

Validating gene expression clusters using the Gene Ontology and other external evidence

Stathis Sideris

Abstract:

A method was developed to allow the use of knowledge on protein functions and protein-protein interactions to identify coherent sets of co-regulated genes suggested by the clustering of gene expression profiles. This was achieved through the development of a gene expression clustering quality metric, which judges the tightness and separation of gene expression clusters, thus providing a quality measure on a clustering or a per-cluster basis. Cluster tightness and separation are assessed by harnessing the manual annotations provided by the Gene Ontology, enriched using integrated biological information available through an in-house data warehouse (BioMap). The metric was tested on a human B-cell gene expression dataset and refined on the basis of the results produced.