

Personalized Medicine in Action: Screening with Semantic Biology Models from Integrated Data Correlation and Knowledge Networks

Jason Eshleman¹, Erich Gombocz¹, Robert Stanley¹

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

Summary

With personalized medicine research spanning multiple information domains, dynamic, flexible and extensible solutions are required to cope with the demand for knowledge management and data sharing across these disciplines. On top of privacy and security concerns, analytical and demographic data from instruments, files, images, databases and the web need to be put together in context of biological mechanisms and clinical relevance in order to generate applicable knowledge. The integration challenges become even more apparent when addressing complex medical data from hospitals, research laboratories, external facilities and public resources with the goal of generating knowledge which can be actively applied in real-time for patient care in personalized medicine. This requires a new approach in order to harvest data relationships and interactions in a meaningful way and to create applicable models for real-time patient-centric decision making.

This poster demonstrates how semantic data integration, where data are placed in their functional context with other data, changes the playing field. Semantics harmonize synonyms and nomenclature to account for common terminology and data relationships across life sciences research and clinical domains, and allows for relationship mining, inference and reasoning in a systems approach. Dynamic, flexible resource description framework (RDF) databases, ubiquitous SPARQL endpoints and "linked open data" initiatives are becoming important tools and are key for any research demanding global collaboration. Graph representations of merged data, including correlation networks and formally mechanistic reference networks, allow for identifying classifiers in biomarker discovery and focusing on relevant dimension-reduced subnets. Knowledge from powerful systems biology-based models fosters understanding of complex biological processes (e.g., diseases & disease states, pre-disposition to certain biological responses, patient stratification for trials and treatment, predictive risk assessment for tumor growth, organ rejection, and severity of drug side effects).

The poster presents several customer cases - decision support in heart transplant patients; assessment of effectiveness of prostate cancer treatment; development of species-independent disease markers. These examples demonstrate capabilities gained from semantic integration of analytical correlation networks with public knowledge networks towards better understanding of complex biologic interactions relevant to personalized medicine. We show how this technology is actively applied in production environments for 1.) combining and reasoning across experimental and published data to create target profiles identifying effective compounds and warning of toxic indication; and 2.) creating actionable inferences and intelligent screening patterns about disease states and treatments. These methods combine clinical, genomic and molecular phenotypic data within Applied Semantic Knowledgebases (ASK™) to assert patient-centric personalized medicine.

The implications of using Applied Semantic Knowledgebases (ASK™) in personalized therapy, treatment, improvement of quality of life and its impact on healthcare and pharmaceutical development are discussed.

Challenges

- Most relational data sources have specific structures and schemas, and exist only in narrow context.
- Dynamic data require flexible integration structures capable of cross-boundary contextualization.
- Research requirements and routine clinical data collection and assessment are not aligned.
- Biomarker discovery often focuses on statistical correlation without addressing biological significance.
- Methods are often tedious and time consuming, with poor results in terms of usable outcomes.

Methodology

- Integrate experimental data (OMICS, tissue images, clinical assays) to generate a semantic data correlation network in an extensible experimental ontology using thesauri and class unification.
- Map and merge public data resources (including canonical pathways and functional biosystems) dynamically for ad-hoc mechanistic enrichment and knowledge expansion using formal concept.
- Selectively merge ontologies (or parts of ontologies) and add data classes, instances and relationships via direct SPARQL endpoint queries, without the need to incorporate entire database content.
- Add information from drill-out searches to semantic RDF-based (directly) or relational (via D2R or similar) public databases where no SPARQL endpoints are available.
- Systematically enrich the data network by merging additional public knowledge resources to create mechanistic and functional connections between the experimental correlation network and knowledge networks to validate the underlying biology.
- Generate a biomarker model from visual SPARQL arrays, validate them and assemble them into an Applied Semantic Knowledgebase for a specific biological problem (e.g. disease state).
- Use the Applied Semantic Knowledgebase (ASK™) for screening and decision-support.

Results

- This poster demonstrates how data from diverse experimental and public resources are merged into a coherent semantic framework for intuitive model building.
- Dynamic, flexible resource description framework (RDF) databases, ubiquitous SPARQL endpoints and "linked open data" provide rapidly deployed insights into complex biological functions through contextualizing of all available information.
- Establish biological models by generating queries interactively from user-selected networks, without requiring knowledge of the SPARQL query language.
- Sets of such SPARQL queries are captured and saved in arrays representative for a specific biological function and have been applied in decision-support in personalized medicine for predictive patient screening.

Discussion

- Semantic integration and merging of all available resources assures coherence and provides a solid base to relevant network analysis.
- As Applied Semantic Knowledgebases (ASK™) represent a novel approach towards complex biological responses, the qualification criteria to select classifiers and the algorithms for the statistical approach are crucial.
- Being able to create complex models in an easy, automated way, makes it universally applicable.
- By providing an array of network-based models, a high degree of confidence can be obtained as responses are accompanied by their closeness of fit to qualify the prediction.
- Using a web portal access to screen patients confidently removes informatics barriers in the clinics.
- While this concept already is actively applied, its function as knowledge application providing decision support impacts largely personalized therapy, treatment effectiveness, patient's quality of life improvement and new avenues for safer, more effective drug development.

Figures

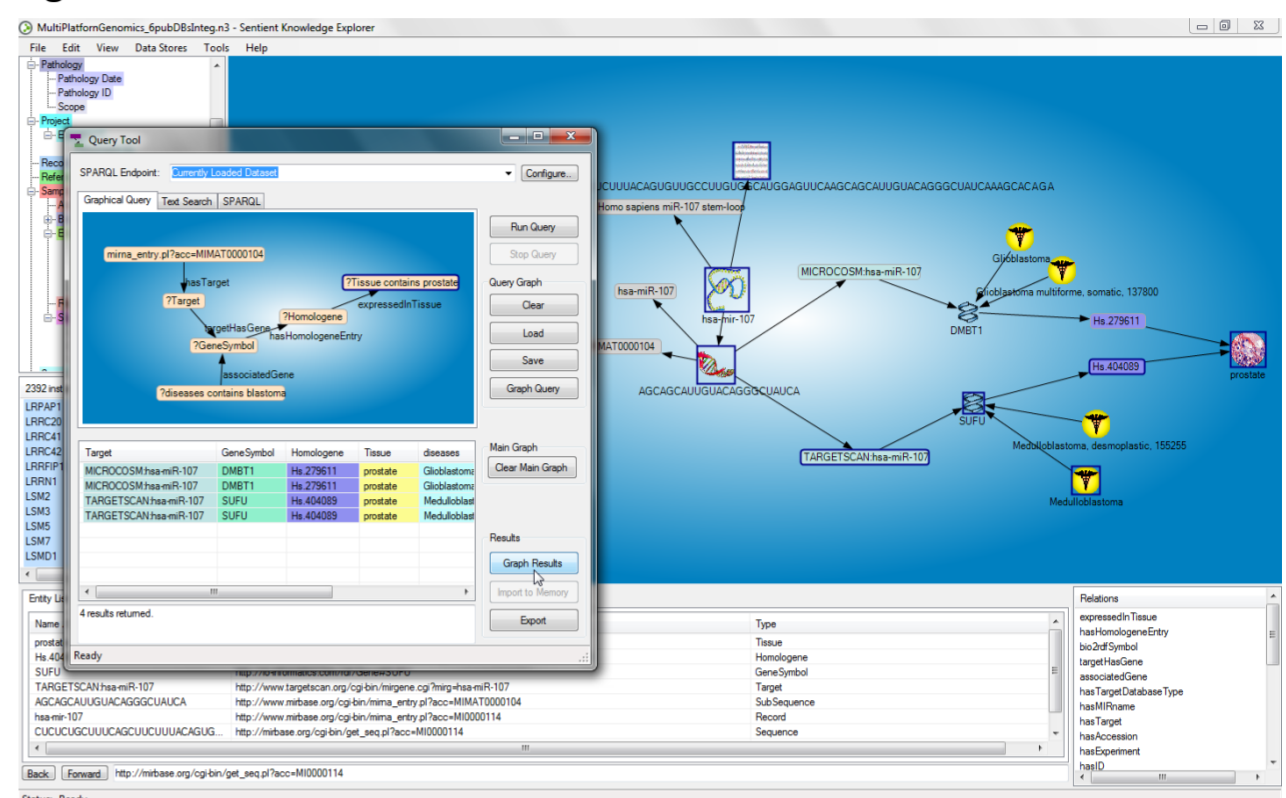


Fig.1: Semantic data merging:
Main: Knowledge Explorer connecting multi-platform genomics (RNA, DNA, MAs) with public resources for external validation of disease-gene relationships in treatment effectiveness studies (Prostate Cancer)
Left overlay dialog: Visual SPARQL query generation to build, refine and qualify semantic model

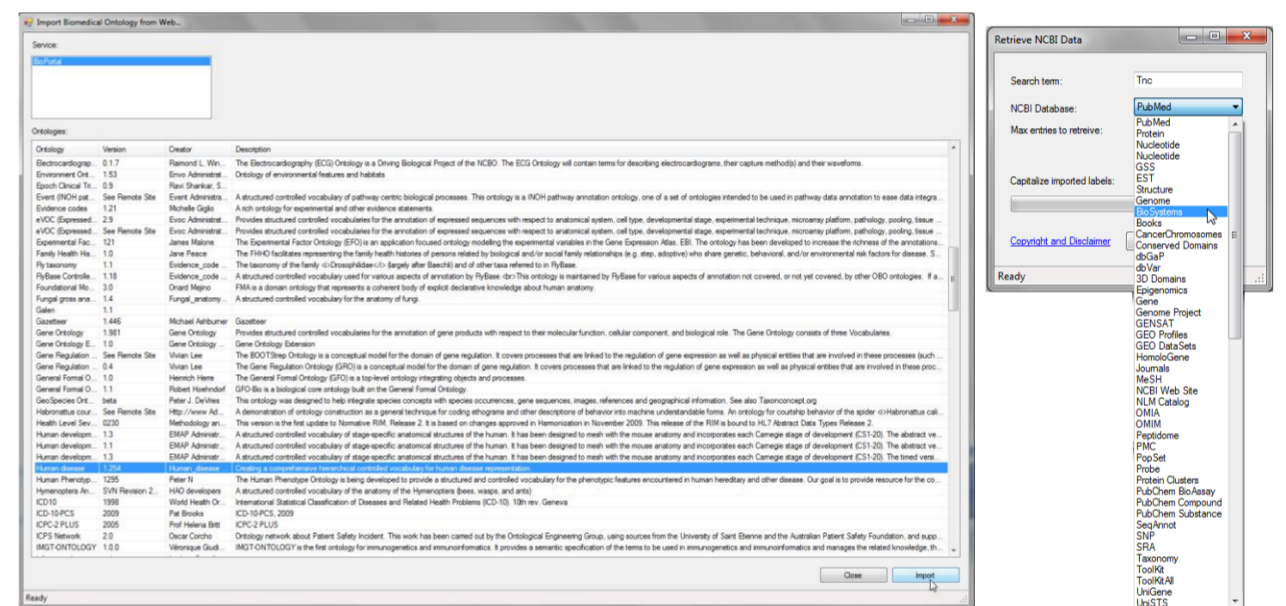


Fig.2: Public ontology and resource integration:
Left: Merging ontologies (example: NCBO's BioPortal) and linked open data via SPARQL endpoints
Right: Adding mechanistic knowledge from public resources (example: NCBI Entrez) directly in RDF

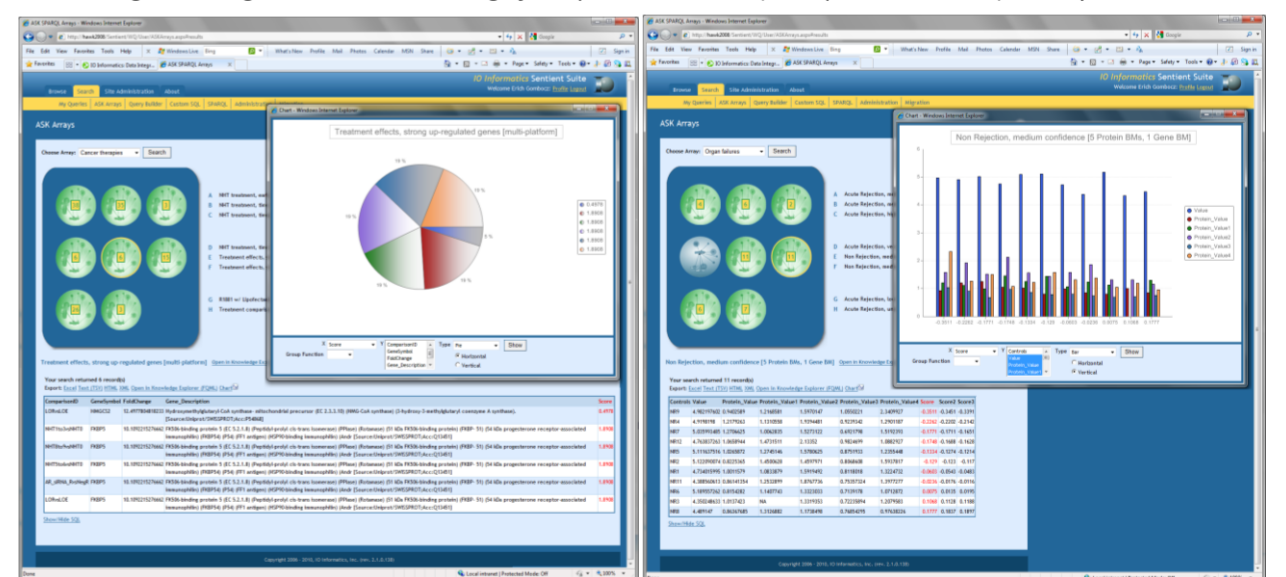


Fig.3: Actionable knowledge from integrated models:
Left: Comparison of effectiveness of combinatorial prostate cancer treatments (genomic biomarkers)
Right: Systems model-based decision support for likelihood of organ rejection in heart transplantation (Combinatorial multi-OMIC biomarker panel)

Conclusions

- Combining and reasoning across experimental data enriched with public domain knowledge provides a new level of insights in complex biological responses. Applied Semantic Knowledgebases (ASK™) uses arrays of network models via an easy-to-use web portal to provide confident decision support for complex biological questions.
- Being able to use, share and apply knowledge via an intuitive web tool hides the underlying complexity. Assessing the closeness of the model to each patient's profile with confidence, is changing the way knowledge is built, refined and applied in life sciences and medicine.

Acknowledgements

The authors thank Colleen Nelson (collaborating Prostate Cancer Centres) and Bruce McManus (PROOF and iCapture Centres) for their leading research and datasets. Many useful discussions in IO Informatics' Working Groups on "Semantics in Life Science" and "Informatics for Personalized Medicine" are attributed for helpful considerations. The authors would like to acknowledge the Working Group members Jonas Almeida, Alan Higgins, Pat Hurban, Bruce McManus, Ted Slater, Mark Wilkinson, Uwe Christians, Jack Collins, Dan Crowther, Amar Das, Herb Fritsche, Kathy Gibson and David Stanley for their various contributions.