Abstract

- Traditional gene enrichment methods for assigning functional attributes to a group of genes identified in an experiment are limited to using Gene Ontology [2] data.
- To uncover broad patterns and determine if gene subset shows over-representation of specific characteristics, current methods need to be extrapolated to other ontologies [1].
- We present here a computational framework for enrichment analyses using experimental datasets from Gene Expression Omnibus [3] by propagation over context and non-context functional maps to Gene Ontology.
- Our preliminary results show that there is a definite need for such translational research to allow biological scientists to better interpret the results of their experiments.

Introduction

- Biological researchers have generally used Gene Ontology [2] to complement microarray analysis.
- Once low level analysis is complete and a group of differentially expressed genes have been identified, enriching these groups for ontology attributes allows researchers to find over-representation of a biological characteristic.
- We enhance such analysis to express enrichments on other ontologies by using pre-computed functional mappings to Gene Ontology.

Methods

- We consider experimental abstracts and raw data from Gene Expression Omnibus.
- We use our in-house pipeline (Figure at Bottom) to compute context and non-context based links using correlation metrics (Figure at Right) present in the abstract [4].
- After determining significant over- or under-expressed genes from raw data, we propagate enrichments using mappings to Gene Ontology.
- To circumvent the complexity of enrichment space, we use a depth first, branch and bound heuristic to prune insignificant attributes.

Results

- We consider 200 ontologies [5], containing about 3 million terms, and about 15,000 experimental abstracts from Gene Expression Omnibus.
- We then apply our in-house algorithm to integrate Gene Ontology (24,987 concepts) with all other ontologies (Figure 3) within the context of Human Disease (12,033 concepts).
- Further, using these maps we compute enrichments for various gene-lists available in the experimental datasets (Figure 4).
- Our preliminary results empirically indicated that our context-based mappings allowed us to leverage more information, and accordingly yielded more accurate descriptions of enrichments.

Conclusions

- We present here a principled framework for domain-wide identification of biological, pharmaceutical or disease based characteristics.
- Our work provides a new approach using translationalized functional spaces for the identification of significantly enriched attributes.
- Such an integrative approach can not help researchers use and apply quantitatively the available biological knowledge to their experiments and analyses.
- We plan to validate and scale the promise of such translational approaches on much larger datasets and experiments.
- For more information please see: http://bcl.med.harvard.edu

References