Incorporation of Ontology-driven Biological Knowledge into Cardiovascular Genomics

Huiru Zheng*, Haiying Wang and Francisco Azuaje

Co. Antrim, N. Ireland, United Kingdom

The integration of multiple biological sources is becoming prevalent in post-genomic research. An important task toward that goal is the quantitative assessment of relationships between relevant predictive resources. Based on the exploitation of the Cardiovascular Gene Ontology (CGO), this study presents a system that enables the incorporation of (CGO) similarity knowledge into cardiovascular research. The implementation of the system is based on the combination of biological function annotations provided by the CGO for more than 4000 genes associated with cardiovascular processes and topological features encoded in the Gene Ontology (GO). As a first step, the number of cardiovascular-related genes associated with each GO term and its child terms was computed. By calculating the probability of finding a child of a GO term in the CGO database, the information content associated with each GO term was then established. In the context of cardiovascular-related annotations, term-term similarity within each of the GO hierarchies, i.e., molecular function, biological process and cellular component, is computed using three GO-driven similarity measures (Resniks, Lins and Jiangs metrics). These provide the foundation for the estimation of semantic similarity between cardiovascular-associated genes. The system allows users to retrieve between-gene similarity using a single query or batch query mode. This study contributes to the development of automated methods for supporting annotation tasks, such as the generation of new annotations for partially-characterized genes associated with cardiovascular disease. The system can also be used to support functional prediction tasks in cardiovascular genomics, such as the validation of gene expression analyses and the identification of false positives in protein interaction networks. As a case study, this paper investigated the relationship between gene expression correlation and functional similarity for a list of 247 priority cardiovascular genes. The system is freely available for non-profit use on request from the authors.