

Interpreting Gene Lists Using a Literature-Derived Protein-Interaction Network

Pankaj Agarwal

Bioinformatics, GlaxoSmithKline

King of Prussia, PA

(Joint work with Dilip Rajagopalan)

A number of omic technologies such as transcriptional profiling, proteomics, literature searches, genetic association, etc. result in the identification of lists of important genes. Interpreting such gene lists and relating them to pathways is a challenging task. Databases of biological relationships between thousands of mammalian genes can help in deciphering omics data. A network of these relationships may then be searched for subnetworks consisting largely of interesting genes from the omics experiment. This subnetwork helps in deciphering the underlying pathways. I will discuss a heuristic algorithm and a scoring function to discover such subnetworks that works well on both simulated data and data from known pathways ([Rajagopalan and Agarwal, 2004](#)). The scoring function is an extension of Ideker et al. ([Bioinformatics, 2002](#)). The method works on reasonably complex curated networks containing about 15,000 biological entities (genes, complexes, & metabolites), and over 70,000 biological relationships. This method can also pick up a pathway signal from data including a moderate amount of noise. I will also show examples of use within the pharmaceutical context.

I will also present a number of open problems which if solved would have a significant impact on developing disease therapies.