From Correlation to Biological Understanding: Multi-modal semantic networks for biomarker discovery and qualification

Alan J. Higgins 1)  
Erich A. Gombocz 2)  
Robert A. Stanley 2)

1) Viamet Pharmaceuticals, Durham, NC 27713, U.S.A.  
2) IO Informatics, Berkeley, CA 94710, U.S.A.
Why a semantic approach?

- Experimental observations from multi-modal –OMICs data may result from the same perturbation, but represent very different biological processes.
- Pharmacodynamic correlations are not necessarily functionally linked within the biological network.
- Genetic and metabolic changes may occur at lower doses and earlier than those that produce pathological changes.
- Data need to be examined in context of their relationships to be able to infer and reason how experimental perturbations affect the organism.
Challenges

Use of multi-modal experimental datasets (even if normalized) requires:

- Data coherence
- Merging across different taxonomies using a controlled terminology for data classes and their relationships
- Integration of conflicting data standards and mapping of relational database query output into a common ontology
- Ability to link-in and drill-out to public reference sources
- Intuitive tools for network visualization, query, inference and reasoning
Integrate -OMICs data for common accessibility into coherent datasets

Find compounds and genes with significant perturbation [LC-MS and GEP MA analysis]

PCA transcriptional and metabolite data independently to enable unbiased sample grouping by dose and time

Query entire project dataset for statistically significantly perturbed metabolites and genes

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, textual and SPARQL queries, then re-plot the resulting sub-networks

Qualified biomarker panels
Hepatotoxicity Study

- Panel of hepatotoxins, 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 metabolic components); 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
Toxicity biomarker panel for 4 different known hepatotoxicants

"is there a common toxic response in liver metabolites?"
Result

- Common toxicity markers for all toxicants have been identified, reviewed and qualified via public domain references.
“Fly-through” Example 2: Hepatotoxicity Study

- Common biomarker panel across tissues for a single toxicant

- “do we need a liver biopsy, if we can do a simple urine test?”
Result

- A panel of 8 metabolic markers with similar response for the toxicant across tissues; the assay can be developed in the most accessible tissue
“Fly-through” Example 3: Alcohol Study

- Responses on Ethanol withdrawal after several days of previous exposure

- “what we can learn from drinking rats who want to quit?”
Result

- Effects of Ethanol withdrawal on metabolites after 4 and 5 days of previous exposure; time-independent and time-dependent responses
“Fly-through” Example 4: Hepatotoxicity + Alcohol Study

- Responses across unrelated studies with different experimental models

- “are there common toxicity responses despite different approaches?”
Result

- Non-obvious correlations in toxic responses across different studies evidence similar biological responses
Take-Home Message

 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 allows to discover hidden or unknown relationships

 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
Results

- We were able to:
  1. Correlate metabolites across different drug treatments – discover commonality of effects
  2. Review effects of a single toxicant across tissues – explore commonality of biomarkers in a more accessible compartment for diagnostics
  3. Correlate genes and metabolites in the same tissue – understand pathway- and systems biology-related interactions
  4. Gain insights in drug dependency effects upon withdrawal – analyze time/dose dependencies in response pattern
Qualification & Validation

- We were also able to:
  - 1. Validate changes in biomarkers associated with known common mechanisms of toxicity
    - Oxidative stress (Glutathion metabolism)
    - Liver function (Bile acid and Urea cycle)
    - Ketoacidosis
  - 2. Discover new pharmacodynamically and biologically linked components; their functional relevance is currently under investigation
Conclusions

- An informatics approach which merges multi-modal experimental datasets to establish correlation networks and associates those with canonical reference sources provides insights in complex processes on the organism level.

- In biomarker discovery and qualification, semantically organized knowledge bridges the gap between analytical research observations and understanding of the underlying biological system.
Next Steps

- Inclusion of clinical endpoints and tissue analytics to further enhance biological understanding

- Creation of a semantic knowledgebase for inferencing and reasoning across complex translational research data

- Application to personalized medicine (patient stratification, pre-symptomatic disease detection, responder profiles based on genotypes, drug interaction 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**

The Alcohol study was conducted under **NIAAA contract # HHSN281200510008C**
Thank you for your attention

ahiggins@viamet.com
egombocz@io-informatics.com
rstanley@io-informatics.com

http://www.viamet.com
http://www.io-informatics.com
http://www.io-informatics.com