High Throughput Prioritization for Integrated Toxicity Testing Based on ToxCast Chemical Profiling

David Reif[1], Keith Houck[1], Richard Judson[1], Robert Kavlock[1], Thomas Knudsen[1], Matthew Martin[1], Ann Richard[1], Imran Shah[1], Thomas Carpenter[2], David Dix[1]

U.S. Environmental Protection Agency, Office of Research and Development, National Center for Computational Toxicology[1], Office of Ground and Drinking Water, Office of Water[2]

Objectives

Develop a prioritization framework for diverse prioritization tasks relating to systemic, cancer, developmental or reproductive toxicity testing that provides:

- Integration over multiple domains of information
- Extensibility to incorporate existing knowledge, prioritization schemes, and different types of data (e.g. measures of biotransformation, exposure, dosimetry)
- Multivariate assessment of toxicity relative to any set of chemicals
- Transparency in relative score for each chemical
- Flexibility to customize components for diverse prioritization tasks

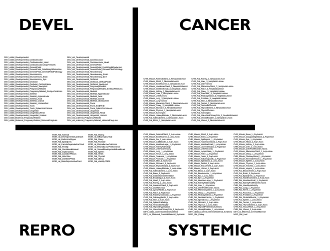
Abstract

The rational prioritization of chemicals for integrated toxicity testing is a central goal of the U.S. EPA's ToxCast™ program (http://epa.gov/ncct/toxcast/). ToxCast includes a wide-ranging battery of over 500 *in vitro* high-throughput screening assays which in Phase I was used to screen a library of 309 environmental chemicals at a cost <1% of that required for extensive animal testing. Various statistical and biological models have been employed to make associations between Phase I *in vitro* bioactivity and other data domains. We have now developed a flexible prioritization support software tool incorporating ToxCast *in vitro* bioactivity profiles, inferred toxicity pathways, *in vitro* to *in vivo* dosimetry estimates, and chemical structural descriptors. This approach calculates a comprehensive toxicity potential and provides multivariate visualizations representing the relative contribution of each data domain to an overall priority score. We demonstrate custom implementations for four prioritization tasks relating to systemic, cancer, developmental or reproductive toxicity testing. ToxCast scores are calculated as a function of specific chemical properties, *in vitro* assays, pathways and dosimetry features selected for each prioritization and type of toxicity testing. Features can be customized to a wide range of specific prioritization tasks (e.g. MOA-specific features relating to endocrine disruption); domains can be added to represent additional data (e.g. exposure potential); and domains can be up- or down-weighted to reflect relative value and give extra emphasis to specific features. Initial results indicate that combining multiple data domains into an overall weight of evidence approach for prioritization produces more robust conclusions than any single type of data taken alone.

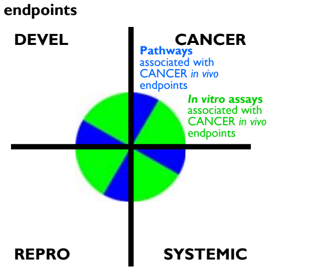
Profiling & Prioritization Strategy

1. Organize *in vivo* endpoints into four sectors

The *in vivo* endpoints from studies captured in ToxRefDB (see Martin et al. 2009) that had at least one statistically significant association with *in vitro* ToxCast assays were organized into four sectors of toxicological relevance. The Cancer (CANCER=C) sector included endpoints such as neoplastic lesions and tumors. The Developmental (DEVEL=D) sector included endpoints such as cleft palate and skeletal abnormalities. The Reproductive (REPRO=R) sector included endpoints such as litter size and offspring survival. The Systemic (SYSTEMIC=C) sector included endpoints such as non-cancer lesions to specific organs and systemic outcomes not specified in another sector.



2. Identify *in vitro* assays and pathways associated with each sector of *in vivo* endpoints



ToxCast assay results (AC_{50} values) and pathway perturbation scores (see Judson et al. 2010) were mapped to sectors if there was a significant statistical association between an assay/pathway and an *in vivo* endpoint for that sector.

3. Define the ToxPi™

Each chemical profile/signature/fingerprint gives a **Tox**icological **P**riority **I**ndex (**ToxPi**) score used for ranking chemicals

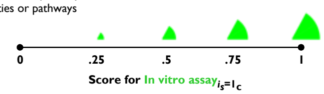
The ToxPi is a weighted sum of the components across all sectors: {C,D,R,S} ∈ S

$$\text{ToxPi} = f(\text{In vitro assays} + \text{Pathways}) = \sum_{1}^{S}\left(\sum_{1}^{I_s} w_{i_s} * \text{assay}_{i_s} + \sum_{1}^{P_s} w_{p_s} * \text{pathway}_{p_s}\right)$$

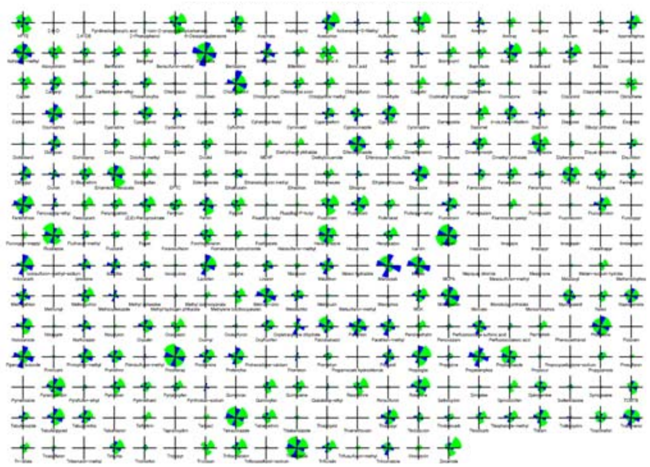
ToxPi terminology:

Sector: area of toxicological concern (In this case, one of four: Cancer, Developmental, Reproductive, Systemic)
Domains: basic data types represented by slices of a given color family (*in vitro* assay slices in shades of green, pathways in blue, etc.)
Slices: represent data from related assays, chemical properties or pathways
Components: data from individual assays, chemical properties or pathways

Here, each component ToxPi "score" is the normalized potency for all assays/pathways in that slice. Thus, scores closer to one indicate higher potency (i.e. lower AC_{50} values for *in vitro* assay components and higher perturbation scores for pathway components).

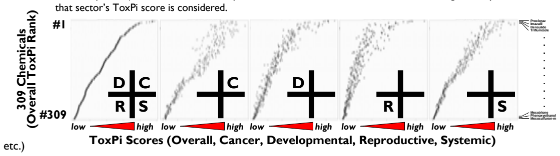


4. Calculate ToxPi scores for all 309 ToxCast Phase-I chemicals



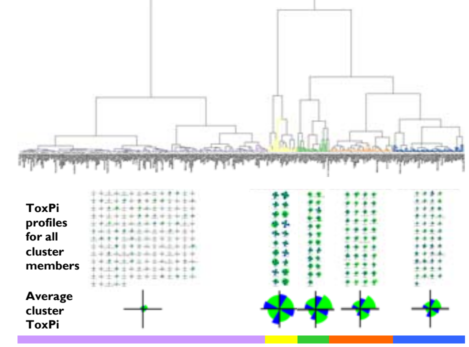
5. Explore ToxPi for sector-specific scores

The vertical axis shows the 309 chemicals ranked by overall ToxPi score in descending order (one chemical per line). The four sector-wise plots show how the score for each chemical changes if only that sector's ToxPi score is considered.



6. Cluster chemicals according to ToxPi scores

Hierarchical clustering (Euclidean distance; complete linkage) was used to segregate chemicals into clusters based upon slice-wise ToxPi scores.



Conclusions

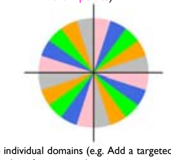
This implementation indicates that an integrated approach, wherein results across multiple data sources are simultaneously incorporated into chemical prioritization, can identify bioactive chemicals worthy of additional scrutiny. While the unsupervised clustering partially recapitulates sectors of *in vivo* activity reflected in ToxCast assays, the unbiased, data-driven assignment of *in vitro* assays into sectors does not provide optimal resolution of sector-specific toxicities.

The framework developed here provides graphical insight into the multiple aspects considered in chemical profiling and prioritization. It is amenable to incorporating extant prioritization schemes and relevant data from diverse sources, thereby facilitating meta-analysis across Agency resources. Because ToxPi score indices are intended for relative ranking, particular implementations of this framework can be continually updated with new chemicals and future data from subsequent phases of ToxCast, Tox21, and other sources.

Future Directions

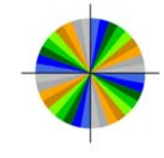
Incorporate additional components (slices) that may be from other domains (e.g. Consideration of exposure potential, chemical properties, and *in vivo* study results)

ToxPi = f(Exposure + Chemical properties + In vitro assays + Pathways + In vivo endpoints)



Customize individual domains (e.g. Add a targeted set of component slices for a particular prioritization task)

ToxPi = f(Exposure + Chemical properties + In vitro assays + Pathways)



Adjust weighting schemes according to specific prioritization tasks or component (slice) meaning (e.g. The weight ($w_{e=1}$) of exposure slice #1 has been increased)

ToxPi = f(Exposure + Chemical properties + In vitro assays + Pathways)

References

Reif et al. (*submitted*) **Endocrine Profiling and Prioritization of Environmental Chemicals Using ToxCast Data**

Judson et al. (2010) **In Vitro Screening of Environmental Chemicals for Targeted Testing Prioritization - The ToxCast Project** *Environmental Health Perspectives*

Martin et al. (2009) **Profiling Chemicals Based on Chronic Toxicity Results from the U.S. EPA ToxRef Database** *Environmental Health Perspectives*

Building a scientific foundation for sound environmental decisions