

Discovery and Validation of Toxicity Biomarkers Using Correlation Networks and Canonical Pathways

Erich A. Gombocz¹, Alan J. Higgins²

¹Dr. Erich A. Gombocz, Chief Scientific Officer, IO Informatics, Emeryville, CA 94608 [Correspondence: egombocz@io-informatics.com]

²Alan J. Higgins, PhD, Senior Director, Translational Medicine, Cogenics (Division of Clinical Data), Morrisville, NC 27560

Summary

This case study describes how experimental data from multiple sources is analyzed and visualized to detect coherent responses to experimental conditions as correlation networks connecting different ('omics) analytical modalities. Such network transformation provide graph visualization that describe and link potential biomarkers according to their coherent relative responses to perturbation – e.g. as “coherent response modules” - in a systems biology environment. The thesis is that biomarker activity for toxicity models are most usefully investigated first according to their correlations across experimental data methods. Once robust correlations are found, associating significant elements of these networks with reference data sources - e.g. with canonical pathway or reaction databases – allows the researcher to better understand how experimental perturbations are affecting the organism.

Changes in biochemical metabolites and gene products may both result from the same toxic insult and contribute equally importantly to the eventual phenotype, but may represent very different biological processes. For example, in the liver, metabolites might measure changes in the urea cycle and intermediary metabolism, while gene expression may pick up alterations in the immune response and signaling pathways. These responses may therefore be pharmacodynamically correlated but not directly functionally linked within the biochemical network. We have therefore grouped these correlated perturbations into coherent response modules as an initial step to search for panels of biomarkers that reflect the biological stimulus and predict the phenotypic outcome.

For novel biomarker discovery, the ability to create correlation networks representing coherent responses across systems-oriented experimental data may be more helpful than applying that data to canonical pathway networks. However, the ability to link these coherent response networks to canonical functional networks and models (such as KEGG, BIGG, Reactome, IntAct, BioGrid, HMDB) allows researchers to see where / how these correlations may be functionally linked. The ability to link coherent statistical responses to canonical functional networks provides a very useful tool for understanding toxicity biomarkers and more generally for bridging analytical research to various areas for biological systems understanding.

Challenges

- Metabolic and gene expression changes may result from same toxic insult, but represent very different biological processes
- Pharmacodynamic correlations are not necessarily functionally linked within the biochemical network
- Biological systems understanding to predict phenotypic outcome on toxic responses is still very incomplete

Methodology

- Select robust correlations between independent analytical results
- Associate significant elements of those networks with reference data sources
- Create correlation networks from coherent response modules to bridge analytical research towards better understanding of biological systems

Experimental Model

- Panel of chemical hepatotoxicants, single oral dose (placebo, low, mid, high) in groups of 4 rats, sacrificed at 6, 24 and 48 hrs.
- Metabolomic analysis of liver, serum and urine (1603 metabolic components); microarray analysis of liver and whole blood (31096 transcript probes).

Analysis

- Find compounds and genes with significant perturbation (LC-MS and GEP analysis)
- PCA of transcriptional and metabolite data independently, enabling unbiased sample grouping by dose and time
- Map results into a semantic framework to visualize, explore and analyze data relationships
- Map pathway enzymes from public sources to experimental data within a common ontology
- Reduce network complexity: apply criteria for connection depth, numeric scaling and weighting
- Perform graphical and SPARQL queries, then re-plot the resulting sub-networks

Results

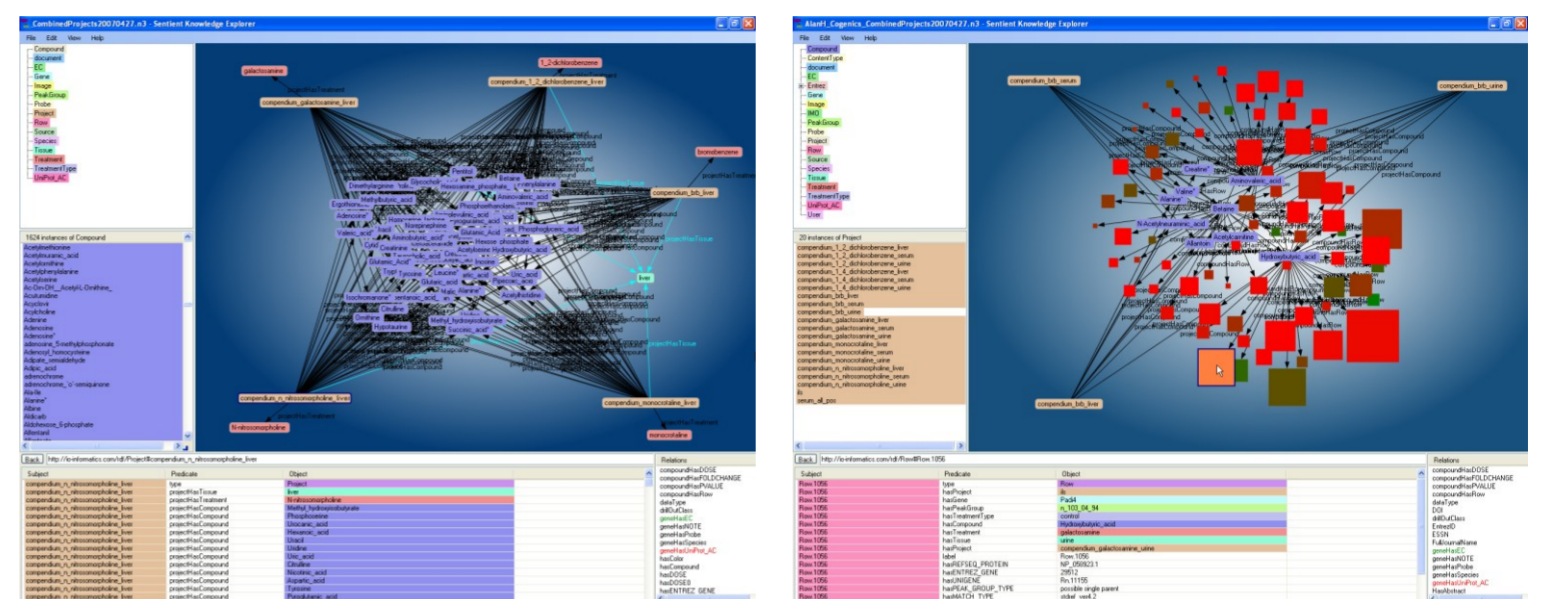


Fig.1 (left): Correlate commonly perturbed compounds across studies with same tissue, but different toxicants: 5 liver toxicity studies (at the corners of the network graph) and their effected metabolites (in the center, violet colored) are easily distinguished from thousands of compounds.

Fig.2 (right): Validate metabolic responses across tissues: Visualize significance of experimental data by scaling compounds to fold changes (represented by box area) and p-value (represented by color)

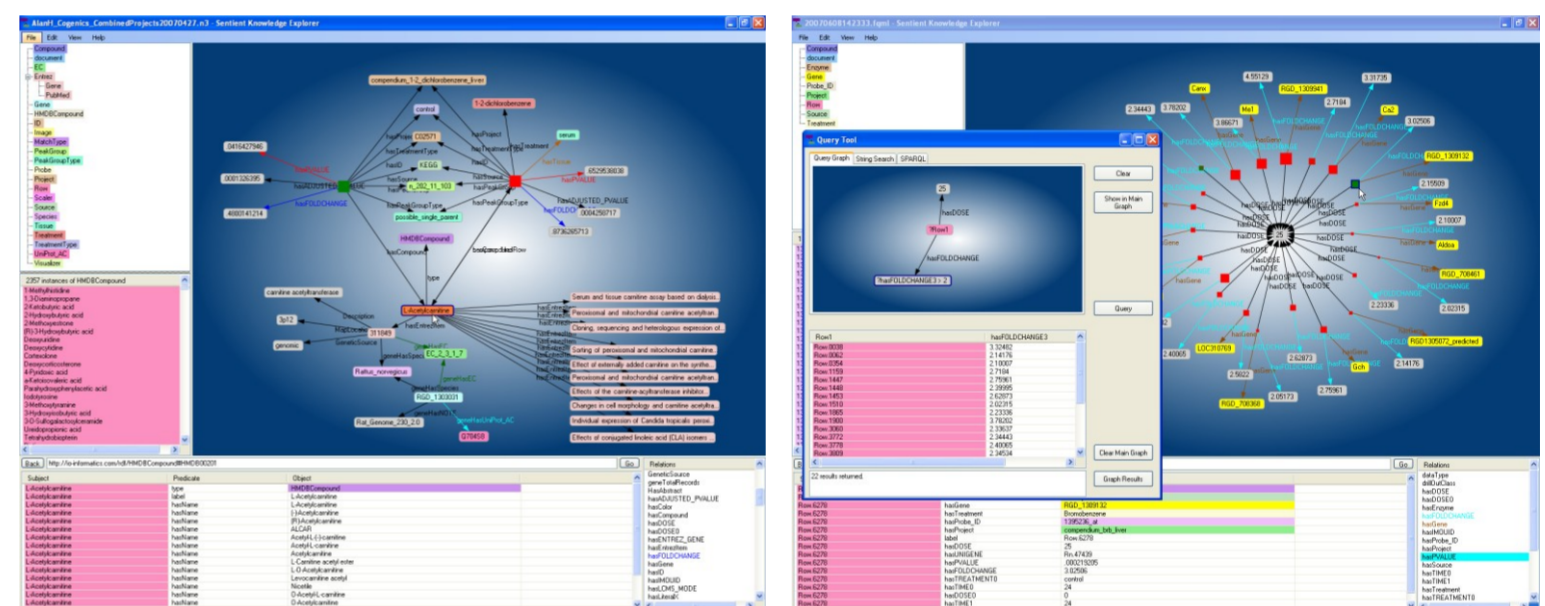


Fig.3 (left): Correlate metabolites and genes: L-Acetylcarnitine network, 1-2-Dichlorobenzene as toxicant

Fig.4 (right): Query by time/dose: Reveal all toxicant-induced gene changes for a single dose, single time point

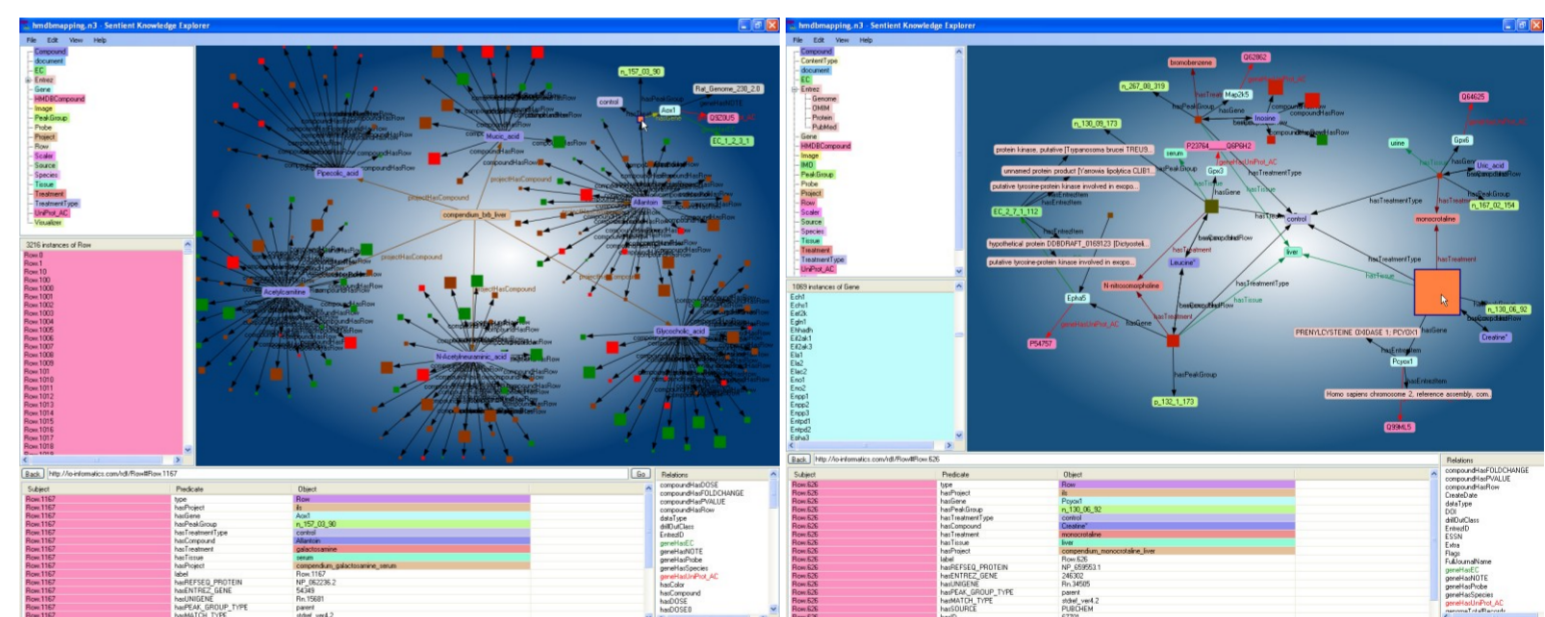


Fig.5 (left): Dose/time classification of experiments for multiple metabolites: Metabolite and gene responses

Fig.6 (right): Map of pathway enzymes and genes from public sources: Contextualize analytical observations

Conclusions

Taking complex, large datasets to explore toxic perturbations across several toxicants and different tissues, we were able to demonstrate the following abilities:

- Correlate metabolites across different drug treatments – discover commonality of effects
- Review effects of a single toxicant across tissues (e.g. liver, serum, urine) – explore commonality of biomarkers in a more accessible tissue for diagnostics
- Correlate genes and metabolites in the same tissue – understand pathway- and systems biology-related interactions
- Analyze time/dose dependencies and effects to optimize experimental conditions for consecutive studies

Acknowledgements

This work was conducted under NIST Advanced Technology Program (ATP), Award # 70NANB2H3009. Microarray studies were conducted under NIEHS contract. The following individuals contributed to this study: A. Heinloth, G. Boorman, R. Paules (NIEHS), I. Shah, H. Mei, K-F. Wang (Computational Biology, Cogenics), K. Powell, J. Corbell (Metabolomics, Cogenics), P. Hurban, E. Lobenhofer (Microarrays, Cogenics), R. Stanley, M. Travers, O. Kent (IO Informatics).