Metabolomics: Open Access
ISSN: 2153-0769

Elkin et al., Metabolomics 2012, 2:4
DOI: 10.4172/2153-0769.1000e115

Editorial Open Access

Completing the Metabolome

Peter L. Elkin[1]*, Mark S. Tuttle[1] and Steven H. Brown[2,3]

[1]Center for Biomedical Informatics, Mount Sinai School of Medicine, New York, NY, USA
[2]Department of Medicine, Vanderbilt University, TN, USA
[3]Department of Veterans Affairs, Vanderbilt University, TN, USA

At the turn of the millennium, genomics was in full swing as scientist's world-wide worked to sequence the human genome. During this time period, we were already aware that the sequenced genome did not tell the entire story. Scientists had begun to discuss functional genomics and metabolomics [1,2]. Of course we have long been aware of metabolic pathways. We have researched and taught pathways such as the Krebs Cycle for generations. In 1996 the Kyoto Encyclopedia of Genes [3] and Genomics (KEGG) published version 1.0 of their online compendium of pathways. KEGG has grown to approximately 165 metabolic pathways (Figure 1) out of a total of 425 metabolic, regulatory and signaling pathways [4]. Despite this progress, many researchers suspect that known pathways may not be completely understood and that additional metabolic pathways have yet to be discovered.

The search for this additional information takes many forms. Knowledgeable researchers follow the progression of their research findings to push the boundaries of our understanding which is a slow but productive process. The medical literature has been online since the 1960's and contains the results from millions of biomedical experiments. With the advent of high performance computing and advances in Ontology and data mining we are now able to ask questions of large datasets such as the biomedical literature.

In the past, data mining was typified by finding co-occurrences between potentially interesting concepts within a corpus of text, such as an article containing information regarding a specific gene and any associated diseases. This data mining approach does not provide the relationship between the gene and the disease, which is an important failing of this technique. The gene may regulate the disease (either up or down regulation), may effect a metabolic pathway that relates to the disorder of interest or may be correlated through some confound or it may simply be co-located with the gene that is truly predictive of the disease that you are studying.

With the advent of contemporary Ontologies combined with automated relationship abstraction we have the ability to ask more semantically complete questions of the medical literature and to draw conclusions or at least interesting hypotheses from data retrieved across two or more articles. Figure 2 shows a graphical rendering of the Cancer disorders hierarchy in SNOMED CT, a large general medical Ontology maintained by the International Healthcare Terminology Standards Development Organizations (IHTSDO) [5]. Ontologies have become repositories for the representation of significant sets of biomedical knowledge. Once the full text contents of the medical literature has been indexed by nationally standard Ontologies we can mine the result for semantically useful information such as a polymorphism of a gene that down regulates a metabolic pathway which causes a particular disorder. This information may be contained in one or more articles. Using the Ontological representations we could ask questions such as what other genes does this gene up-regulate or if this gene is turned off how does it affect protein synthesis? For a gene known to affect a disease by a specific mechanism one might ask what other disorders are affected by this mechanism of action and therefore may also be related to a polymorphism in this gene?

There are approximately 6,000 articles published every month in the biomedical literature. There are >19,000,000 citations in MedLine [6]. Researchers do not have the time or ability to follow all of the research results that relate to their area of research whose metabolic functions may span multiple target disorders. Ontologic knowledge, combined with automated entity and relationship abstraction via natural language processing, allows us to ask more semantically complete questions of the medical literature and to draw conclusions or at least interesting hypotheses from data retrieved across two or more articles. Data mining techniques using natural language processing technologies aimed at knowledge representation with standardized Ontologies have the potential to find novel synergistic information contained within single or across multiple articles. In our recent article, entitled "BioProspecting the Bibleome: Adding Evidence to Support the Inflammatory Basis of Cancer" we discussed the use of this technique to identify genes related to the basic development of cancer or apoptosis [7].

We believe that the scientific community should adopt a program which densely indexes the free text of all articles and makes the resultant database freely available to investigators for rapid cycle research and development. Such a capability would support novel hypothesis generation and additional in silica testing. This environment would facilitate meta analyses leading to greater understanding of research results from a series of experiments. These techniques when combined with basic science experimentation become a powerful mechanism to speed our ability to generate novel research results including target identification leading to drug discovery and in so doing to expand our ability to positively affect the human condition.

References

1. Wixon J, Brancia F (2000) Meeting highlights: beyond the genome 2000: the 18th International Congress of Biochemistry and Molecular Biology. Yeast 17: 314–321.
2. Downs DM, Escalante-Semerena JC (2000) Impact of genomics and genetics on the elucidation of bacterial metabolism. Methods 20: 47-54.
3. Kanehisa M (1996) Toward pathway engineering: a new database of genetic and molecular pathways. Science & Technology Japan 59: 34-38.
4. http://www.kegg.jp/kegg/atlas/?01100
5. http://www.ihtsdo.org/
6. Névéol A, Wilbur WJ, Lu Z (2012) Improving links between literature and biological data with text mining: a case study with GEO, PDB and MEDLINE. Database (Oxford) Print 2012.
7. Elkin PL, Frankel A, Liebow-Liebling EH, Elkin JR, Tuttle MS, et al. (2012) Bioprospecting the Bibleome: Adding Evidence to Support the Inflammatory Basis of Cancer. Metabolomics 2:4.

***Corresponding author:** Peter L. Elkin, M.D., MACP, FACMI, Center for Biomedical Informatics, Mount Sinai School of Medicine, New York, USA, Tel: 212-860-3837; Fax: 212-824-2329; E-mail: ontolimatics@gmail.com

Received June 29, 2012; **Accepted** July 03, 2012; **Published** July 05, 2012

Citation: Elkin PL, Tuttle MS, Brown SH (2012) Completing the Metabolome. Metabolomics 2:e115. doi:10.4172/2153-0769.1000e115

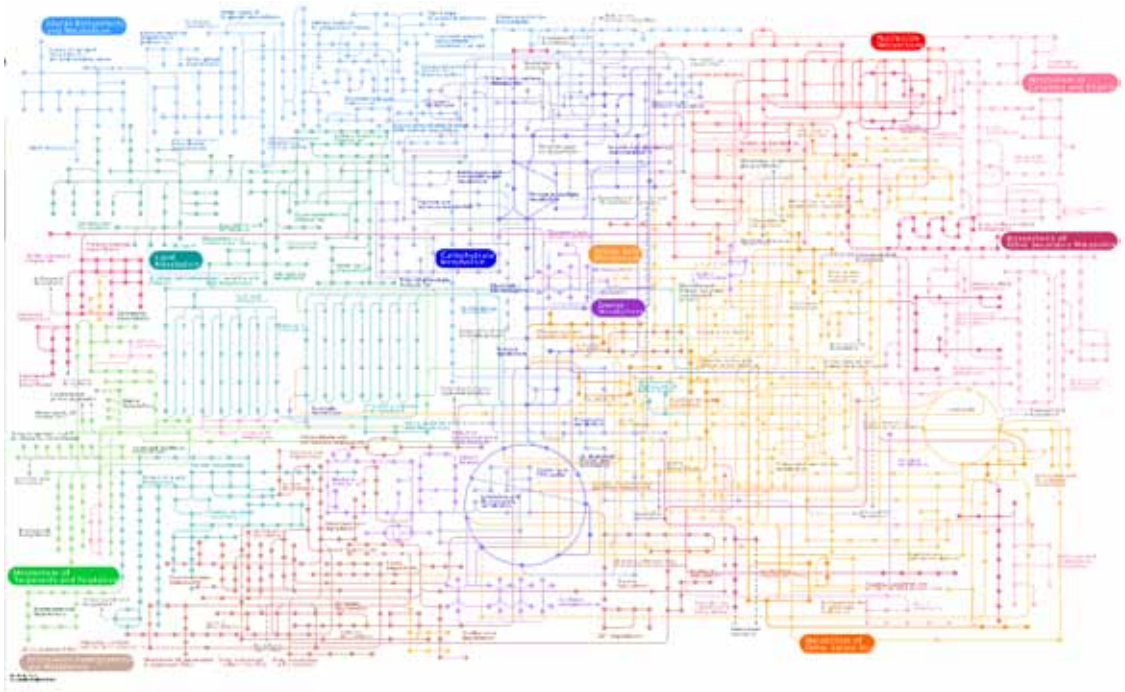

Figure 1: KEGG map of Metabolic Pathways.

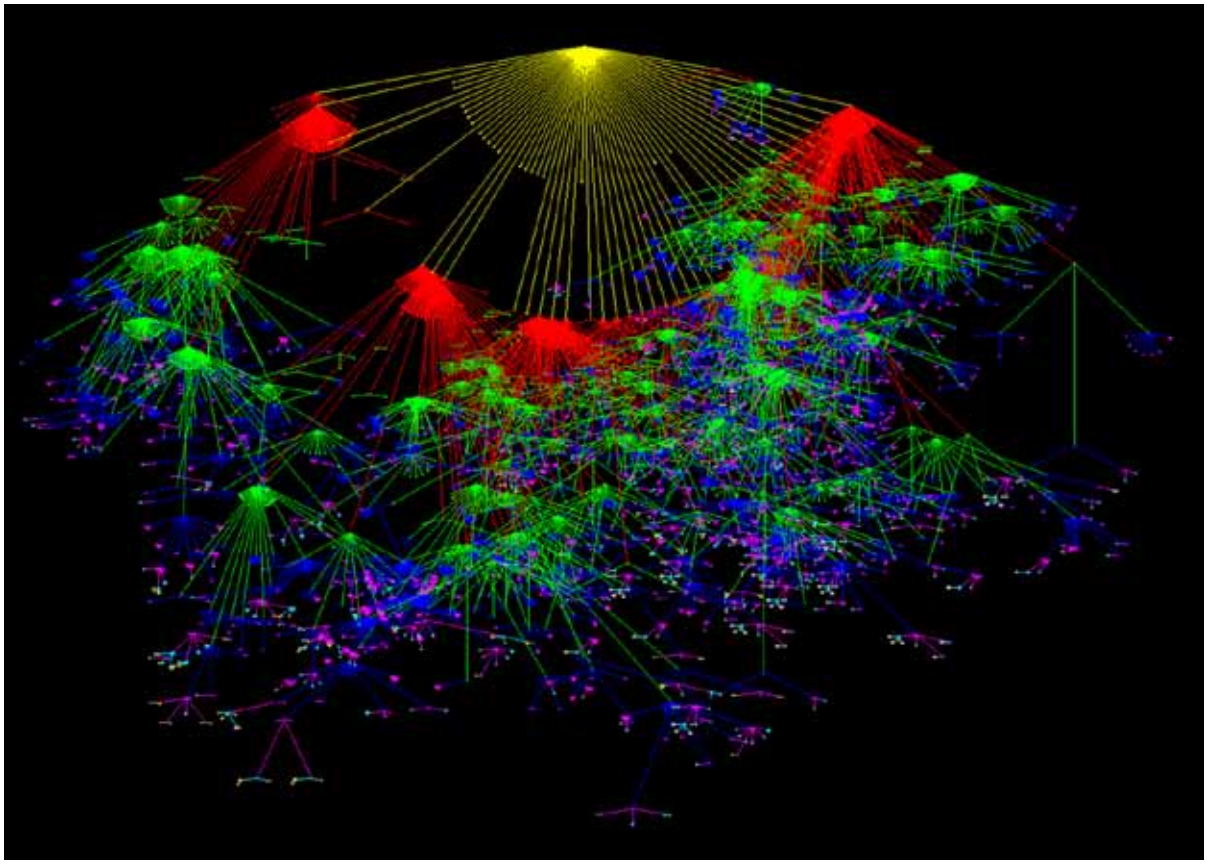

Figure 2: Graphical Representation of the Conceptual Distance between Cancer Diagnoses in SNOMED CT.