GO-SCAN: Analysis and Visualization of Gene Ontology Annotation

Gene Ontology Significant Collection of ANnotations

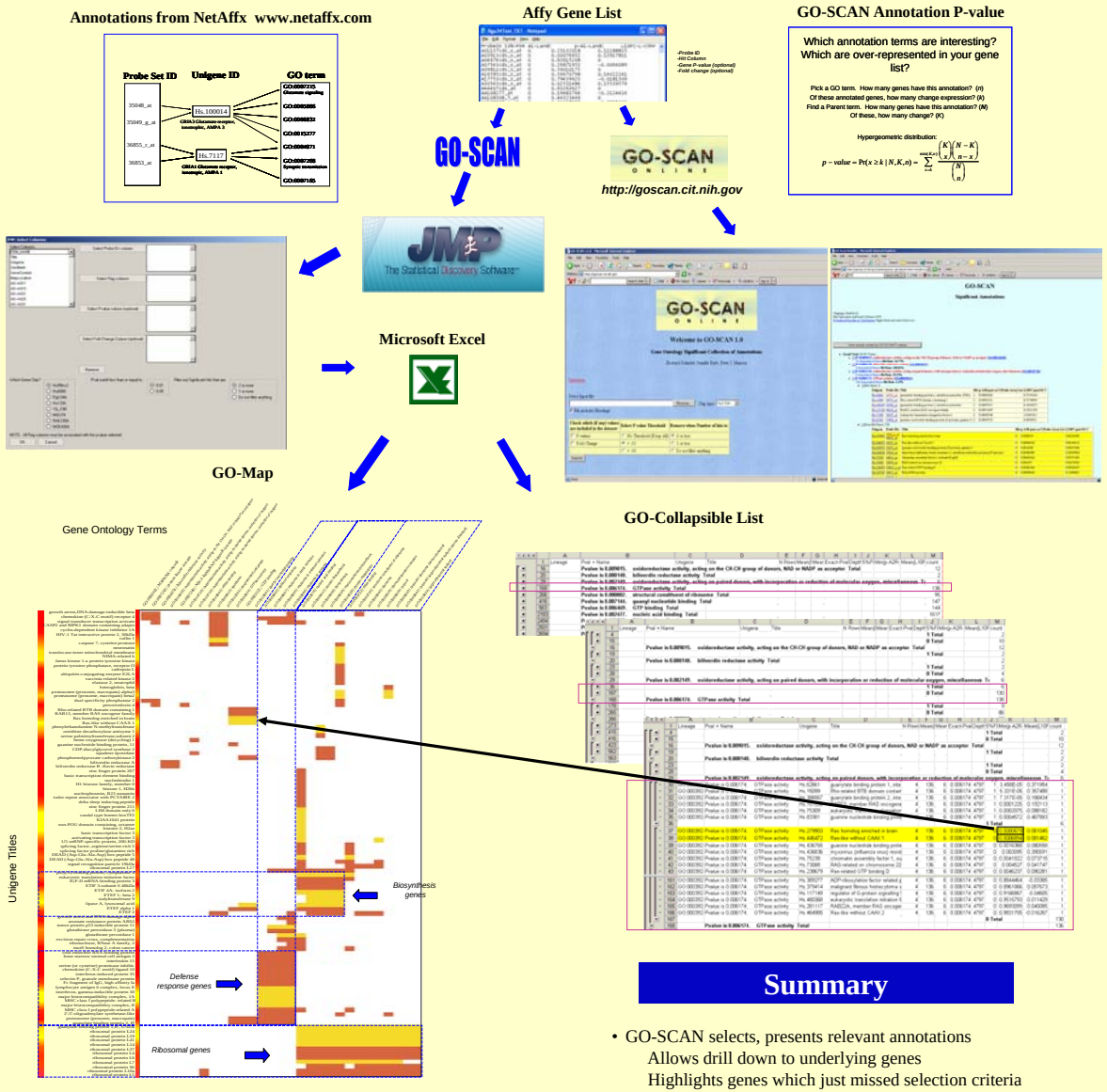
Jennifer J. Barb, M.S., Howard Schindel, Peter J. Munson, Ph.D.
Mathematical and Statistical Computing Laboratory, DCB/CIT/NIH/DHHS

Abstract

We have developed GO-SCAN, a bioinformatics tool that selects and presents relevant Gene Ontology (GO) annotations for a list of differentially expressed genes ("hit" list) from an Affymetrix microarray experiment. GO-SCAN provides two novel visualizations. The first is an expandable list structure presenting the relevant annotations in an order according to their position within the ontology hierarchy. Optionally, the fold-change or p-value can be used to order the genes within each term. This visualization is available through the web (GO-SCAN *online*) or as a text file resulting from a JMP script that must be formatted in MS Excel. The second visualization is a two-way hierarchical cluster heatmap (GO-Map) displaying genes ordered into groups sharing similar GO annotation. GO terms are clustered when they annotate a similar list of genes. We use a Fisher's exact test to select statistically relevant genes and report a p-value for each gene. Probe sets on the Affymetrix chip are first mapped to their corresponding Unigene ID, then to their GO annotations. Annotations used in GO-SCAN are derived from the Affymetrix web site and from the Gene Ontology consortium. In applications, GO-SCAN has allowed users to quickly discover themes and pathways related to the experimental perturbation or disease under study. GO-SCAN has been tested on human, rat, mouse and yeast chips. We illustrate GO-SCAN findings with a study of sickle cell disease.

Conclusion

GO-SCAN is a sophisticated bioinformatics tool used to assist in functionally annotating lists of differentially expressed gene lists. GO-SCAN *online* is easily accessible and works on any platform running any web browser. GO-SCAN can be used running JMP and then opened in Excel after some small formatting changes. GO-SCAN contributes to the analysis of the gene list by breaking it into functional themes where genes become parts of clusters and not single entities. Future work with the online version of GO-SCAN will add a graphical representation of the collapsible list thus giving the researcher an even greater of understanding of their gene lists.



Red boxes identify gene-annotations for genes selected with less than 5% False Discovery Rate (FDR) and additional filter criteria. Yellow boxes identify gene-annotations for genes with a relaxed selection criteria (7% FDR, no other criteria). Additional genes tend to fall into same annotation groups, providing additional evidence that they are differentially expressed. Major annotation groups include ribosomal genes, defense response genes and biosynthesis genes. The order of the genes and the order of the annotations has been adjusted to maximize the coherence of annotation groups using two-way hierarchical clustering. Original study of differential expression between peripheral blood mononuclear cells (PBMCs) from a group of sickle cell patients compared to comparable control patients. Jison ML, Munson PJ, Barb JJ, Suffredini AF, Talwar S, Logun C, Raghavachari N, Beigel JH, Shelhamer JH, Danner RL, Gladwin MT., Blood mononuclear cell gene expression profiles characterize the oxidant, hemolytic, and inflammatory stress of sickle cell disease. Blood. 2004 Jul 1;104(1):270-80. Epub 2004 Mar 18.

Summary

- GO-SCAN selects, presents relevant annotations
  - Allows drill down to underlying genes
  - Highlights genes which just missed selection criteria
  - Allows flexible formatting and editing of annotation list
- GO-Map organizes annotations and genes in clear visualization
  - Permits easy identification of major themes
  - Identifies "near misses" on gene list
  - Reduces redundancy in annotations
- GO-SCAN *online* gives flexibility to run GO-SCAN anywhere
- Available to NIH users at *http://affylims.cit.nih.gov*
- Also avaialble at *http://abs.cit.nih.gov/*