Relevance-Ranking and Inferencing of Disease Phenotype, Etiology, Pathophysiological Mechanism, and Therapeutic Opportunities Using Biomedical Semantic Web Methodologies

Ranga Chandra Gudivada[1,2], Xiaoyan A. Qu [1,2], Anil G Jegga[2,3], Eric K. Neumann[4] ,

Bruce J Aronow[1,2,3]

Departments of Biomedical Engineering[1] and Pediatrics[2], University of Cincinnati and Division of Biomedical Informatics[3], Cincinnati Childrens Hospital Medical Center, Cincinnati OH-45229, USA

Teranode Corporation [4], Seattle, WA 98104

An important goal for biomedical research is to elucidate causal, modifier, and suppressor networks underlying human disease. Integrative genomics-based approaches have shown considerable success in the identification of linked pathways and biological processes in normal and disease states. However, a huge bottleneck for the accomplishment of deductive and predictive medicine is the failure to represent knowledge and observations in a sufficiently standardized manner to permit assertions of causality between disparate disease studies and biomedical knowledge. Semantic Web principles, standards and technologies provide an ideal platform to integrate such heterogeneous information and allow the detection of implicit relations embedded in biomedical and genomic datasets. Semantic Web query languages such as SPARQL can be effectively used to mine the biological entities underlying complex diseases through richer and complex queries on this integrated data. First order results are generally large and unmanageable and we have thus sought to develop techniques to rank relationships based on specificity and uniqueness. Such ranking can be used to prioritize inferred disease–gene, disease–pathway or disease–processes novel relationships. We implemented an existing semantic web-based knowledge mining technique which not only discovers underlying genes, processes and pathways of diseases but also determines the importance of the resources to rank the results of a search while determining the semantic associations. For preliminary analysis we have sought to model and rank disease entity relations by integration of phenomic and genomic network entities mining UMLS, OMIM and genomic databases.