NETwork-Based Analysis of Genomic variations

Technology #cu14362

Identifying and understanding the complex interactions that underlie widespread human phenotypes is amongst the most important challenges in medicine and biology. Two or more different genes can lead to the same phenotype or perform the same function. This has important consequences in medicine and pharmaceutical treatments. For instance, if two genes perform the same or similar functions, and a drug only targets one of these genes, then it will not be effective. This technology can help solve such problems. The NETBAG phenotype network finds relationships between genes by using a naïve Bayesian network. NETBAG scores the predicted likelihood that two human genes share the same phenotype. In doing so, NETBAG can uncover disease risk genes among a list of mutations observed in probands. The algorithm searches for cohesive clusters of genes perturbed by disease-associated genetic variations.

Motivation

Identification of complex molecular networks underlying common human phenotypes is a major challenge of modern genetics. We have developed a principled approach for integration of diverse sources of genome-wide genetic variation under a unified framework to address this problem. To identify affected molecular networks, we have developed an algorithm that searches for cohesive clusters of genes perturbed by disease-associated genetic variations.

Phenotype network

The NETBAG algorithm hinges on our previously described phenotype network, in which each pair of human genes are assigned a score proportional to the likelihood of a shared genetic phenotype. The likelihood contains a naive Bayesian integration of various protein-function descriptors: shared annotations in Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), protein domains from the InterPro database, tissue expression from the TiGER database; direct protein-protein interactions, or shared interaction partners in a number of databases (BIND, BioGRID, DIP, HPRD, InNetDB, IntAct, BiGG, MINT and MIPS); phylogenetic profiles and chromosomal co-clustering across genomes.
Network search

Among a list of provided genes, NETBAG will search for a strongly-interconnected subset of genes. Starting with each input gene as a seed node, a greedy search algorithm will choose the most strongly-connected gene to add to the candidate set. Candidate networks are assigned a score based on a weighted sum of their edges, representing the likelihood that the respective genes participate in the same genetic phenotype. Network significance is then determined by comparing this score to a distribution of scores obtained by applying the same search algorithm to sets of random genes.

Publications


Inventors

Dennis Vitkup