From observations to understanding: A semantic systems biology approach for biomarker discovery in translational medicine

Erich A. Gomboczˢ[1)], Robert A. Stanley[1)], Toshiro Nishimura[1)], Alan J. Higgins[2)], Pat Hurban[3)]

[1] IO Informatics, 2550 Ninth Street, Suite 114, Berkeley, CA 94710, U.S.A. [Correspondence: egombocz@io-informatics.com]
[2)] Viamet Pharmaceuticals, Durham, NC 27713, U.S.A.
[3)] Cogenics, Division of Clinical Data, Morrisville, NC 27560, U.S.A.

Summary

While many of the observed changes in metabolites, gene expression and tissue analytics can result from the same treatment, most likely they 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 translational medicine to be successful, it is mandatory to move from the bare collection of observed statistical correlations towards an integrated relationship network which helps to differentiate the effects and by doing so to understand the underlying biology.

This report describes a semantic systems biology approach which fulfills this requirement. On examples of toxicity 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 joining coherent findings despite their different methodologies allow researchers to better understand how the organism responds to experiment-induced perturbations and to recognize mechanistic aspects of toxicity biomarkers at a systems biology level.

Challenges

- Multi-OMICs expression changes – despite all resulting from the same toxic insult – can represent very different biological processes and can exhibit the net of multiple overlapping alterations.
- Biological system's networks and pharmacodynamic correlations are not necessarily functionally linked.
- Understanding of 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
- Complexity of network analysis in general and lack of intuitive, science-driven tools makes such approaches non-appealing to researchers.
- Scalability and performance issues confine most network approaches to relatively small datasets which restricts query, inference and reasoning across the entire datasets.

Experimental Models

Hepatotoxicity study

- 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 (1603 metabolites).
- Microarray analysis of liver and whole blood (31096 transcript probes).

Alcohol study

- High doses t.i.d. for four days, with and without 24h withdrawal
- Metabolic analysis of plasma, liver and brain
- Microarray analysis of liver and brain

Methodology

- Identify metabolites and genes with significant perturbation (LC-MS and GEP analysis).
- Select robust correlations between independent analytical results (PCA to enable unbiased sample grouping by dose and time)
- Map results into a semantic framework to visualize, investigate and analyze data relationships.
- Associate significant elements of those networks with reference data sources, using thesauri to consolidate data class and relationship synonyms, and combine experimental data with literature
- Map pathway enzymes from public sources to experimental data and merge into a common ontology network across taxonomies and standards
- Use import mapping and thesauri to consolidate data classes and synonyms.
- Apply criteria for connection depth, numeric scaling and weighting to reduce network complexity fro visualization.
- Perform graphical, textual and SPARQL queries to specify conditions (such as, time / dose), then re-plot the resulting sub-networks to qualify potential biomarker panels.
- Output the resulting knowledgebase to RDF, N3 or triples store backend for inference, reasoning and iterative improvements of the underlying model.

Results & Discussion

- A single observation is most likely the result of many different and intertwined biological responses.
- Merging of multi-modal datasets into a common ontology knowledgebase and association with canonical reference sources provides insights in complex processes on the organism level
- In a semantically integrated relationship model, discovery and qualification of biomarkers are based on experimental evidence in context with systems biology networks.
- Reduction to relevant sub-networks provides insights into complex biological responses, allowing qualifying hidden or previously unknown relationships.
- Using network analysis, we can detect commonalities in biological responses even from datasets obtained via different experimental models or study designs.

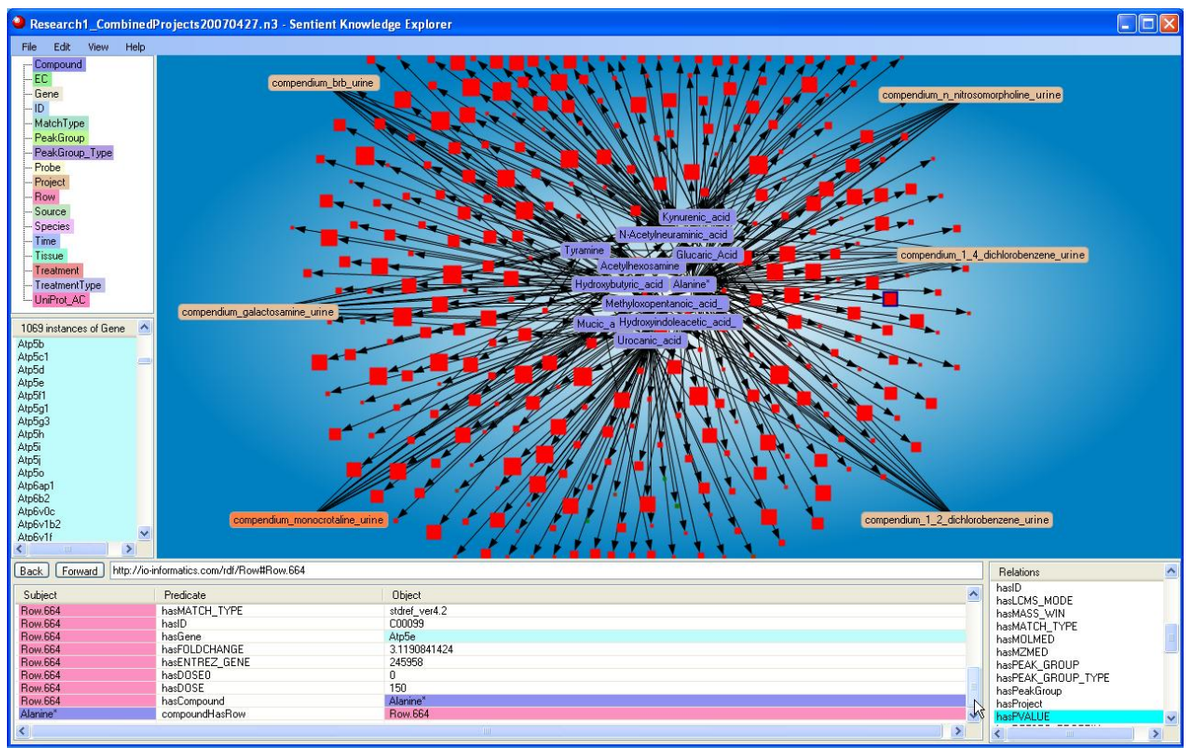

Fig.1: Merged multi-OMICs projects (Hepatotoxicity study): Commonly perturbed metabolites (center, violet) across 6 different hepatotoxicants and their experiments scaled by dose (area) and fold change (color) in Sentient Knowledge Explorer™. Gene expression data and metabolic enzymes have been mapped to the knowledgebase.

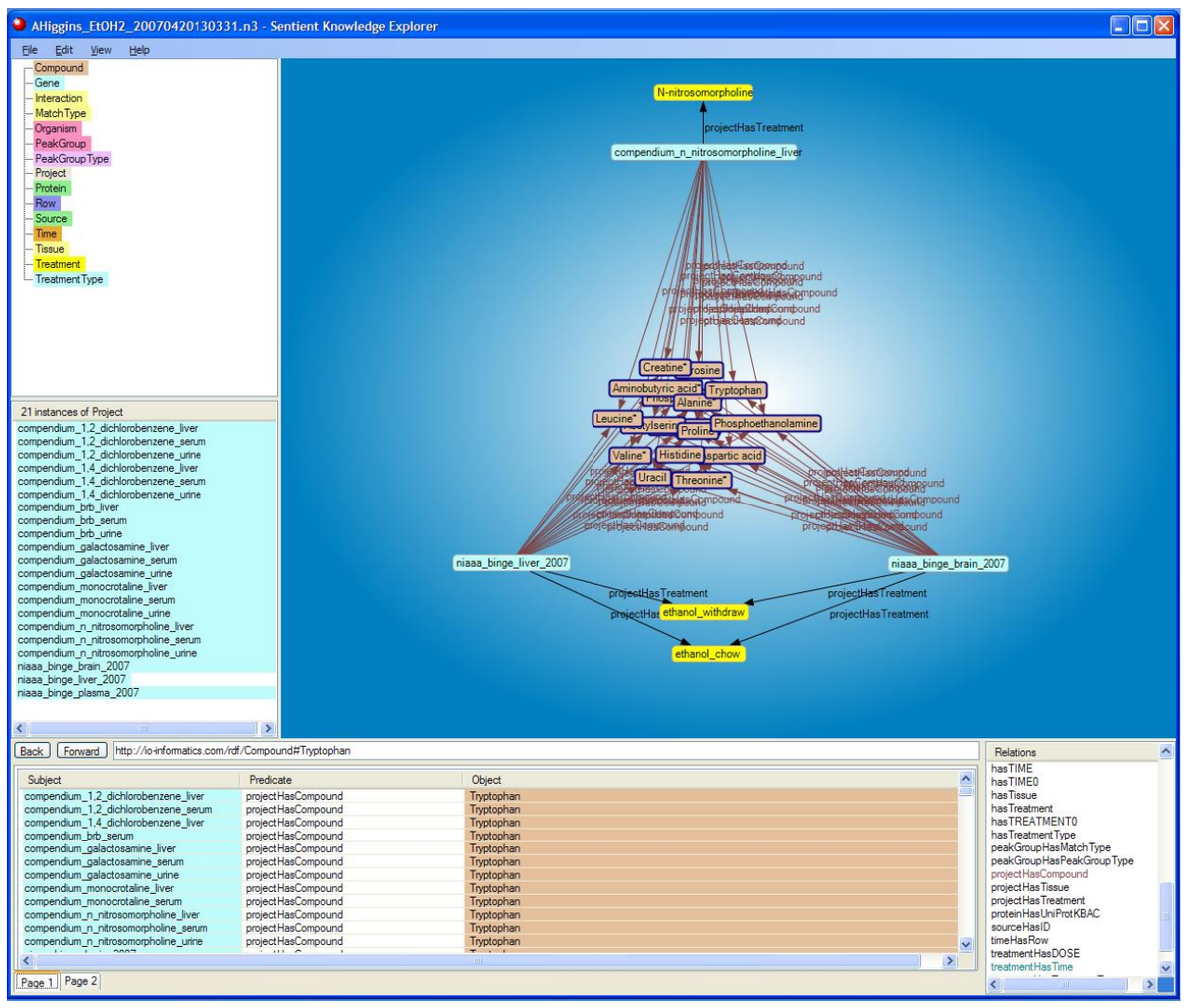

Fig.2): Merging across different experimental models: The combination of Hepatotoxicity and Alcohol studies reveals similarity in metabolic responses indicating involvement of related biological function (even across different tissues).

Conclusions

Using a semantic, common ontology framework to explore toxic perturbations in large metabolic and genomic datasets across several toxicants and different tissues, 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 biomarker panels for toxicity across tissues and toxicants and gain insights in complex biological functions involving multiple pathway dependencies.
- Validate changes in biomarkers associated with known common mechanisms of toxicity such as oxidative stress (Glutathion metabolism), liver function (Bile acid and Urea cycle) and Ketoacidosis
- Discover new pharmacodynamically and biologically linked components and investigate their functional relevancy.

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**. The Alcohol study was conducted under **NIAAA** contract # **HHSN281200510008C**.