

# Predictive Toxicology: Applied Semantics with a major impact on drug safety

Erich A. Gombocz, Robert A. Stanley

IO Informatics, Inc., 2550 Ninth Street, Suite 114, Berkeley, CA 94710, USA. [Correspondence: [egombocz@io-informatics.com](mailto:egombocz@io-informatics.com)]

## Summary

While predictive toxicology has been a major goal in the development of safer drugs for quite a while, its promises have – despite major advances in analytics – been challenged by the complexity to better understand the biological functions involved. Specifically, while many of the observed changes in metabolites, gene expression and tissue analytics can result from the same treatment, they most likely represent very different biological processes. In the liver, for example, metabolites might reflect changes in the urea cycle and intermediary metabolism, while gene expression may quantify alterations in the immune response and signaling pathways. As such, despite being pharmacodynamically correlated, the effects may not be directly functionally linked within the biological system. Thus, for biomarker discovery in predictive toxicology to be successful, it is mandatory to move from the bare collection of observed statistical correlations towards an integrated relationship network, which helps differentiating the effects, and by doing so understanding the biological functions involved.

This report describes a semantic systems biology approach to predictive toxicology which fulfills this requirement. One example is of toxicity-induced perturbations in metabolites and genes in studies with different experimental models. The effectiveness of integrating data and their relationships in a common ontological framework is demonstrated. Being able to merge such observations with public resources across taxonomies and semantic standards, an integrated network visualization and analysis can reveal the simultaneous impact on several pathways independent of the experimental layout and qualify common biomarker panels which otherwise would go undetected.

The insights gained from semantically joining coherent findings despite their different methodologies allow researchers to better understand how the organism responds to experiment-induced system changes and to recognize mechanistic aspects of toxicity biomarkers at a functional level. The resulting combinatorial marker profiles can then be refined, validated and applied to predictive screening of new drugs for toxicity.

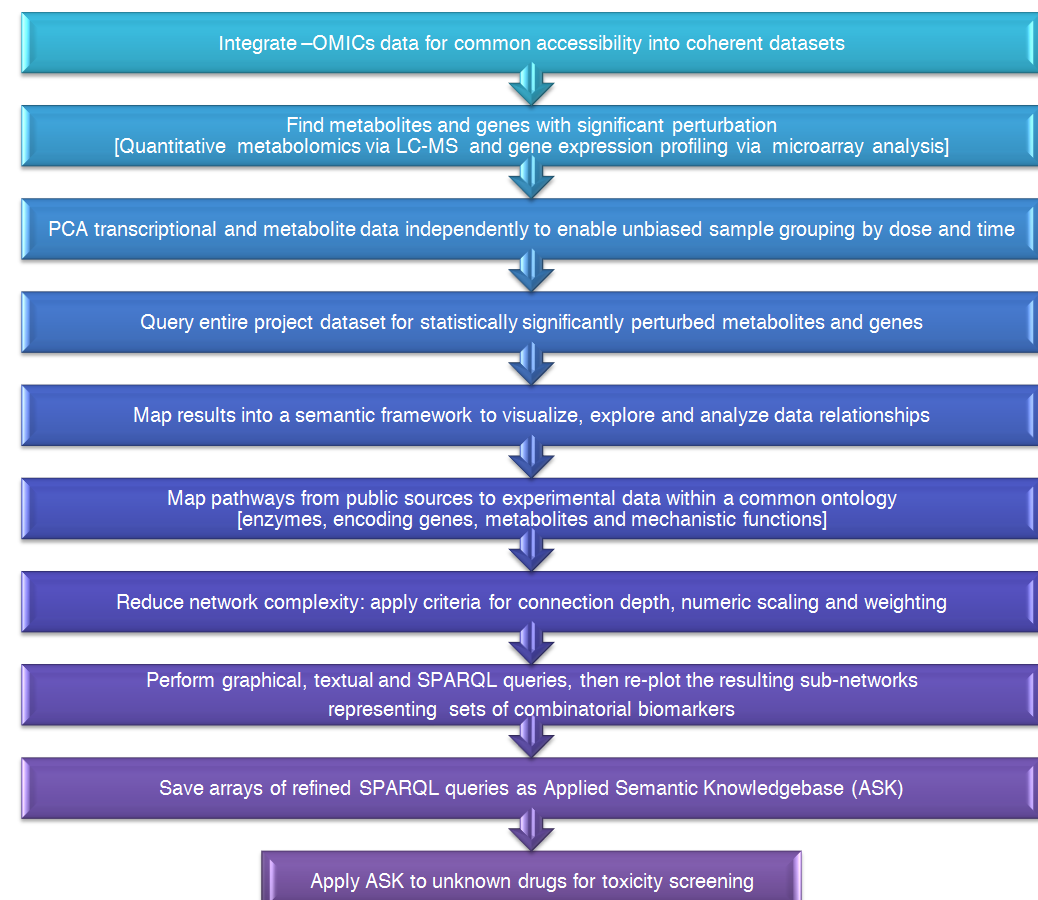
## Challenges

- Biological system networks and pharmacodynamic correlations are not necessarily functionally linked.
- Understanding biological systems to predict phenotypic outcome on toxic responses is still very incomplete and in its infancy.
- In many cases, data relationships are not *a priori* contained in the data sets.
- Data coherence across different taxonomies, ontologies, non-standardized vocabularies is challenging.
- Data type-specific approaches and pre-defined taxonomies are not generally applicable across experimental and public domain data resources.
- Complexity of network analysis in general and lack of intuitive, science-driven tools makes such approaches non-appealing to researchers.
- Scalability and performance of RDF/semantics-based knowledgebases may restrict query, inference and reasoning across the entire dataset.

## Experimental Models

- Panel of several hepatotoxicants, single oral dose (placebo, low, mid, high) in groups of 4 rats, at 6, 24 and 48 hrs.
- Metabolomic analysis of liver, serum and urine (1,603 metabolites).
- Microarray analysis of liver and whole blood (31,096 transcript probes).

## Methodology



## Results & Discussion

- A single observation is most likely the result of many different and intertwined biological responses.
- Merging multi-modal datasets into a common ontology knowledgebase and association with canonical reference sources provides insights in processes involved in toxicity on the organism level.
- The semantic approach overcomes most obstacles for meaningful data integration by applying import mapping, thesauri for synonyms or naming inconsistencies, and ontology merging across classes and hierarchies.
- Being able to take advantage of existing ontologies (such as GO, Pathway Ontology), the analysis of the resulting network of multi-OMICs experiments combined with public reference resources and pathway information enabled us to visualize, query and qualify biomarker profiles with regard to their systems biological relevancy for predictive toxicology.
- Creating, refining and validating combinatorial biomarker models for toxicity can be done by any domain expert via direct selection of network nodes on the graph to automatically generate a visual semantic query (SPARQL) containing the characteristic pattern without programming.
- Arrays of such SPARQL queries representing different types of toxicity are stored as Applied Semantic Knowledgebases (ASK), which are directly applicable as drug safety decision support in screening.

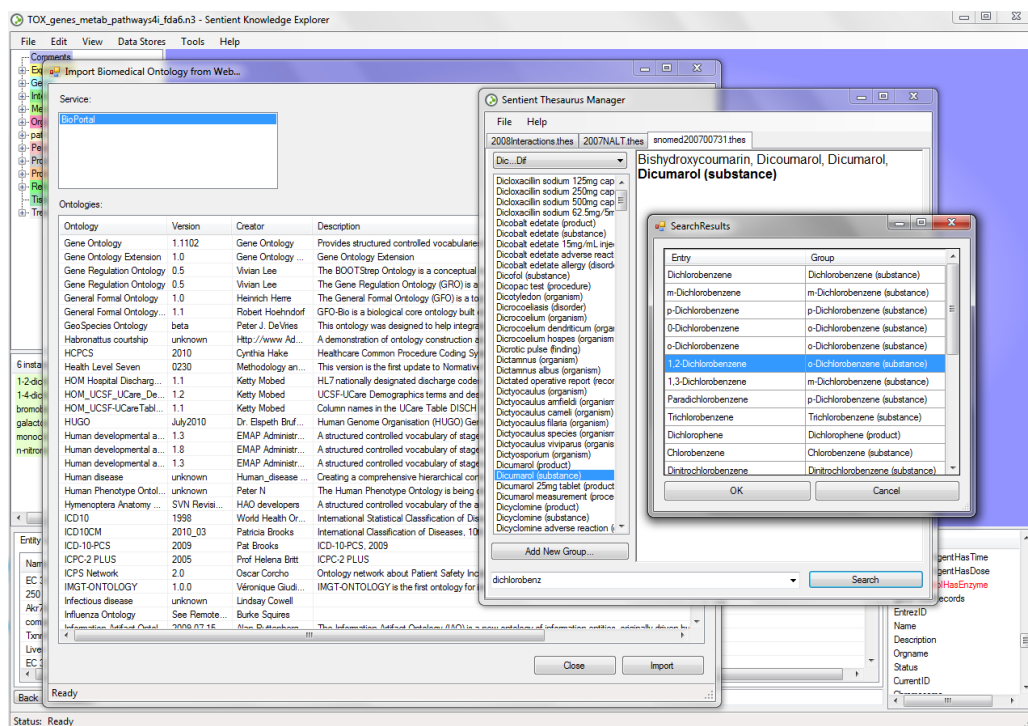


Fig. 1: Data integration framework: Merging ontologies and harmonizing with thesauri

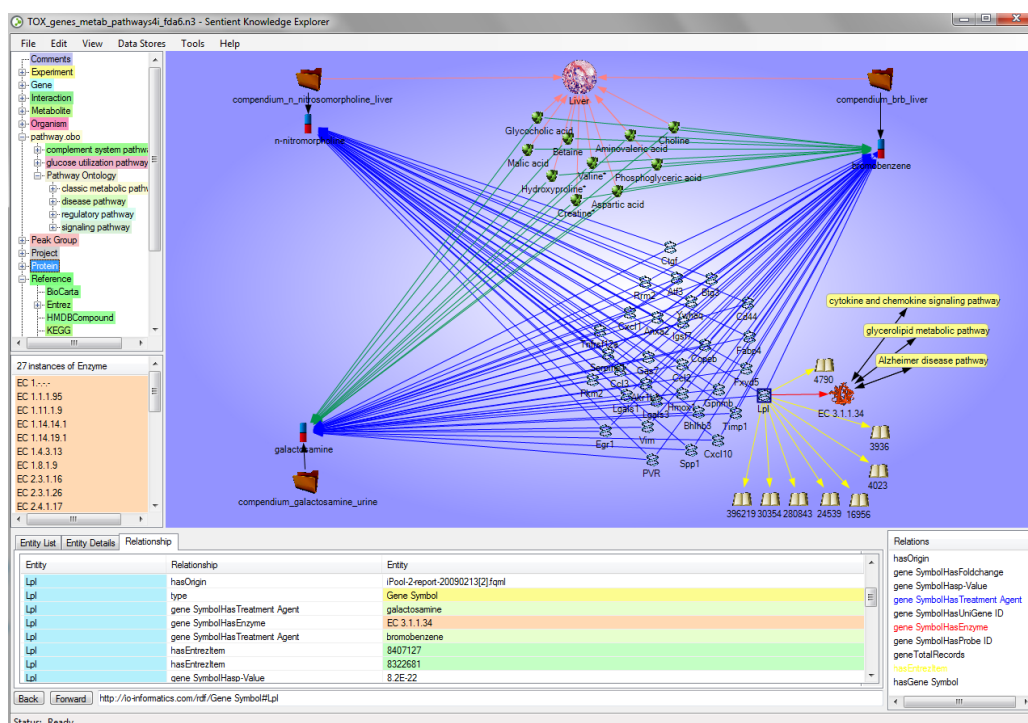


Fig. 2: Ontology merging and incorporation of public resources and pathway information

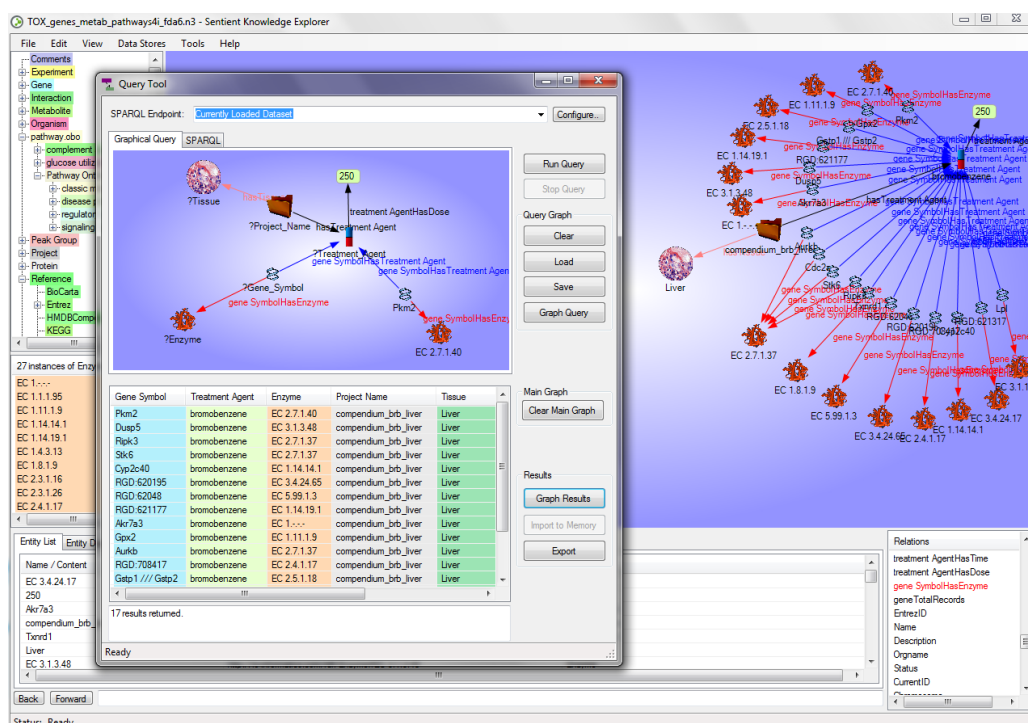


Fig. 3: Graphical SPARQL query: Test, refine and apply profiles for predictive toxicology

## Conclusions

Using a semantic, common ontology framework to explore toxic perturbations in large metabolic and genomic datasets across multiple toxicants and tissues to obtain marker profiles for predictive toxicology, we were able to:

- Correlate metabolites across different treatments to ascertain commonality of effects for a class of drugs.
- Review effects of a single toxicant across tissues to explore commonality of biomarkers in the best tissue for diagnostics.
- Qualify potential combinatorial biomarker profiles for toxicity across tissues and categories of toxicants and gain insights in complex biological functions involving multiple pathway dependencies.
- Validate those toxicity marker profiles on other known toxicants in their response to known common mechanisms of toxicity such as oxidative stress (Glutathione metabolism), liver function (Bile acid and Urea cycle) and Ketoacidosis.
- Refine the combinatorial marker profile with additional pharmacodynamically and biologically linked components and investigate their functional relevancy for inclusion into a broad predictive toxicology marker set.
- Applied Semantic Knowledgebases represent a major advance in combinatorial marker-based prediction of drug safety at a very early stage in the drug discovery and development process and facilitate informed risk assessment for go/no-go decisions based on biologically relevant response profiles.

## Acknowledgements

This work was conducted under NIST Advanced Technology Program (ATP), Award # 70NANB2H3009 as a Joint Venture between Icoria / Cogenics (Division of CLDA) and IO Informatics. Microarray studies were conducted under NIEHS contract # N01-ES-65406. This work was also made possible through contributions from members of IO Informatics' Working Group on "Semantic Applications for Translational Research."