI (South San Francisco, CA).

Assay Data Collection

We have evaluated biochemical and cellular high throughput screening (HTS) data from a total of 611 ToxCast Phase I assays, which use a range of technologies. Each assay target was evaluated and corresponding *Entrez* geneIDs were identified. For the human data specifically, a total of 318 unique genetic loci were associated with reported assay targets.

Database Generation

The ToxMiner workflow is illustrated in Figure 2. Pathway information for the ToxCast chemical dataset was obtained from several publicly available databases, shown in Figure 3. These data are preprocessed and consolidated into the ToxMiner database via scripts written in perl, and implemented using MySQL. The ToxMiner database extends the currently available ACToR (Aggregated Computation Toxicology Resource) database, which captures information on chemicals and assays of chemical-biological effects (Judson, *et al.*, 2008).

Pathway Network Construction and Analyses

Pathway data was downloaded from ToxMiner and formatted for network analysis using Cytoscape version 2.6.1. For the GO data specifically, inferred distances from root nodes (*e.g.* "hierarchical" classifications), were obtained directly from the GO website. Root node distances of less than four, where one is equivalent to all GO processes and seventeen represented the most distant node, were omitted from further analysis. Additionally, GO processes were limited for each ToxCast assay in terms of the number of loci reported for each process, in that no GO process included less than two genes and no more than 50 genes.

Figure 3: Pathways/Processes Identified by Analyses of ToxCast Data

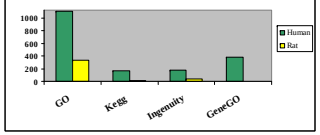
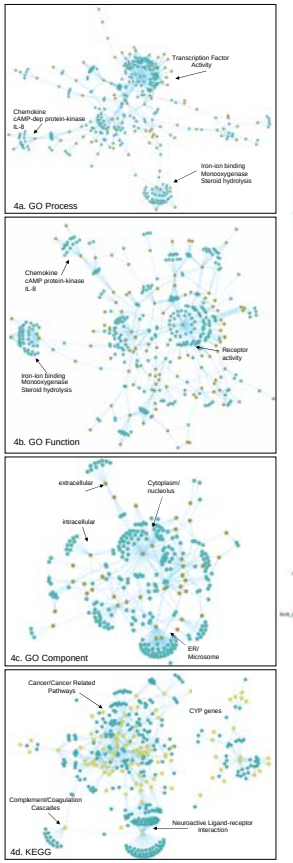


Figure 3 illustrates the number of distinct pathways or processes, for both human and rat, identified from each publicly available data source when queried against Entrez geneIDs identified for ToxCast Phase I assays.

Table 3 Assay Abbreviations

ToxCast Assay Source	*Abbreviation*
Novascreen	NVS
BioSeek	BSK
Attagene	ATG
NIH Chemical Genomics Center	NCGC
CellzDirect	CZD

Table 3 lists the ToxCast assay source abbreviations used in Figures 5 and 6.



Figures 4 (a-d): Pathway/Process Networks generated using Cytoscape 2.6.1 illustrating the overall structure and relationship of ToxCast Phase I Assays (blue nodes) to GO (Genetic, Biological and Cellular) processes (brown nodes) and KEGG Pathways (yellow nodes)

ToxCast Assay-Phase I GO Process KEGG Pathway

Figure 5: Exploded view of the ToxCast Assay (brown nodes) associated with GO nuclear receptor activity (blue nodes) (also see Figure 4b and Table 4).

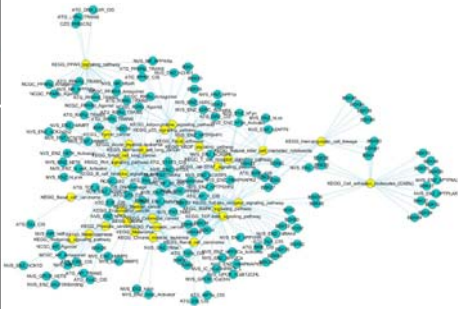


Figure 6: Exploded view of the ToxCast Assay (blue) associated with KEGG Cancer and cancer related pathways (yellow) (also see Figure 4d and Table 5).

Table 4: GO Receptor Activity Processes Associated with ToxCast Assays

Species	GO Receptor Activity	Genes	Assays	Unique Genes in Assays	Ratio
M.musculus	beta-3 adrenergic receptor binding	2	1	1	50%
M.musculus	PLC activating metabotropic glutamate receptor activity	3	1	1	33%
R.norvegicus	A2A adenosine receptor binding	2	1	1	50%
R.norvegicus	corticotropin-releasing hormone receptor 1 binding	1	1	1	100%
R.norvegicus	corticotropin-releasing hormone receptor 2 binding	1	1	1	100%
R.norvegicus	V1A vasopressin receptor binding	2	1	1	50%
H.sapiens	beta3-adrenergic receptor activity	3	2	2	67%
H.sapiens	glutamate receptor binding	16	2	2	13%
H.sapiens	CD4 receptor binding	6	1	1	17%
H.sapiens	CD8 receptor binding	3	1	1	33%
H.sapiens	dopamine D2 receptor activity	3	1	1	33%
H.sapiens	dopamine D4 receptor activity	3	1	1	33%
H.sapiens	endothelin-A receptor activity	2	1	1	50%
H.sapiens	estrogen receptor activity	4	5	1	25%
H.sapiens	gastric inhibitory peptide receptor activity	5	1	1	20%
H.sapiens	hematopoietin/interferon-class (D200-domain) cytokine receptor signal transducer activity	4	1	1	25%
H.sapiens	interleukin-8 receptor binding	3	5	1	33%
H.sapiens	leukotriene B4 receptor activity	7	1	1	14%
H.sapiens	macrophage colony stimulating factor receptor binding	5	3	1	20%
H.sapiens	MHC, class II receptor activity	13	3	1	8%
H.sapiens	thromboxane A2 receptor activity	1	1	1	100%
H.sapiens	type II transforming growth factor beta receptor binding	10	3	1	10%
H.sapiens	U-plasminogen activator receptor activity	2	4	1	50%
H.sapiens	very-low-density lipoprotein receptor activity	6	1	1	17%
H.sapiens	vitamin D3 receptor activity	2	3	1	50%

Table 4 lists all NRs identified from ToxCast HTS using the network analysis presented in both Figures 4b and 5.

Table 5: KEGG Cancer Related Pathways Associated with ToxCast Assays

Species	KEGG Cancer Activity	Genes	Assays	Unique Genes in Assays	Ratio
R.norvegicus	MAPK signaling pathway	229	3	3	1%
R.norvegicus	Prostate cancer	84	1	1	1%
R.norvegicus	Small cell lung cancer	74	1	1	1%
H.sapiens	MAPK signaling pathway	258	27	19	7%
H.sapiens	Toll-like receptor signaling pathway	104	26	14	13%
H.sapiens	Prostate cancer	89	18	13	15%
H.sapiens	Colorectal cancer	86	17	12	14%
H.sapiens	T cell receptor signaling pathway	93	14	12	13%
H.sapiens	Chronic myeloid leukemia	77	16	11	14%
H.sapiens	Natural killer cell mediated cytotoxicity	136	13	11	8%
H.sapiens	Acute myeloid leukemia	57	14	10	18%
H.sapiens	Bladder cancer	42	10	10	24%
H.sapiens	Pancreatic cancer	72	14	10	13%
H.sapiens	Small cell lung cancer	87	14	10	11%
H.sapiens	Wnt signaling pathway	148	16	10	7%
H.sapiens	B cell receptor signaling pathway	64	12	9	14%
H.sapiens	Cell adhesion molecules (CAMs)	133	19	9	7%
H.sapiens	Non-small cell lung cancer	54	13	9	17%
H.sapiens	Renal cell carcinoma	69	13	9	13%
H.sapiens	Endometrial cancer	52	10	8	15%
H.sapiens	Jak-STAT signaling pathway	156	8	8	5%
H.sapiens	TGF-beta signaling pathway	86	10	8	9%
H.sapiens	Melanoma	71	9	7	10%
H.sapiens	Thyroid cancer	29	10	7	24%
H.sapiens	Cell cycle	112	11	6	5%
H.sapiens	p53 signaling pathway	68	9	6	9%
H.sapiens	VEGF signaling pathway	73	7	6	8%
H.sapiens	Basal cell carcinoma	55	8	4	7%

Table 5 lists all cancer and cancer related pathways identified from ToxCast HTS using the network analysis presented in both Figures 4d and 6.

Discussion and Future Directions

Though this preliminary analysis of the ToxCast Phase I HTS data using the newly constructed ToxMiner database focuses only on a subset of pathway sources, we have confirmed the relevance of the HTS assays in testing the metabolic, biological and cellular systems for which they were intended. The organization of this extensive dataset, and visualization via network analyses, has highlighted several directions for future inquiry. Through the use of the ToxMiner database and the analysis framework presented here, we hope to address relationships between potential disease states in humans and environmental chemicals, as well as contribute to the larger goals of toxicogenomics by clarifying the role of gene-environment interactions in disease states.

Acknowledgements

We would like to give express thanks to David Scoville for his meticulous identification of geneIDs associated with the ToxCast Assays, Amar Singh for his thoughtful assistance with the initial data manipulation and thoughts on database design, and David Reif for helpful discussion.

References

Dix, D., Richard, A., Houck, K. and Judson, R. 2008. ToxCast Phase I Chemical Library. Available: http://www.epa.gov/ncct/toxcast/chemicals.html.

Knudsen, T., Houck, K., Judson, R., Singh, A., Mortensen, H.; Reif, D., Dix, D., Kavlock, R. 2009. Biochemical Activities of 320 ToxCast Chemicals Evaluated Across 239 Functional Targets (In Prep).

Judson, R., Richard, A, Dix, D., Houck, K., Elloumi, F., Martin, M., Cathey, T., Transue, T.R., Spencer, R., Wolf, M. 2008. ACToR-Aggregated Computational Toxicology Resource. Toxicology and Applied Pharmacology. Nov 15;233(1):7-13.