

Using the ToxMinerTM **Database** for Identifying Disease-Gene Associations in the ToxCastTM Dataset Holly M. Mortensen, David Dix, Keith Houck, Robert Kavlock, and Richard Judson



Introduction

The US EPA ToxCastTM program is using in vitro, highthroughput screening (HTS) to profile and model the bioactivity of environmental chemicals. The main goal of the ToxCast program is to generate predictive signatures of toxicity that ultimately provide rapid and cost-effective methods to prioritize chemicals for targeted in vivo testing, thus improving efficiency of the use of animals in those

bioassays.

The ToxMiner database was created in order to link biological, metabolic, and cellular pathway data to genes and in vitro assay data for the chemicals screened in the ToxCast Phase I HTS assays. Also included in ToxMiner is human disease information, which correlates with ToxCast assays that target specific genetic loci. We have implemented initial pathway inference and network analyses, which allow linkage of the types of adverse health outcomes with exposure to chemicals screened in Phase I. This approach permits exploration of disease at a higher level of cellular and organismal organization, focusing on multiple, related disorders, and may aid in the understanding of common disease outcomes (e.g. cancer or immune disorders) that are characterized by locus heterogeneity. Through the use of the ToxMiner database and the analysis framework presented here, we hope to gain insight in to 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 pathobiology.

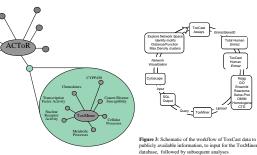


Figure 2: Schematic of the components of the T database and its relation to the ACToR database

Primary Aims:

- To assess application of HTS methods for use in the characterization of bioactivity profiles of the ToxCast Phase I To derive pathway (biological, metabolic, and cellular) information that links each ToxCast HTS assay and target
- chemical to its corresponding set of genes, proteins, transcription factors, etc.
- To provide a comprehensive resource of publicly available pathway information, with specific focus on chemicals of
- To 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 adverse disorders.

Methods

PACELIUMS

Chemical Selection
Phase I of the EPA ToxCast program employs a chemical library containing 320 compounds
(Table I). 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 miscellamous 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* genelDs were identified. For the human data specifically, a total of 318 unique genetic loci were associated with reported assay targets.

Pathney Network Construction and Aualyses
Pathney and 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.

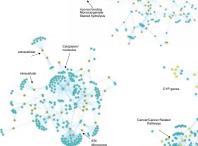
2 pesticide inerts Table 2 MOA Classes (>3 chemica nazole fungicides dium channel modulators



Figure 4: Pathways/Processes Identified by Analyses of ToyCast Data



5a GO Pro



5d. KEGG



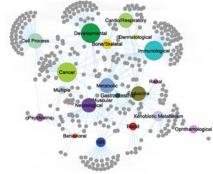
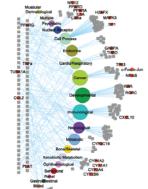


Figure 6: Network diagram of the disorder classes probed by ToxCast Phase I HTS Assays. Colors denote grouped disease types defined by OMIM disease-gene associations and based loosely on disease classes used by Goh et al. PNAS (2007). Node size corresponds to the number of distinct loci present in each disorder class.





Genetic Loci Associated with Rat Liver Proliferative Lesion:

Figure 7: Network diagram of the genetic loci and corresponding disorder classes probed by ToxCast Phase I HTS Assays Colors denote grouped disease types defined by OMIM disease-gene associations and based loosely on disease classes used by Goh et al. PNAS (2007). Node size corresponds to the number of distinct loci present in each disorder class. All genetic loci are illustrated in gray, with the exception of those shown in red that are discussed in the recent article by Jadoon et al.

Figure 1: Adapted from Collins et al.,

Public Data Release June 2009 in ACToR: http