

# “Does network analysis of integrated data help understanding how alcohol affects biological functions?”

## - Results of a semantic approach to biomarker discovery

Erich A. Gombocz<sup>1)</sup>, Alan J. Higgins<sup>2)</sup>, Pat Hurban<sup>3)</sup>, Edward K. Lobenhofer<sup>3)</sup>, Fulton T. Crews<sup>4)</sup>, Robert A. Stanley<sup>1)</sup>, Chuck Rockey<sup>1)</sup>, Toshiro Nishimura<sup>1)</sup>

<sup>1)</sup> IO Informatics, Berkeley, CA 94710, U.S.A. [Correspondence: [egombocz@io-informatics.com](mailto:egombocz@io-informatics.com)]

<sup>2)</sup> Viamet Pharmaceuticals, Durham, NC 27713, U.S.A.

<sup>3)</sup> Cogenics, A Division of Clinical Data, Morrisville, NC 27560, U.S.A.

<sup>4)</sup> University of North Carolina. Bowles Center for Alcohol Studies (CAS), Chapel Hill, NC 27599, USA.

### Summary

Investigating ethanol toxicity in different rat models and using the combined analysis of complex perturbations in metabolites and genes in studies with different experimental models as example, 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.

Although functional understanding is still in its infancy, joining coherent findings despite their different methodologies allow researchers to gain insight how the organism responds to experiment-induced perturbations and to recognize mechanistic aspects of toxicity biomarkers at a systems biology level. The results of such an approach towards understanding which biological functions are affected by alcohol are presented and discussed.

### Challenges

- Metabolic and gene expression changes may result from same toxic insult despite representing very different biological processes.
- Pharmacodynamic correlations and functions in biological systems networks are not directly related.
- Complexity and lack of intuitive, science-driven tools for network analysis makes systems biology approaches non-appealing to researchers.
- Multiple, incompatible semantic standards make merging towards a common ontology extremely difficult; this also impacts query and reasoning.
- Conventional pathway boundaries do not account for overlaps and involvement of the same enzyme in multiple pathways
- Scalability and performance issues confine most network approaches to relatively small datasets.
- Biological understanding of complex processes is in its infancy.

### Experimental Models

1. Chronic administration
  - Tsukamoto-French model
  - Continuous infusion via gastric cannula
  - Adjust dose according to behavioural scores
  - Concomitant fat and vitamin supplementation in both treated and controls
  - Liver, brain, serum, serial urine samples for metabolomics (1603 metabolites, LC-MS)
2. Binge model
  - Four days t.i.d., oral gavage ± 24h withdrawal/fasting
  - Adjust dose according to behavioural scores
  - Water controls
  - Liver, brain, plasma samples for metabolomics (1603 metabolites, LC-MS)
  - Liver and brain samples for gene expression profiling (31096 transcript probes, MA)

### Process

- Identify metabolites and genes with significant perturbation (LC-MS and GEP analysis).
- Select robust correlations between independent analytical results
- Map results into a semantic framework for visualization and data relationship analysis.
- Associate significant elements of those networks with reference data sources.
- Use thesauri to consolidate synonyms, and combine experimental data with public literature.
- Map pathway enzymes from public sources to experimental data and merge into a common ontology network across taxonomies and class hierarchies.
- Reduce network complexity for visualization through weighting and connection depth criteria.
- Perform graphical, textual and SPARQL queries to specify multi-parametric conditions (such as time/dose or tissue dependencies).
- Re-plot the resulting sub-networks to qualify potential biomarker panels and their pathway involvement.
- Output the knowledgebase to a triples store backend for inference, reasoning and iterative improvements of the underlying model to refine understanding of biological processes.

### Results

Correlating similarities in metabolic and genetic response via their relationships to common pathway enzymes is the key to better decoding of complex biological functions. New discoveries in the One-Carbon metabolism and the involvement of signaling pathways indicating influence on long-term memory storage in brain are under further investigation.

#### Brain

- Changes are more conserved than in liver and plasma
- Most changes occur during withdrawal
- No major changes take place in urea cycle or purine pathway.

#### Liver

- Signs of alcoholic ketoacidosis
- Depletion of biogenic amine precursors could relate to alcohol dependency and withdrawal
- Major changes in purine metabolism suggest inhibition of xanthine oxidase by oxidative stress.

#### Plasma

- Changes in biogenic amine precursors rebound during withdrawal
- Selective depletion of cytosine and cytidine vs. thymidine

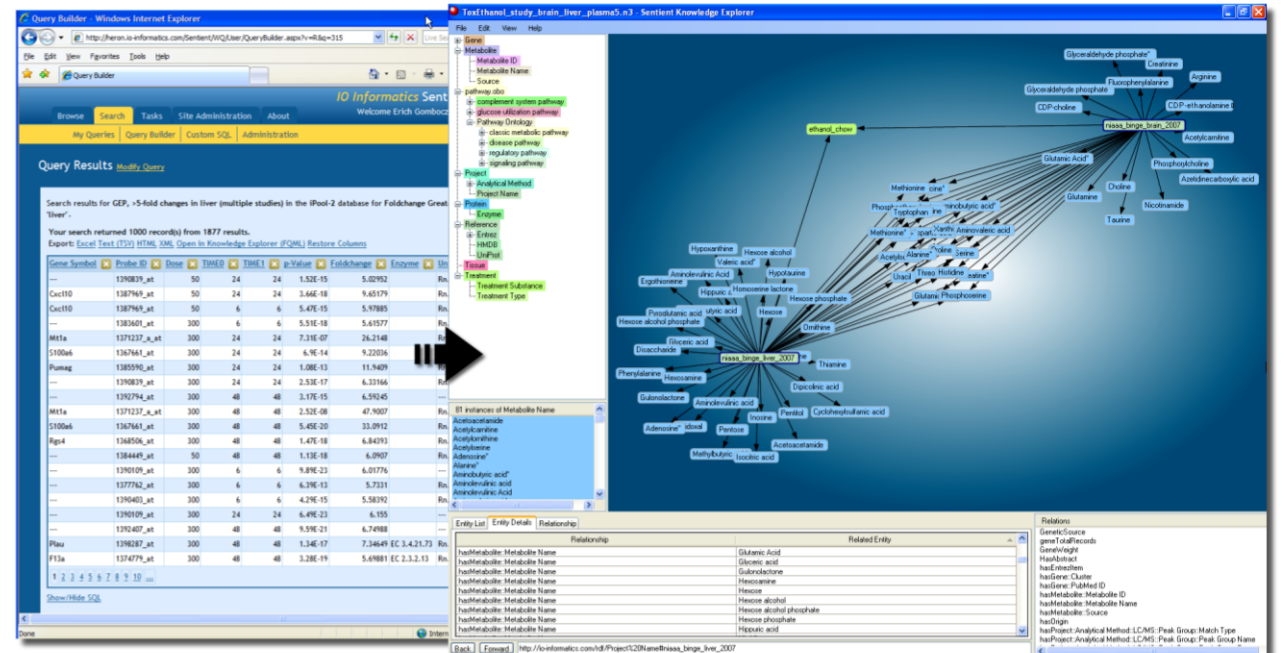


Fig.1: Semantic data merging: Query gene expression (left), merging with metabolic changes into a common ontology in Sentient Knowledge Explorer™. Metabolic changes across liver and brain are easily revealed (right).

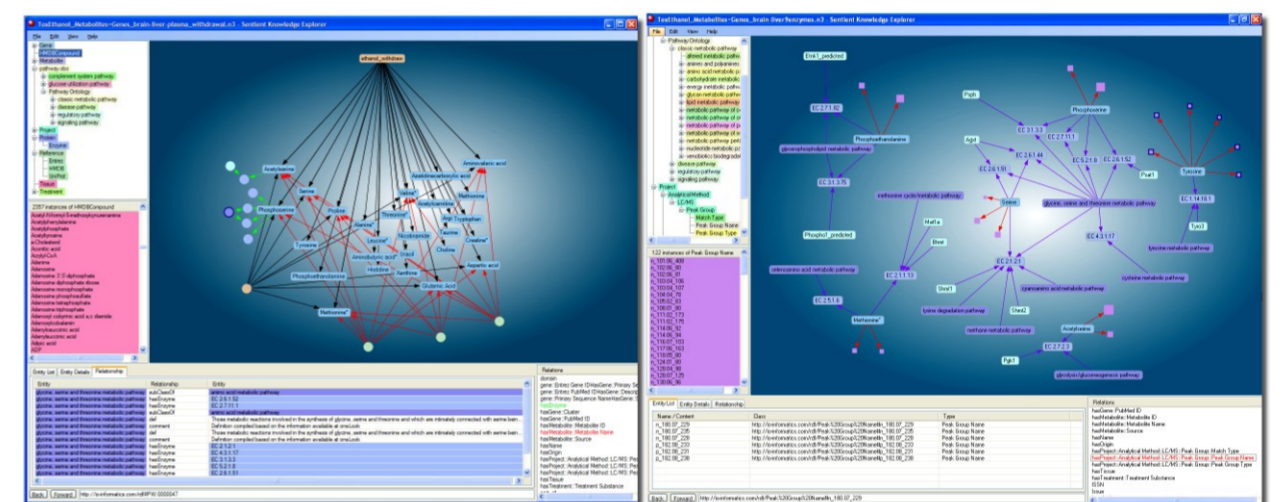


Fig. 2: Analyzing withdrawal: Sub-network of perturbed metabolites and enzymes in brain, liver and plasma (left). Metabolites and gene mapped to their Pathway enzyme links to affected genes and metabolites (left, square areas scaled to fold-changes).

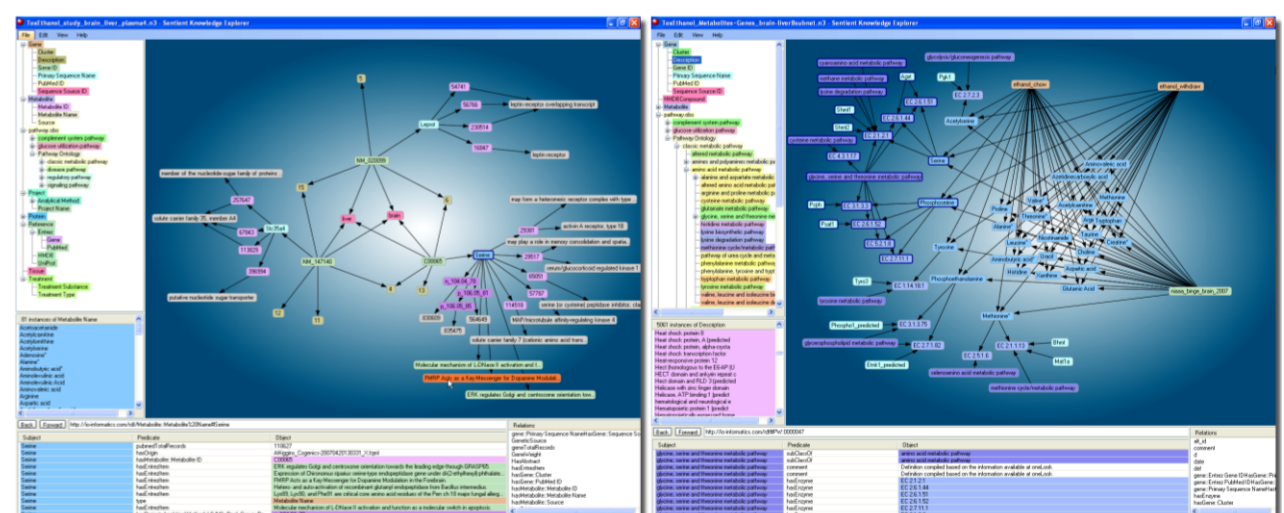


Fig.3: Integrating references: Drill out to HMDB and NCBI's Entrez databases (left). Mapping common pathway enzymes to NCBO's BioPortal Pathway ontology reveals all pathways for each enzyme involved (right).

### Conclusions

Using a semantic approach to biomarker discovery, toxic perturbations were analyzed for commonalities in responses at the transcription and metabolic level.

- Mapping of significantly changed genes and metabolic changes to pathway enzymes involved in both responses provides meaningful insights in the underlying biology of Ethanol exposure.
- Non-obvious correlations in toxic responses across different studies evidence similar biological responses.
- Need to view metabolism as an entire network as boundaries between pathways are misleading.
- Semantic data integration methods offer means for discovery of biological similarities across different experimental models.
- Discovery of new pharmacodynamically and biologically linked components provides clues to downstream interactions involving multiple pathways.

“Network analysis of integrated data helps understanding biological functions, but we still have a lot to learn.”

### Acknowledgements

Broad toxicity biomarker studies were 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 in collaboration with Bowles Center for Alcohol Studies (CAS) at UNC. This work was also made possible through contributions from members of IO Informatics' Working Group on “Semantic Applications for Translational Research”.