Ontology Integration: In the Framework of Translational Informatics

Sirarat Sarntivijai[1], Yongqun He[1,2], Matthias Kretzler[1,3], and Brian D. Athey[1,4]

[1]The National Center for Integrative Biomedical Informatics and the Center for Computational Medicine and Bioinformatics, [2]Department of Microbiology and Immunology, [3]Department of Internal Medicine, and [4]Department of Psychiatry, University of Michigan, Ann Arbor, MI 48109, USA

Abstract

Biomedical ontology (bio-ontology) was first created out of the needs for systematic annotation. Most bio-ontologies residing in the OBO Foundry today were created *de Facto* at the laboratory of origin. Therefore, computing with logical reasoning embedded in individual bio-ontology can be challenging due to the divergence of individualism, especially when mapping multiple bio-ontologies for knowledge discovery. While such reusability and interoperability for knowledge transfer and discovery should be promoted, working with multiple bio-ontologies requires a sophisticated operating model that can overcome the issues of structural definition discrepancy of ontologies describing similar elements in the same domain, inconsistent and error-prone information within an ontology, and bridging across different information layers. We demonstrate that by mapping and integrating bio-ontologies of different biological layers from molecular genotype to molecular phenotype to clinical phenotype, bio-ontology processing plays an important role in knowledge discovery. Examples of use cases given in this study are integration of vaccine ontology in health care research, and using ontology integration to identify key disease factors of Diabetic Nephropathy. The framework proposed here utilizes graph matching theory, natural language processing, and ontology alignment to create a novel approach of ontology integration that drives ontology processing forward from annotations to computations, to translations for the next-generation translational informatics.

Motivation

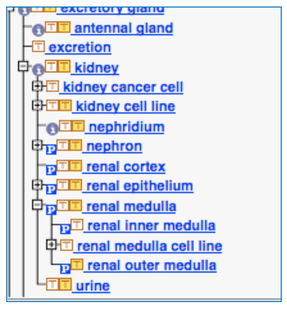
Integration of biomedical informatics knowledge has always been an important tool contributing to understanding driven-biology problem. Although all of these informatics studies focus on biomedical domain, they vary in the context of studies. Rapid growth of available information does not always promise a solution to obtain knowledge. In this study, we describe a strategy to extract knowledge from the pool of data using ontology and natural language processing, and discuss the current issues in limitation of knowledge integration for an improvement in the on-going study.

Overview

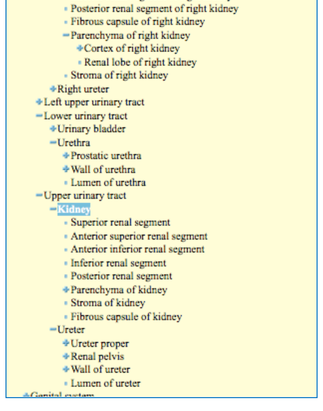
- **Goal** – practical integration and information retrieval of data from ontology and literature for an in-depth understanding of a driven-biological problem being asked.
- **Challenges** – Contradiction of ontology definition and problem of standardization, very large information set from various sources, method to find such overlapping amongst the different pieces of biomedical informatics and their interactions.
- **The approach** – model for biomedical informatics integration

Challenges

- **The controversy of the practice of ontology**
  - *"Divergence of Individualism"*
  - Reusability-Interoperability VS. Bottom-up Creation
  - One domain, multiple ontologies, multiple structures
- **Social engineering of the biomedical community**
  - Enforcement VS. Incentive
  - Case study of MIAME: GEO requirement to conform MIAME
- **BRENDA**
  - Glomeruli (listed under Nephron)
  - Tubulointerstitium not listed
  - Should Glom. & Tubul. be listed under 'Cortex'?



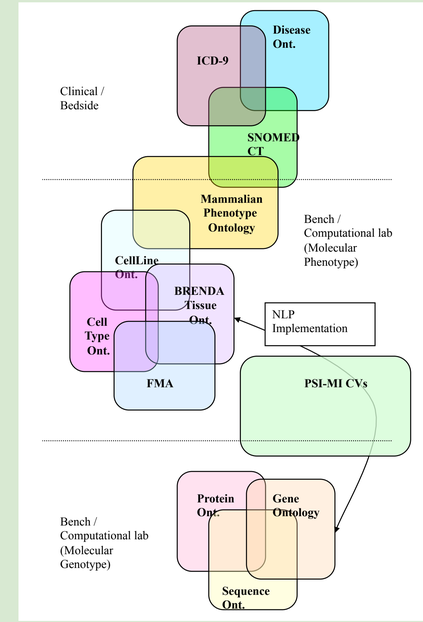
- **Foundational Model of Anatomy**



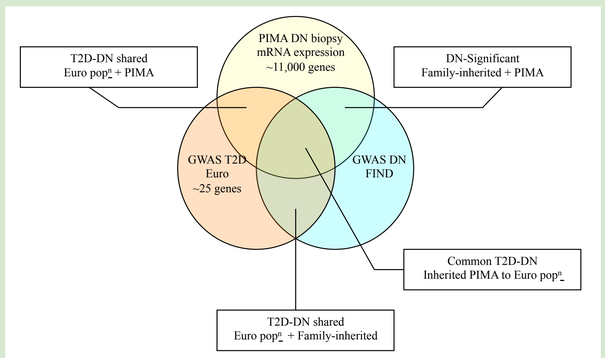
The Framework/Model

Use Case 1: Diabetic Nephropathy

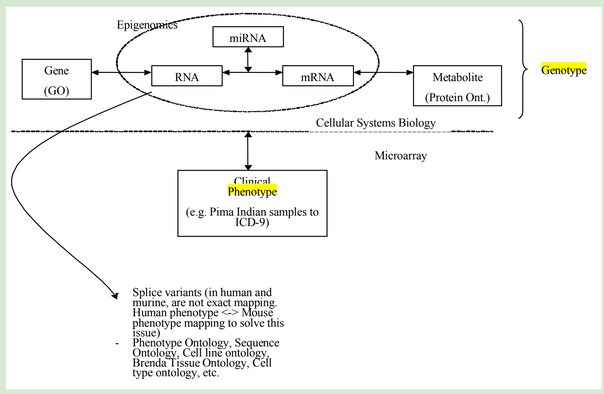
- One domain, heterogeneous data
  - PIMA Biopsy: mRNA expression (cortex/glum/tube) + clinical phenotype data + morphometric data
  - GWAS - SNP study on T2D
  - GWAS - SNP study on Diabetic Nephropathy

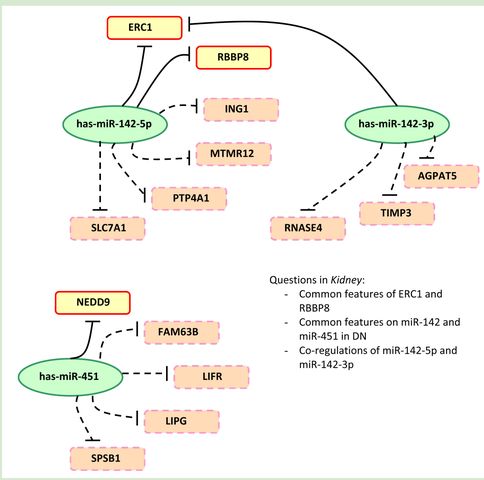


Multi-dimensional data: what are we looking for?



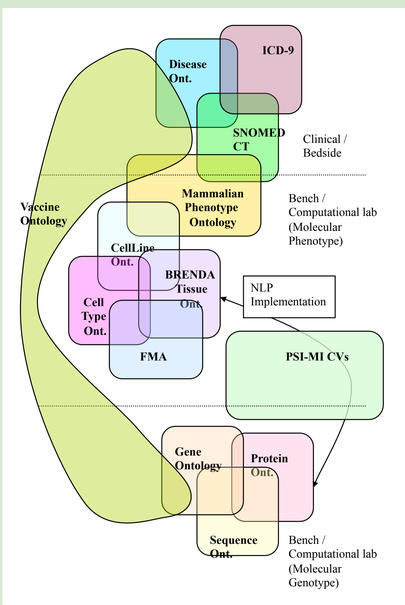
Systems Bio -> Transcriptome -> Metabolome



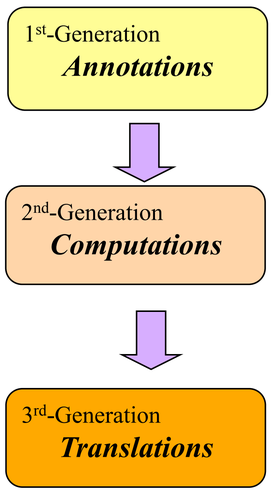


Use Case 2: Vaccine Ontology

- Getting to the interactome
- Integrating VO into Ontology mapping framework
- The low-hanging fruits
  - Cell line ontology <-> Vaccine ontology
    - How do different subclones respond to a vaccine?
    - How do different derivatives of a vaccine affect one cell line?



The Practice of Biomedical Ontology



Individual ontology created for the purpose of annotation, no real connections to other ontologies

Mapping of 2 or more ontologies requires a significant higher level of computation. Useful information can be extracted; advanced indexing system can be implemented.

High-level integration is required to drive towards translation. Implementation requires multiple computational methods. Linking low-level molecular genotype to molecular phenotype to clinical phenotype is the key mechanism.

References

- Sarntivijai, S., Ade, A.S., Athey, B.D., States, D.J. (2008) A bioinformatics analysis of the cell line nomenclature.. *Bioinformatics,* **24**(23), 2760-2766.
- Kingsmore, S.F., Lindquist, I.E., Mudge, J., Gessler, D.D., Beavis, D. (2008) Genome-wide association studies: progress and potential for drug discovery and development. *Nat. Rev. Drug Discovery.,* **7**, 221-230.
- Zeggini, E., Scott, L.J., Saxena, R., Voight, B.F. (2008) Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat. Gen. Adv. Onl. Pub.* 1-8.
- Bodenreider, O. (2008) Biomedical Ontologies in Action: Role in Knowledge Management, Data Integration, and Decision Support. *Yearb. Med. Inform.*, 67-79.
- Burgun, A., Bodenreider, O. (2008) Accessing and Integrating Data and Knowledge for Biomedical Research *Yearb. Med. Inform.*, 91-101.

Acknowledgements

This work has been done with the support from NIH under grant U54 DA021519 for the National Center for Integrative Biomedical Informatics.

We would also like to thank Yongsheng Bai, Viji Nair, and Inhan Lee for their experimental support and insights.