From multiple biomarkers to patient-centric personalized medicine:
An "Applied Semantic Knowledgebase" for decision support

Erich A. Gombocz[1)], Robert A. Stanley[1)], Jason A. Eshleman[1)], Zack Rhoades[1)]

[1)] IO Informatics, 2550 Ninth Street, Suite 114, Berkeley, CA 94710, U.S.A. [Correspondence: egombocz@io-informatics.com]

Summary

The use of biomarkers from multiple modalities (OMICs, imaging, clinical endpoints) – while increasing in the scientific community –has, for the most part, lagged behind the promise of their use in patient screening as decision support for treatment options. This is partly because of the difficulties of meaningful semantic integration of heterogeneous experimental and public data, and the complexity in understanding the involved biological functions, both of which are essential for predictive biology in the development of safer drugs and more effective therapies in the clinics.

Based on core data access and integration capabilities, Sentient software applies semantic patterns to create, refine and qualify biomarker network models using virtually any combination of internal experimental data and / or external published information. Such patterns apply extended semantic "Visual SPARQL" query technology to build complex searches across multiple information sets, even if the initial datasets are not formally joined under any common database schema or federation method. An Applied Semantic Knowledgebase (ASK) unique to a specific research focus contains arrays of such patterns, thus providing a collection applicable to screening and decision making.

In this poster, we present 3 recent customer examples using ASK to demonstrate its power as decision support tool:
- *the assessment of treatment effectiveness for combinatorial prostate cancer therapies;*
- *the pre-symptomatic detection, scoring and stratification of patients at risk of heart, lung or kidney failure; and*
- *the impact of inflammatory responses in high risk plaque rupture*

The semantic joining of experimental correlation networks enriched with curated public domain knowledge networks helps researchers to gain a better understanding of mechanistic aspects of biomarkers at a functional level. This knowledge is then applied to highly sensitive, specific, scored patient screening – providing a confident decision support for life sciences and personalized medicine.

Challenges

- Data coherence: sources, taxonomies, ontologies, non-standardized vocabularies.
- Complexity in meaningful integration of multi-OMIC data sets.
- Lack of intuitive, science-driven tools for hypothesis and intuitive decision support.
- Biomarker classifier validation with mechanistic and functional insights is demanding.
- Unifying public resources and internal datasets and proper weighing of markers is non-trivial.

Methodology

- Integrate data sets from multiple modalities, providing a unified access point for all data.
- Merge and map results into a semantic framework to visualize, investigate and analyze data relationships.
- Scale potential markers using numerical properties to reduce network complexity and pre-select classifiers.
- Identify statistically significant changes in multiple modalities with robust correlation between independent analytical results via query.
- Validate biological viability of experimental networks through incorporation of mechanistic public knowledge
- Save the resulting sub-network as SPARQL query, and represent your model as an array of such queries.
- Refine each model with test cases; apply it to unknowns for screening and use scoring of the match ("hits-to-fit") for informed decision-making with high confidence.

Approach

- Combining web-based federated query and semantic network approaches to create unified knowledgebases.
- Create an easy way to explore knowledge networks of biological data, identify potential markers and build refinable hypotheses describing complex biological functions as biosystem subnets.
- These hypotheses can be exported to a lightweight web-based application which allows access from any location for screening and decision support.
- Scientists have the ability to apply the hypothesis across existing data sets as well as expand it to new information as it is generated, thereby constantly refining the pattern.

Results & Discussion

- This poster demonstrates an innovative approach towards integration of experimental and public data sources, visual exploration of such an integrated environment, and construction of characteristic biomarker patterns.
- Semantic integration and merging of data assures coherence, harmonizes synonyms and different terminologies, and provides a consistent base to relevant network analysis.
- SPARQL queries are captured visually and saved in arrays representative for a specific biological function. Being able to create complex models in an easy, automated way makes the method universally applicable.
- ASK has been actively applied for decision-support in predictive pre-symptomatic patient screening for organ failure (PROOF[2]), identification of plaque stage prior to rupture in high risk plaque (BG Medicine HRP[3]), and the comparative effectiveness of cancer treatment for prostate cancer (PCC[4]).

Impact & Future

- Applied Semantic Knowledgebases (ASK™) provide researchers in life sciences and clinics, who are faced with complex biological questions and rely daily on decision-support, with a novel, directly applicable approach to address their needs.
- While this technique already is applied in several areas of interests in pharmaceutical research, life sciences and personalized medicine, its function as knowledge application to provide decision support ranges from targets to compounds to patient treatment and screening.
- Actionable inferences about disease states and treatments using combined clinical, genomic, proteomic, metabolic and molecular phenotypic data in conjunction with mechanistic insights from public knowledge networks via a web-based tool is a remarkable step in patient-centric personalized medicine.
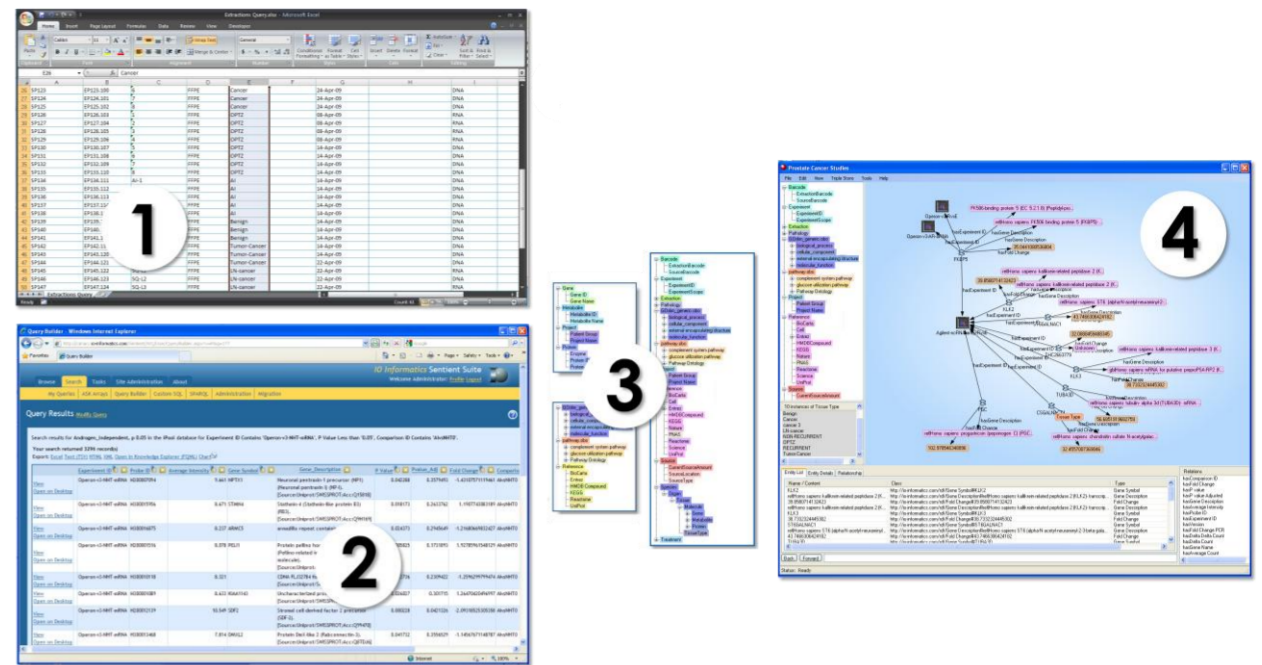
Figures

Fig. 1: Semantic data merging of experimental correlation networks and knowledge networks: From spreadsheets and database query (1,2 left) to ontology merging (3) and a reference-enhanced biomarker network (4, right)
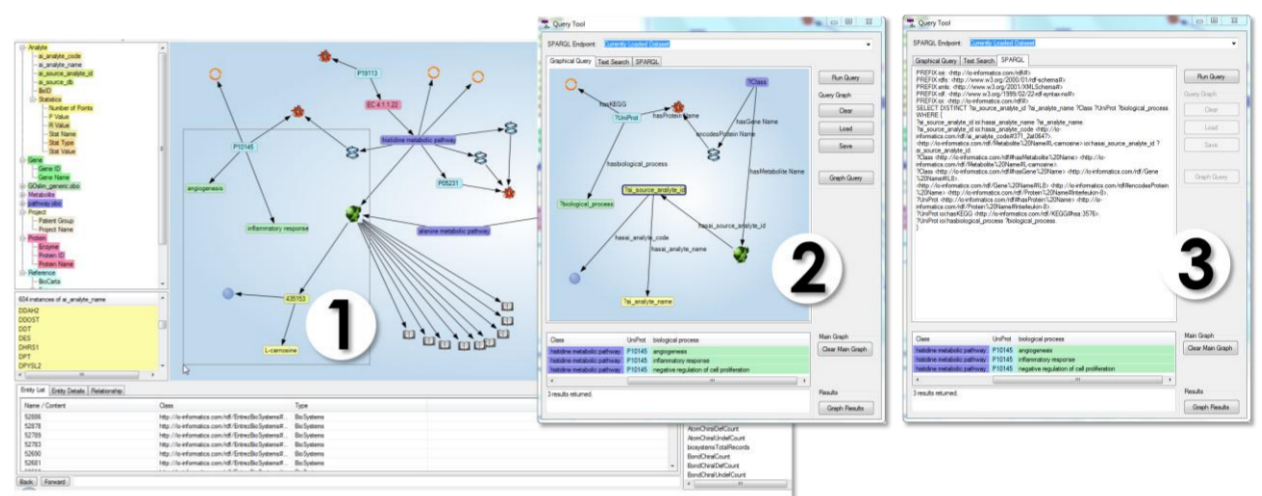

Fig. 2: SPARQL creation from graph: Node selection in the main network (1) generates a visual SPARQL representation of the query(2) and the actual SPARQL statement (3) automatically
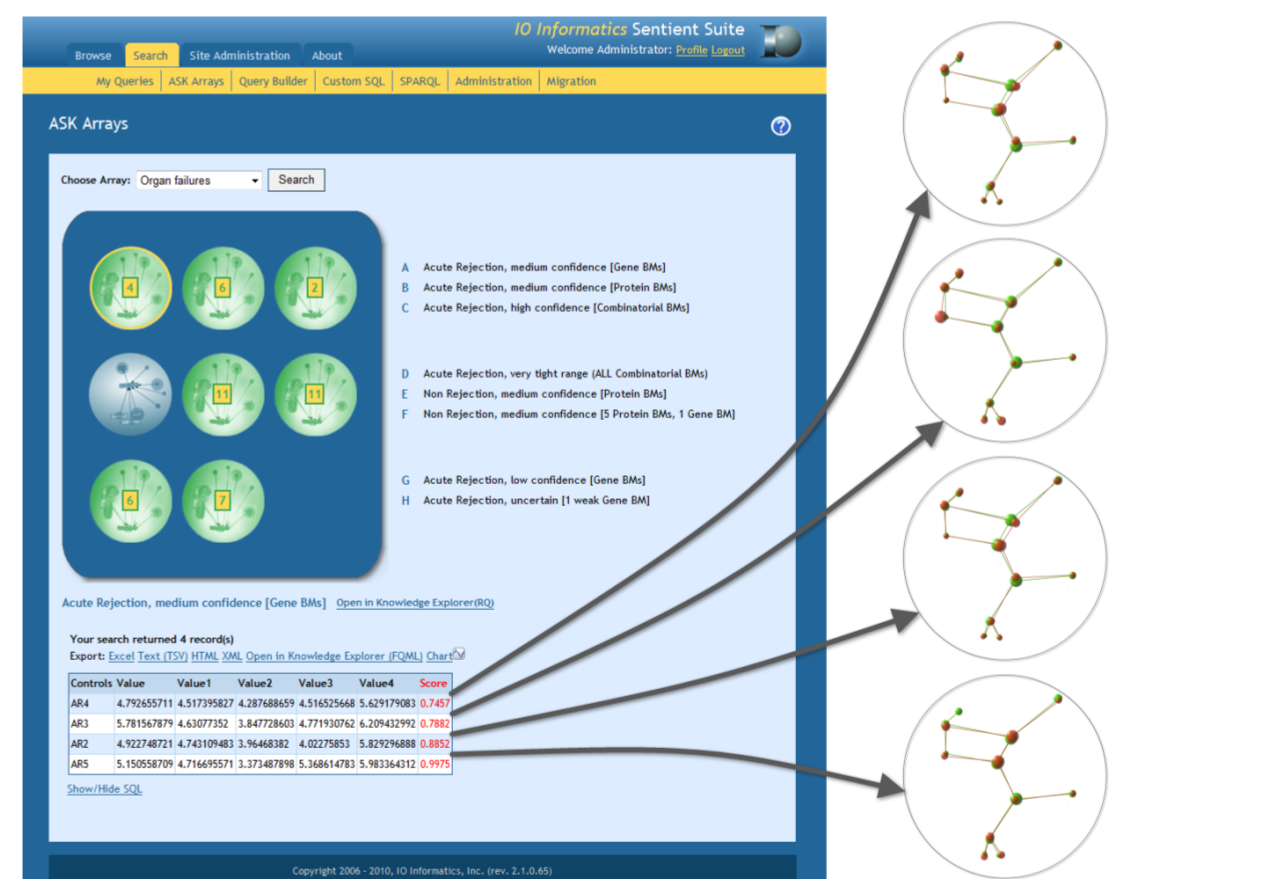

Fig. 3: Web-browser accessible ASK arrays: Predictive screening as decision-support for pre-symptomatic heart transplant failures (left: interface with scoring) and "hit-to-fit" representation for confidence in prediction (right)

Conclusions

- Applying arrays of semantically-based models in an easy-to-use fashion accounts for its appeal to researchers in life sciences and personalized medicine, who are faced with complex biological questions and rely on confident decision-support day-by-day.
- Being able to use, share and apply knowledge based on sophisticated network models via an intuitive web tool which hides the underlying complexity from the user, yet provides concise information which data (targets, compounds, diseases, patients) fit the model, and how good the match is in each particular case, is changing the way how knowledge is built, refined and applied in life sciences and personalized medicine.

Acknowledgements

References

[2)] Prevention of Organ Failure (PROOF) Centre of Excellence, Vancouver, BC, Canada.
[3)] BG Medicine Consortium on High Risk Plaque (HRP), Waltham, MA, U.S.A.
[4)] UBC Prostate Cancer Centre (PCC), Vancouver General Hospital, Vancouver, BC, Canada.