

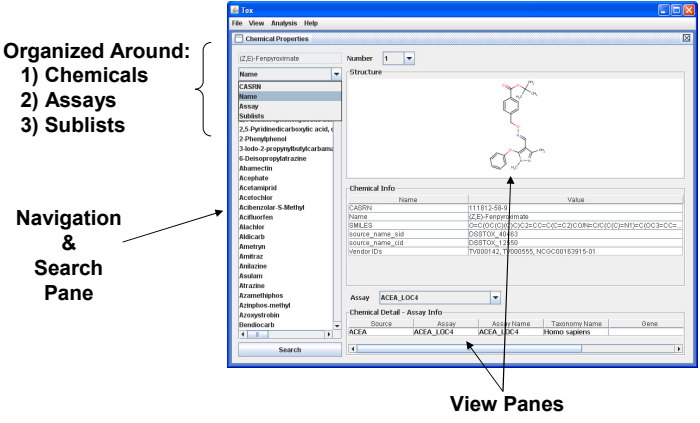
ToxMiner™ Interface for Visualizing and Analyzing ToxCast Data

Matthew T. Martin¹, Ling-Chieh Tsai², David M. Reif¹, David J. Dix¹, Richard S. Judson¹, Russell S. Thomas²
¹ U.S EPA, ORD, Computational Toxicology Research Program; ² The Hamner Institutes for Health Sciences, Research Triangle Park, NC

Science Question

The ToxCast dataset represents a collection of assays and endpoints that will require both standard statistical approaches as well as customized data analysis workflows. To analyze this unique dataset, we have developed an integrated database and Java-based interface. The software is organized around both chemical and assay centric visualization and analysis tools. Currently, these visualization and analysis tools include standard views of chemical and assay properties and results, the ability to define subsets of chemicals and assay results based on specified criteria, correlation matrices across assays, hierarchical clustering, and relative risk calculations. Machine learning algorithms have been added using Weka. The software will be made freely available.

Figure 1: ToxMiner is organized around chemicals, assays, and user created lists of chemicals or assays (sublists). Multi-pane views enables easy searching and visualization of results.



Organized Around:
 1) Chemicals
 2) Assays
 3) Sublists

Navigation & Search Pane

View Panes

Figure 2: Software features include (A) multiple data visualization modes customized to specific data types; and (B) an array of data analysis tools.

Methods/Approach

Figure 3: Screen shots of the different data visualization modes including (A) Data Collections; (B) Assay Properties; (C) In Vitro Pharmacokinetics; and (D) Phenotypes Associated with Assays.

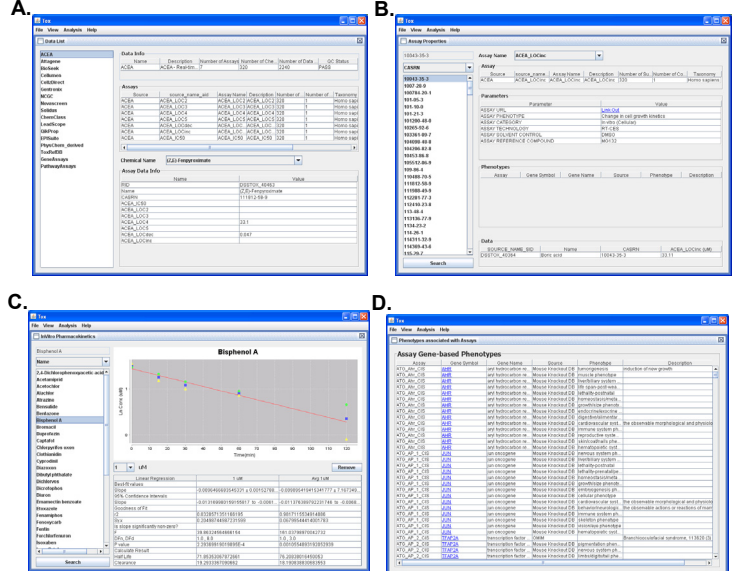
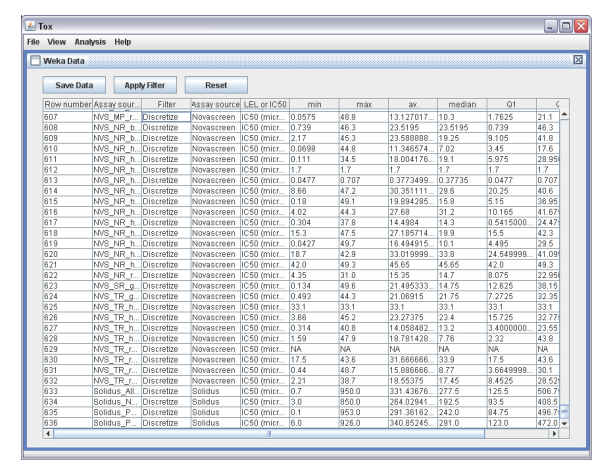


Figure 5: Weka Data: Screen shots of the ToxCast data with a Discretization Filter Applied. Weka Data allows the user to perform custom filtering and transformation before running a predictive algorithm or other analysis methodologies.



Results/Conclusions

Java-based program with Swing graphical user interface (GUI). The software is designed to be both chemical and assay centric with multiple visualization and analysis tools. Generate summaries of chemical and assay properties, pharmacokinetic properties, and assay associations with genes and phenotypes. Perform basic comparisons between and across assays and with endpoints in ToxRefDB and a limited assessment of assay predictivity of ToxRef endpoints using relative risk calculations.

Impact and Outcomes

Although still in the testing and development stages, this tool has shown promise to serve as a primary route of data release to the scientific community.

The ToxMiner Interface creates an environment for data experimentation and for testing the feasibility and understandability of new methods and downstream applications.

Research Goals

Organizing, visualizing, and analyzing the large amount of data generated by the ToxCast effort is going to be a significant challenge for the toxicology community. There are many unique characteristics and applications of the data that make currently available software solutions suboptimal for analysis. Therefore, building a versatile software platform that is specifically designed to interact with the ToxCast data and evaluate hazard, dose-response, and dosimetry for chemicals of interest will enable researchers to make the most of the rich dataset and enable regulators to make better decisions.

A. Data Visualization Modes

B. Data Analysis Tools

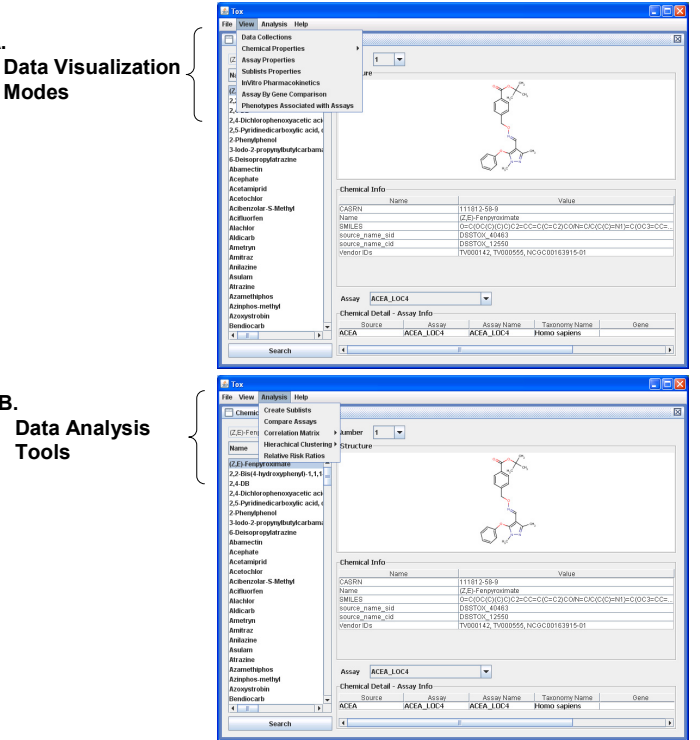


Figure 4: Screen shots of the different data analysis tools including (A) Venn diagrams; (B) Correlation Matrices (both Assay-to-Assay and Assay-to-ToxRefDB); (C) Hierarchical Clustering; and (D) Relative Risk Calculations.

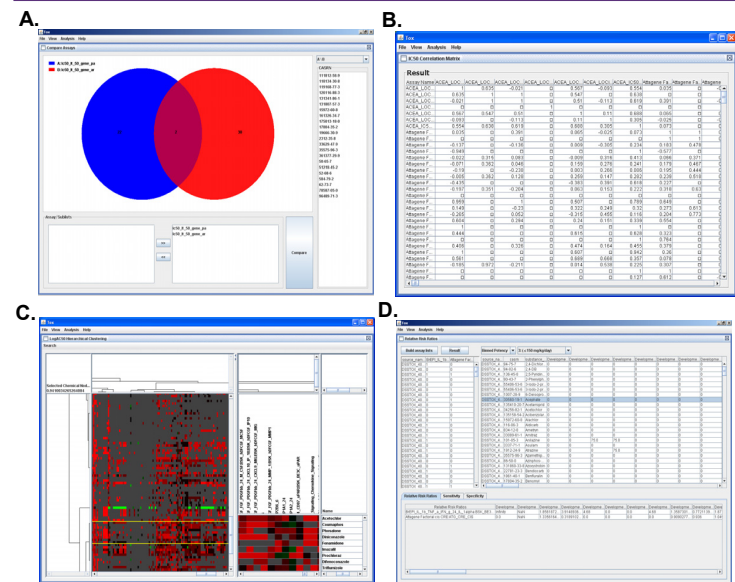
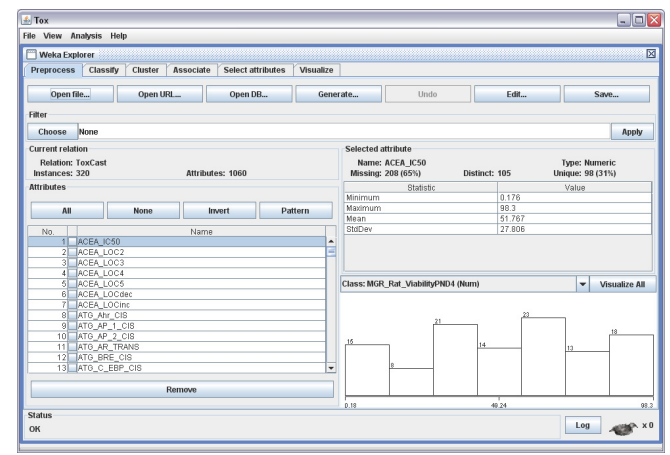


Figure 6: Weka Explorer: Screen shots of the imported ToxCast data. The Weka Explorer implementation provides all the necessary tools for descriptive and discovery analyses. Continued work in creating a seamless environment from creating sublists and subsetted data to sophisticated classification algorithms is underway.



Future Directions

Implement additional features increasing the usability and user-base of the tool, including prioritization schemes.

As the ToxMiner DB begins to store efficacy and concentration response data, additional tools and methods will be implemented into the ToxMiner Interface

Further implement sophisticated machine learning and data analysis algorithms using with R and Weka integration.

Make ToxMiner software package publicly available.

References

Ian H. Witten and Eibe Frank (2005) "Data Mining: Practical machine learning tools and techniques", 2nd Edition, Morgan Kaufmann, San Francisco, 2005.