

ToxPi GUI: an interactive visualization tool for transparent integration of data from diverse sources of evidence

David M. Reif^{1,*}, Myroslav Sypa², Eric F. Lock², Fred A. Wright³, Ander Wilson¹, Tommy Cathey⁴, Richard R. Judson¹ and Ivan Rusyn²

¹National Center for Computational Toxicology, U.S. Environmental Protection Agency, Durham, ²Departments of Environmental Sciences and Engineering and ³Department of Biostatistics, University of North Carolina, Chapel Hill and ⁴Lockheed Martin Corporation, Information Systems & Global Services, Durham, NC, USA

Associate Editor: Olga Troyanskaya

ABSTRACT

Motivation: Scientists and regulators are often faced with complex decisions, where use of scarce resources must be prioritized using collections of diverse information. The Toxicological Prioritization Index (ToxPiTM) was developed to enable integration of multiple sources of evidence on exposure and/or safety, transformed into transparent visual rankings to facilitate decision making. The rankings and associated graphical profiles can be used to prioritize resources in various decision contexts, such as testing chemical toxicity or assessing similarity of predicted compound bioactivity profiles. The amount and types of information available to decision makers are increasing exponentially, while the complex decisions must rely on specialized domain knowledge across multiple criteria of varying importance. Thus, the ToxPi bridges a gap, combining rigorous aggregation of evidence with ease of communication to stakeholders.

Results: An interactive ToxPi graphical user interface (GUI) application has been implemented to allow straightforward decision support across a variety of decision-making contexts in environmental health. The GUI allows users to easily import and recombine data, then analyze, visualize, highlight, export and communicate ToxPi results. It also provides a statistical metric of stability for both individual ToxPi scores and relative prioritized ranks.

Availability: The ToxPi GUI application, complete user manual and example data files are freely available from <http://comptox.unc.edu/toxpi.php>.

Contact: reif.david@gmail.com

Received on May 20, 2012; revised on September 18, 2012; accepted on November 16, 2012

1 INTRODUCTION

Prioritization decisions must be made in any field where available time, financial and logistical resources cannot address requirements (e.g. scientific, commercial or regulatory). Proper formulation, evaluation and communication of such prioritizations require effective decision support tools. The National Research Council's report *Science and Decisions* (NRC, 2009) underscored the urgent need for safety testing of tens of thousands of commercial chemicals that lack information on potential hazards to human health. Effective prioritization of chemicals is needed to bridge the gap between available knowledge and information

needed to meet legislative and public health mandates. Prioritization of chemicals for screening is necessary for linking downstream decisions with information from traditional toxicity studies, as well as data from new high-throughput screening (HTS) and other technologies (Rusyn *et al.*, 2012). The decision-making contexts typically revolve around two broad types of questions: 'In what order do we test chemicals?' and 'Which subsets of chemicals deserve specific, targeted tests?'

The ToxPi (Reif *et al.*, 2010) approach addresses both of these needs by yielding an explicit prioritized order of chemicals, as well as a visually interpretable and transparent weight-of-evidence analysis. This approach can be tailored to diverse sets of chemicals (or other entities of interest), evidence (data) and prioritization tasks. Its initial applications have been to aid in the prioritization and profiling of chemicals by integrating data from *in vitro* assays, chemical properties and inferred biological pathway activity collected for the ToxCastTM program (www.epa.gov/nceet/toxcast).

2 FEATURES AND IMPLEMENTATION

The ToxPi GUI is implemented in Java (freely available from <http://www.java.com>) to allow cross-platform sharing of data and results. All installation details and features described here are fully detailed in the user manual at the download site.

On opening the GUI, the main data screen (*Home* window) is displayed. This screen includes buttons for adding data files, then for selecting chemicals and data sources ('components') of interest. Example data files in a simple spreadsheet-compatible format are provided alongside the ToxPi application download. If a single ToxPi-format file is selected, users can recreate an entire ToxPi analysis in one-step. This functionality is for users who wish to share analyses or adjust/explore analyses of others. For new formulations, data from multiple files may be flexibly recombined from the *Home* window. The next step proceeds through an explicit formulation of ToxPi components, slices (aggregations of one or more individual data components), weights (relative emphasis across slices of data), coloration and options for mathematical scaling via the *Slices Information* window.

The *Results* window shows the sortable ToxPi index values and graphical profiles for all chemicals (Fig. 1, top panel). Users can toggle between steps (*Home* or *Slices Information*) to modify source data or formulation options. The entire set of results

*To whom correspondence should be addressed.

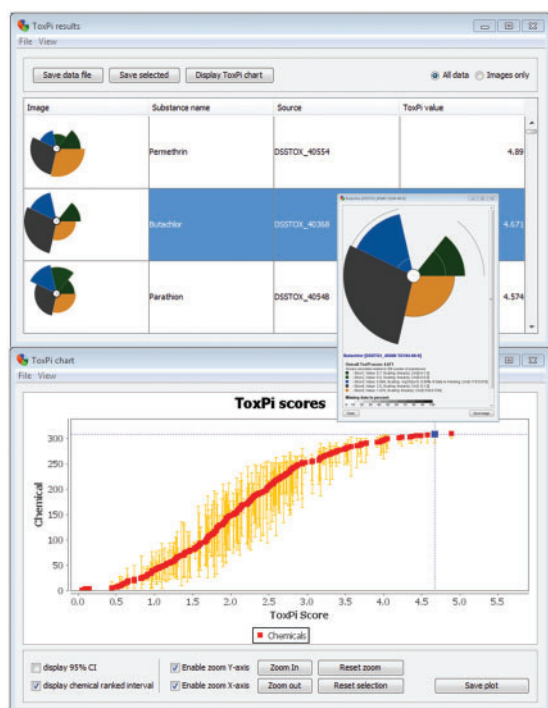


Fig. 1. Example of relationship between *Results* and *Chart* windows. The upper panel shows sorted ToxPi results, with the highlighted reference chemical (rank #2) signified by the bold square and cross-hairs on the lower panel. *Inset*: pop-up high-resolution window showing individual chemical's ToxPi and information

and parameters can be saved to an external spreadsheet-like file for sharing, publication or later use. All ToxPi profiles can be saved as an image, with individual profiles arranged in a sortable grid at publication-quality resolution. Alternatively, particular chemicals can be highlighted, with selected results saved.

Clicking on any chemical in the *Results* window brings up the detailed *Chemical* window (inset in Fig. 1). For the selected chemical, this window includes an exportable high-resolution ToxPi image containing details of slice-wise scores, associated confidence intervals and scaling. To represent the proportion of missing data in each slice, a small inner circle near the origin is shaded.

Confidence intervals are calculated by bootstrap sampling of the component values (source data) within each slice. The values are sampled with replacement 1000 times, and a new ToxPi score is calculated for each sample, yielding a distribution of bootstrapped scores. The 95% confidence interval (CI) is then estimated from this distribution, with the upper and lower bounds included in all tabular results and ToxPi graphics. These CI estimates convey the stability of results and make no distributional assumptions. Effectively, the method described here will penalize (by creating large CIs) slices that contain extreme outliers, rely on few source components or combine measurements on discordant numerical scales within a single slice. Such penalization is appropriate to convey inherent uncertainty and to provide incentives to find appropriately uniform scaling for within-slice measurements or create additional slices to appropriately group data types.

The *Chart* window shows all results in the context of the entire set of chemicals, where each square plots the ToxPi score for a

given chemical on the horizontal axis versus its relative rank on the vertical axis (Fig. 1, lower panel). The dynamic chart displays any combination of the following interactive capabilities: display/hide CI bars (score and/or rank); zoom in/out; pan/drag in any direction; highlight chemicals within the distribution. The *Chart* window is the graphical complement to the tabular prioritization results, and it can be saved to a file at any step. If users highlight appropriate reference chemicals—akin to ‘standards’ in a typical experiment—within this distributional chart and display CI bars, this window can be used to assess both the quality (correct order of reference chemicals) and stability (non-overlap of CI bars between key sections of the distribution) of a given prioritization scheme.

3 IMPACT AND FUTURE DIRECTIONS

ToxPi profiles could also be used for chemical *read-across*, analogous to chemical structure-alert ‘likeness’ groupings. The read-across can be implemented in terms of overall ToxPi bins/clusters of chemicals having similar profiles, or subsets of slices can be interpreted as ‘alerts’ (e.g. specific *in vitro* data) to support targeted testing decisions or tiered assessment strategies. The output of the ToxPi GUI can be directly used for such clustering exercises, and tabular files can be saved for transparency.

The current applications of ToxPi have focused on prioritization of chemicals; however, any units of categorization with varied output profiles could be used. Thus, the application may be used, for example, to compare groups of subjects across clinical, molecular, behavioral, socio-economic indicators and environmental data, facilitating additional novel integrated analyses and displays. The major advantage of the ToxPi framework stems from its flexibility as diverse types of data can be incorporated (e.g. personalized exposure measurements), and results can be updated as more information becomes available or if the underlying assumptions on data integration strategies are modified. This flexibility comes with the same caveats as with any generally useful analytical tool, whereby users must exercise sound judgment in the numerical representation and intended interpretation of input data.

Future prioritization approaches will have an increasing demand for transparency and interactivity. This demand is driven by pressures from a more informed public on both commercial and government entities. Transparency in both data and models is facilitated by web-accessible databases and software tools. The interactive ToxPi prioritization interface allows formal input into decisions by stakeholders, regulators and the public.

Funding: U.S. EPA STAR Program (RD83382501).

Conflict of Interest: none declared.

REFERENCES

- National Research Council. (2009) *Science and Decisions: Advancing Risk Assessment*. The National Academies Press, Washington, DC.
- Reif, D.M. *et al.* (2010) Endocrine profiling and prioritization of environmental chemicals using ToxCast data. *Environ. Health Perspect.*, **118**, 1714–1720.
- Rusyn, I. *et al.* (2012) Predictive modeling of chemical hazard by integrating numerical descriptors of chemical structures and short-term toxicity assay data. *Toxicol. Sci.*, **127**, 1–9.